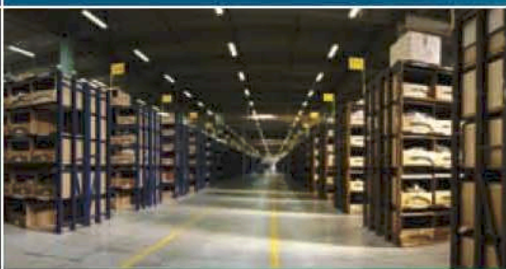
INTERNATIONAL FORUM ON SHIPPING, PORTS AND AIRPORTS (IFSPA) 2009



Post-Financial Tsunami: The Way Forward for Shipping, Transport and International Trade

Conference Proceedings

24 – 27 May 2009
The Hong Kong Polytechnic University
Hong Kong

CONFERENCE ORGANIZERS:

The Hong Kong Polytechnic University
香港理工大學

CY TUNG International Centre for Maritime Studies
董浩雲國際海事研究中心

Proceedings of the

International Forum on Shipping, Ports and Airports

(IFSPA) 2009

Post-Financial Tsunami:

The Way Forward for Shipping, Transport and International Trade

24 -27 May 2009

The Hong Kong Polytechnic University, Hong Kong

Proceedings of the International Forum on Shipping, Ports and Airports (IFSPA) 2009

24 – 27 May 2009

ISBN: 978-962-367-653-3

August 2009

Preface

This Proceedings covers the papers presented at the Third International Forum on Shipping, Ports and Airports (IFSPA) 2009, held at The Hong Kong Polytechnic University (PolyU) from 24 to 27 May 2009.

IFSPA is a reputed international conference dedicated to maritime, aviation and logistics research. This year, the Conference theme was "Post-Financial Tsunami: The Way Forward for Shipping, Transport and International Trade".

Building upon its previous success, IFSPA 2009 attracted more than 80 papers and presentations, (co-)authored by more than 100 scholars coming from 17 countries and territories. All manuscripts submitted to the Conference were peer reviewed by academic scholars around the world, and papers were only accepted with positive recommendations from reviewers.

With the massive support from academies and industries, IFSPA 2009 had a strong presence of world-known Keynote Speakers and Honorary Guests. We are also privileged to have hosted two Special Sessions "Dry Port Development in Asia and Other Regions: Theory and Practice" and "Logistics and Supply Chain Modelling Research" respectively with the United Nations Economic and Social Commission for Asia and the Pacific and the Logistics Research Centre of PolyU. The Conference also comprised 4 Industrial Forum and Academic Keynote Sessions as well as 16 Parallel Sessions.

Our packed programme successfully managed to thrash out some of the critical issues affecting the shipping, transport and international trade sectors arising from the global financial tsunami, and presented with recommendations which offer our participants a fresh angle to cope with the challenges ahead.

IFSPA 2009 is indebted to many people and organizations, notably the sponsors and partners. Without their generous support, a conference of this magnitude could never be materialized. Sincere thanks are also given to all speakers, guests, paper reviewers and, of course, authors. As the President of the Organizing Committee, I would also like to give my gratitude to all committee members, for their indigenous efforts striving to make this conference a success.

John Liu

Conference President, IFSPA
Director, C.Y. Tung International Centre for Maritime Studies
Chair Professor of Maritime Studies
Head, Department of Logistics and Maritime Studies
The Hong Kong Polytechnic University

Contents

IFSPA 2009 Organizing Committee

Conference President	Prof. John J Liu
Chairman	Dr Koi Yu Adolf Ng
Co-Chairs	Dr Xiaowen Fu Dr Ling Zhu
Committee Members	Dr Pengfei Guo Dr Meifeng Luo Dr Hengqing Ye Dr Tsz Leung Yip
Secretariat	Ms Joey Wong Ms Priscilla Lau Mr Justin Wong

Conference Sponsors and Partners

We would like to express our deep appreciation to the support of the following companies and organizations.

Conference Sponsors	Hong Kong Airport Authority ECORYS Research & Consulting
Conference Partners	Cathay Pacific Airways Limited The Chartered Institute of Logistics & Transport, Hong Kong Hong Kong Shipowners Association International Association of Maritime Economists Logistics Research Centre, The Hong Kong Polytechnic University Marine Department, Hong Kong SAR Government

List of Paper Reviewers

- Dr Stephen Cahoon
  Australian Maritime College, Australia
- Dr Ismail Cetin
  Dokuz Eylul University, Turkey
- Dr César Ducruet
  University of Paris 1 Panthéon-Sorbonne, France
- Prof. Henry Fan
  Nanyang Technological University, Singapore
- Prof. Mitsuo Gen
  Waseda University, Japan
- Prof. Paul T-W Lee
  Kainan University, Taiwan
- Prof. Chung-Lun Li
  The Hong Kong Polytechnic University, Hong Kong
- Dr Venus Lun
  The Hong Kong Polytechnic University, Hong Kong
- Prof. Katta Murty
  University of Michigan, United States of America
- Prof. Takeshi Nakazawa
  World Maritime University, Sweden
- Prof. Theo Notteboom
  University of Antwerp, Belgium
- Dr Anthanasios Pallis
  University of the Aegean, Greece
- Prof. Brian Slack
  Concordia University, Canada
- Prof. Dong-Wook Song
  Heriot-Watt University, United Kingdom
- Prof. Wayne Talley
  Old Dominion University, United States of America
- Prof. Jose Tongzon
  Inha University, South Korea
- Prof. Eddy van de Voorde
  University of Antwerp, Belgium
- Prof. Willy Winkelmans
  University of Antwerp, Belgium
- Dr Zhou Xu
  The Hong Kong Polytechnic University, Hong Kong
- Dr Jia Yan
  Washington State University, United States of America
- Dr Bibo Yang
  The Hong Kong Polytechnic University, Hong Kong

A probe into relevant strategies for railway in developing modern logistics

Yun Bao, Zhaoxia Yang, Haiying Li and Lingyun Meng

State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, 100044, China.

Abstract

Capable of transporting large volume of freight and passenger flows over long distance with efficient fuel consumption, railway plays a vital role for developing modern logistics, especially in China. In recent years, Chinese railway has taken unprecedented reforms and constructions, and has gained notable results, however, it is still facing many challenges. In this paper, based on the reform of railway transportation in China, first, we analyzed the currently existing problems of railway in the reforming process and also in the future, then we proposed the corresponding solution – to develop railway modern logistics. Furthermore, we discussed the methods to develop railway modern logistics, including establishing railway "trackless-station" by utilizing the freight stations which were abolished freight service in the reform. In addition, railway can establish logistics hubs and hub-and-spoke transportation networks as a leader role together with other modes of transportation with the development of multimodal transportation. Finally, we discussed how to reform the traditional function of railway freight transportation with the advantage of technology and network of railway, and how to change the railway freight transportation into modern logistics which integrates transportation, storing, packaging, distribution, information and so on.

Keywords: railway, modern logistics, trackless-station, railway logistics hubs

1. Introduction

In recent years, China railway has taken unprecedented reforms and constructions, such as the adjustment of railway transportation productivity layout(The State Department of Transportation of Ministry of Railways, 2006), the strategy of "two coordination and one construction"(coordinate stations with low volume of freight and dedicated lines, construct strategic loading sites) (Jifeng Liu, 2008), the large scale construction of passenger dedicated lines(Ministry of Railways, 2008), and so on, which offer favorable term for the development of railway, however, it is still facing many challenges. During the process of railway transportation productivity layout adjustment, many railway stations with low volume of freight were closed, in addition, the business of less-than-carload lot has been canceled from 2006. If no measures are carried out, these freight centralized reforms must cause inconvenience to the owner of low volume of freight. Meanwhile, the policy of constructing strategy loading sites focused the big customers rather than normal customers, which may make the owner of low volume of freight feel discrimination. Furthermore with the construction of the passenger dedicated line in China, it will enable passenger trains and freight trains travelling on different lines. At that time, a great deal of railway freight transportation capacity will be released and railway transportation companies will face the challenge of market pressure changing from monopolization to competition, then the questions is that how Chinese railway makes full use of freight transportation capacity and how provides whole trip service rather than the trunk line transportation. What's more, with the development of economic, railway stations, with the traditional function, can't meet the demand of transportation market. With the fierce competition of transportation market, the management conception will be outdated, and it is the inevitable tendency to the developing of railway freight transportation to reform the traditional function of railway freight transportation with the advantage of technology and network, and change the railway freight transportation into modern logistics which integrates transportation, storing, packaging, distribution, information and so on.

2. The Development State of Modern Logistics in Railway

Modern logistics is an integrated service of fully using modern technology and integrating the links of transportation, storing, packaging, distribution, information service and so on. Now, railway logistics is carried out by railway bureaus, transportation specialized corporations and railway diversified transportation enterprises. The structure is illustrated as below:

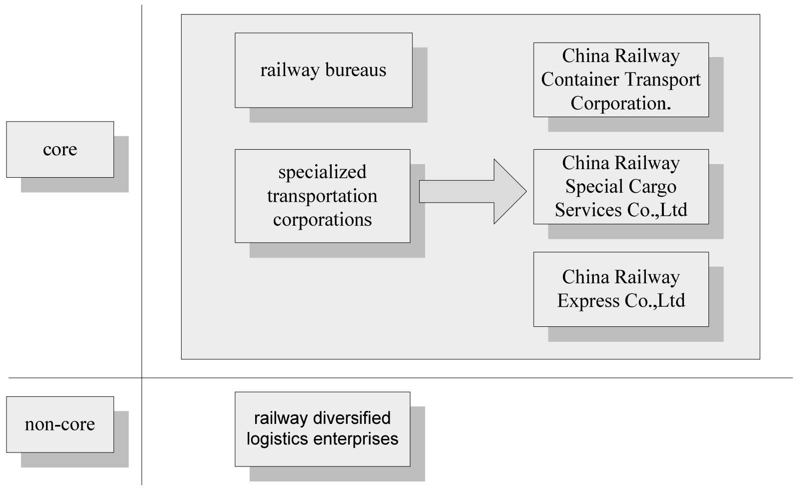


Figure 1: The Organization Structure of Logistics Development in Railway

2.1. The State of Railway Bureaus

Now, the main task for every bureau is trunk line transportation, so the development of logistics is very slow. However, some railway bureaus have realized the competition and began to pay attention to the development of modern logistics in railway. Such as Hohhot Railway Bureau, they have planed to construct six logistics strategy bases.

2.2. The State of Railway Corporations

China railway established the three specialized corporations (China Railway Container Transport Corporation, China Railway Special Cargo Services Co.,Ltd and China Railway Express Co.,Ltd) in the year-end of 2003 and defined that they are the railway specialized corporations who have the right of carry the goods on railway, meanwhile, they can provide the service of storing, loading and unloading, packaging, distribution, information and so on. China Railway Parcel Express Co.,Ltd and China Railway Express Co.,Ltd reorganized into China Railway Express Co.,Ltd in 2006. As the new principal parts of the market, the three specialized corporations have ranked among the logistics fields and entered the first thirty strongest logistics enterprises in China.

2.3. The State of Railway Diversified Transportation Enterprises

Railway diversified transportation enterprises are the enterprises that established by railway stations or railway sub-bureaus (before they were canceled) under the background of railway diversification. They were separated from the core transportation and competed in the market under the polity of the core and non-core transportation separation. Railway diversified enterprises involve many industries, such as logistics enterprises, enterprises that belong to logistics, business enterprise, construction enterprises, tourism, advertising, and so on. Among these enterprises, logistics enterprises is the enterprises which can provide two or more than two kinds of logistics service, while enterprises that belong to logistics are such enterprises which provide single or less service such as forwarding, storing, expansion service and so on (Jianneng Jiang, et al. 2006).

3. The Problem in the Development of Modern Logistic in Railway

Though railway has got a lot achievement in the process of developing modern logistics, there are many problems on organization, mechanism and cognition in the development, which even became restraining factors.

1. The less integration in the inner of railway itself. Railway didn't integrate and divide among railway bureaus, transportation specialized corporations and railway diversified transportation enterprises. Railway bureaus still engage in trunk line transportation, and don't control the collecting and dispatching of the goods, furthermore, they didn't meet the command of the market. Railway diversified transportation enterprises, for small scale and been scattered, are difficult to form scale effect, even though they can meet the market well.
2. The disparity (gap) in cognition. Railway departments simply consider that the expand service of transportation or storing is logistics, and don't integrate the links of modern logistics effectively. They also don't control the collecting and dispatching of goods, and can't meet the demand of the market to provide integrative service to the customers.
3. The timeworn management system. For a long time, with typical planned economy and timeworn management system, there are administrative barriers in different departments in the inner of railway. Because of the paucity of cooperation, it is different to exert the advantage of the network platform.
4. The low standardization of the technical equipments. As the history backend of railway transportation standardization managed separately and each department going its own way, railway technical equipments are difficult to connect with road, water or air effectively. This reduces the operation efficiency of modern logistics and increases unnecessary cost.
5. The bad service. As the exclusive patent in the market of railway, the whole industry is used to the boos from top to bottom. It lacks professional and technical personnel to engage modern logistics, especially operational stages. The works' professional competence is generally not so high that causes the low working performance and reaction speed and the low level of service.
6. The awareness of freight transportation marketing remains underdeveloped. For a long time, the capacity provided by China railway cannot meet the demand because of the resource layout and the capacity of railway, which caused the weak of the awareness of freight transportation marketing. With the improvement of railway network and realization of travelling on different lines of passenger trains and freight trains, a great deal of railway freight transportation capacity will be released, then the question is how China railway strength its marketing to gain more market shares.

Based on the above problems, we think that China railway should change the present state, make the best of technical-economical advantage to develop modern logistics in railway.

4. The Ways to Develop Modern Logistics for China Railway

4.1. Develop Railway Trackless-stations

During the process of the adjustment of railway transportation productivity layout, many stations with low volume of freight were closed(Table 1). To these stations, for one thing, the resource is wasteful; for another thing, the customers who delivered their goods at the stations in the past have to turn to highway or to railway stations that are a little farther to them, and these may cause the increase of the cost, furthermore, this may cause the loss of customers. Hence, we proposed to establish trackless-stations in the places where there is freight transportation demand but there are no railway stations to carry out logistics service in order to attract the owners with low volume of freight. The trackless-stations can not only provide traditional service, but also can provide logistics service, including goods collecting and dispatching, storing, packaging, distribution processing, information service, and so on.

Table 1: The changes before and after the adjustment of railway transportation productivity layout

	Before (June, 2003)	After (March, 2006)	Decrease
The number of vehicle depots	216	125	42.1%
The number of passenger transportation sections and train crew districts	56	33	41.1%
The number of locomotive depots	165	57	65.5%
The number of passenger car depots	44	22	50.0%
The number of freight car depots	100	29	71.0%
The number of permanent way sections	249	115	53.8%
The number of electricity sections	132	42	68.2%

Source: Haiying Li (2008)

The establish of trackless-stations can not only full use of the resource (space, personnel) of railway stations that have been abolished, but also provide storing, packaging, distribution, etc. as the form of modern logistics. It can reach the goals of the transportation as "door-to -door", which expend the service and the market of railway freight transportation and abstract more customers. The trackless-stations can intensify market competition, strengthen the role of railway radiation service and driving, service and support the development of regional economies effectively.

4.2. Establish Railway Logistics Hubs

The development situation of Chinese logistics is under the mode of area divided and disordered management. The plan and construction of logistics infrastructural facilities lacks necessary coordination, which caused the low supporting and compatibility of logistics infrastructural facilities and the weakness of the whole system. Moreover, the input in cooperating different type of transportation and connecting transportation system in different areas is fewer than foreign countries that logistics develops well. So in China, the development of integrated freight transportation hubs, logistics bases and logistics centers is so slow that it affects the development of logistics industry.

Capable of transporting large volume of freight over long distance with efficient fuel consumption, railway plays a vital role for developing modern logistics. The collecting and dispatching of goods in traditional railway is by district trains or pick-up and drop trains. Low efficiency wastes railway capability and also can't meet the customers' demand about time. Railway transportation productivity layout made the allocation of stations much more centralized; moreover the construction of passenger dedicated lines makes railway release much more capability; with the development of integrated transportation, more attention has been paid on the development of cooperating different modes of transportation. The reforms and constructions as above offer favorable term and guarantee for developing railway logistics, and railway should lead other modes of transportation to develop modern logistics hubs. The transportation between hubs is by railway, and the collecting and dispatching of goods in a hub is by road which is much more limber. This can strengthen road-and-railway coordinated transport to exert the advantage of railway—long distance and large volume. There are two type of road-and-railway coordinated transport: the road belongs to railway and the road doesn't belong to railway. Railway should expend the capability of the road that belong to it and strengthen the co-operation with road enterprises that do not belong to it.

We propose to classify railway logistics hubs as network railway logistics hubs, regional railway logistics hubs, and local railway logistics hubs.

1. Network railway logistics hubs
   Network railway logistics hub should establish at big city where traffic is very developed and the exchange activities are quite frequent, such as Beijing. To establish network railway logistics, railway must lead the other modes of transportation (road, water, air), employ multiple financing methods and take standard management to establish and develop large-scale socialized logistics hubs, furthermore, as the leader, railway should make it to the main based points to construct

special and socialized logistics network. Railway should make the best of the exist freight transportation hubs to develop network railway logistics hubs which are mainly depend on road-and-railway, and sea-and-railway.

2. Regional railway logistics
   Regional railway logistics should be established at one or more economic regions. The establishment of regional railway logistics is to meet the needs of the development of the local market economy and industries. Railway can establish the Yangtze River Delta logistics hub, the Zhujiang Delta logistics hub, and so on.
3. Local railway logistics hubs
   Local railway logistics hubs are on a small scale just as a local marshalling station in traditional railway system, but they can provide logistics service besides goods transfer and collecting.

4.3. Establish Hub-and-spoke Transportation Networks as a Leader

The hub-and-spoke transportation network is an importation effective network to integrate transportation resource and improve the use efficiency of the resource. The system of hub-and-spoke(HS) just like the wheel of bicycle, and it is a node-route system. In order to collect the quantity of good to achieve economies of scale , goods that from different original (spoke) to different destination(spoke) or from the same original to different destination should be collected to one middle place(hub) to transship, and then enjoy the service of discount through traffic. Today, there is growing evidence in the transportation literature that other deregulated industries such as the trucking industry and the freight industry generally have adopted some form of HS networking as method of delivery (Giovanni Nero,1999). It has been became the main tendency of the development of network structure among the cooperative third-party logistics providers (Zapfel Gunther, Wasner Michae1, 2002).

Based on big logistics hubs, transportation network of hub-and-spoke is a centralized system. In this system, the transportation is not from original to destination but to the hubs. Goods is first centralized at the hubs and then transported (Figure 2). The transportation of hub-to-hub can get economies of scale and decrease the cost.

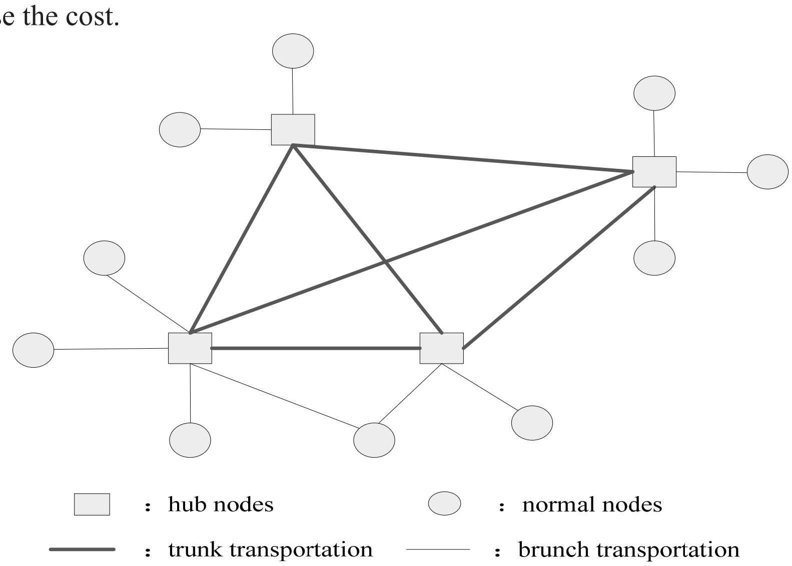


Figure 2: The System of Hub-and-Spoke

The structure of hub-and-spoke demands transport centralized in time and space. The whole process was managed by logistics devanning, sorting, transshiping and distribution in the hubs. The system of the hubs needs rational traffic division among different modes of transportation to keep seamless connect with each other in the whole logistics network, especially at the nodes. The distinguishing feature of the network is the collection of traffic flow in the main lines, and this is helpful to decrease the cost of the whole network. The system of hub-and-spoke owns the advantage of economies of scale and varietal, and can offer selective service to the customers and decrease the cost, which can intensify market competition, expend the economic hinterland and maximize the welfare of transportation enterprises and the areas.

4.4. Improve the Efficiency of Regroup

The efficiency of railway regroup includes the exterior regroup with different modes of transport and the inner regroup with different enterprises.

As an important part of national economic system, railway transportation meets the demands of social production, the circulation of commodities and people's travelling. There are five type transportation modes, namely railway, road, water, air and pipeline. Each transportation system has its own techno-economy characteristic features to meet different geographical conditions of nature and demands of transportation. The more development of economy, the more diversification of transportation demands, which needs different modes of transportation to meet the demands. The connection and integration of different modes of transportation must become the core of the construction integrated transportation system. Railway, road, water, air and pipeline must abandon the mode of developing by itself as past, and they should comply with the demand of the construction of integrated transportation system and the general requirements of promoting the integration and common development of transportation.

To railway itself, China MOR(Ministry of Railway) should integrate the departments of diversified logistics enterprises, forwarder and provide integrative service. Cooperation and division should be adopted to realize goods collecting, dispatching and transportation.

4.5. Strengthen the Integrative Logistics Service of Railway

The definition of integrative logistics service in logistics terminology is: The integrative logistics service is the result of the planning, designing and management of logistics programs according the demands of the customers. Railway should strengthen integrative logistics service to provide the program planning, designing, the process optimizing and operational management of the supply chains. Railway department should provide integrative logistics solutions for big customers and good logistics service, especially in the near future when it owns much more capacity on fright transportation.

5. Relevant Strategies for Railway in Developing Modern Logistics

Transportation is traditional function of railway. Based on improvement the functions of storing, packaging, distribution, distribution processing, railway should do much work as follows:

5.1. Change the Structure of Transportation Products to Meet Different Command of Customers

In order to meet the different level commands of customers, the departments of railway should provide the products with different arriving speed, different level of cost, and so on, based on an overall consideration of various factors, such as the time value of the goods, the whole transportation time and cost, the conveniece of the goods collecting and dispatching, and so on.

It is the main developmental direction for railway to provide long-distance, convenient, safety, and low cost service for customers. In recent years, in order to meet different market demands, China Railway Container Transport Corporation, China Railway Special Cargo Services Co.,Ltd and China Railway Express Co.,Ltd have went into the marketplace and competed, and they have developed many transportation service programs, such as the five-fixed trains, the through trains about mass freight, the parcel express trains, the parcel express trains about fresh and live freight, and so on. These products have got good social effects. Railway should expend products development to meet different demands of customers, for example, in order to meet the fast development of cities and the demands of exchange of consumer products, railway should consider to develop express though trains in different cities. In the major thoroughfare, in order to meet the demands of the goods with high value-added and high time limited, railway can organize freight container express trains with quickly loading and unloading on lines(Heng Du, Kai Liu, 2006).

5.2. Develop Modern Information System and Establish Logistics Information Network

Based on the present service, railway transportation departments need to strengthen the construction of information system to keep close touch with enterprise customers, give information about the place and state of their goods in the process of transportation. The information system focuses on decreasing the logistics cost and management cost, using information system to integrate different aspects of management to realize management information-based and improve the service level and capability of railway logistics. First, in order to maximized share railway freight transportation information resource and raise the utilization rate of resources effectively, railway logistics management information system should be established though advantage information technology. Meanwhile, railway department should develop storing in to the direction of intelligent and informatization. It can check the stores quickly in different storages in one computer by computer network, automatic recognition technique of bar code, radio-frequency technique, and so on, which can help manager to manage and allocate the goods. In the process of transportation, freight cars and goods can be followed though the system of TMIS, GPS, and GIS to get the information of goods at any time. In the process of the interchange of information, EDI should be adopt to increase the speed of file delivering and avoid the file typed repeatedly and wrong. Based on the actual conditions, the construction of railway information system should be promoted step by step.

Railway should realize intelligent train dispatching, socialization of passenger and freight transportation marketing, and modernization of management as soon as possible. Railway electronic commerce system should be established to provide all aspects of service, such as to reserve freight cars, to charge, to check the information of the goods when it is in the process of transportation, to inform customers when it arrived, to well accept and hear cases and to protect the legitimate rights of customers. The information and goods is blent together in traditional railway process. The owner of the goods has to come to the railway station to check much business, such as applying for freight cars, loading, unloading (or supervising the work), and so on. The work wastes much time of them. With the development of information, the traditional process doesn’t meet the commands. Railway departments should separate the flow of information and goods to decrease the cost. The owners of the goods don’t take part in the process of goods. After they make a (phone) call or get to the internet at home, their goods would arrive the destination. They also can get the information of the goods at home when the goods is on the way.

5.3. Strengthen the Logistics Infrastructure and Integrate Railway Freight Transportation Service

Railway construction has a momentous consequence in the construction of China transportation. By the year-end of 2008, China's total mileage of railways opened to traffic is expected to 79,000 km. There was an adjustment of 《Mid-and-Long Term Railway Network Planning》in 2008, and the plan that has been adjusted accelerates the speed of railway construction, by 2020, China's total mileage of railways opened to traffic is expected to more than 120,000 km (MOR, 2008).

Railway should strengthen the resource integration, besides the construction of the infrastructure. The integration of railway freight transportation resource is to integrate railway logistics resource and the related resource, to achieve the aims of decrease the cost and improve the efficiency. The integration contains the aspects as follows: (1). integrate the organizational units. Railway logistics entity will be established by annexation and reorganization of enterprises of different logistics enterprises. The entity will provide integrative logistics service which takes the third-part logistics enterprises as the core, and combines transportation, forwarder, express, storing enterprises, and so on. (2).integrate the facilities and equipments. Railway department should integrate, expend the scale and close the facilities and equipment which are different in scale, level, and distributed in everywhere to construct logistics centers and bases. (3).integrate the process of work. Railway department should systematize and simplify the work of packaging, loading and unloading, transportation, storing, and so on, then realize logistics integration by removing the process of extra and inefficiency, shortening the channels. (4).integrate railway stations with the yards of road and air, the port for docking to an integrated and effective logistics center, meanwhile, in order to realize logistics rationalization in a bigger area, they can connect, cooperate with proper division of labor with logistics centers(distribution center) in cities.

5.4. Establish Express Freight Transportation System to Accelerate the Arrived Speed of Goods

Express will be the direction of railway freight transportation development, and the express of high value-added goods is the new growth point for railway transportation. The establishment of railway express freight transportation system will contribute to transport high value-added goods and gain bigger market share. With the development of economy in China and the input of railway, China railway should establish express transportation network with container-based and combined transport-based to intensify market competition, cement market share and expend new market. Railway express network can not only meet the commands of the customers, but also drive the reform of railway management and technique.

The service mode of traditional railway (station to station) must change into the fast and convenient service—door to door, railway logistics enterprise should adopt a proactive approach to cooperate with road, water and air. The establishment of the three dimensions freight transportation network and the standard of combined transport shall be accelerated, developing one principle and smooth interface backup for the service of high value-added goods, such as integrated logistics and the integration of supply chains, and so on.

5.5. Improve the Function of Logistics

Based on the analysis of railway scope of management and the situation of modern logistic in railway, railway should take transportation, forwarder, the expend service of transportation as the main business, then combine the advantage of railway transportation and the resource of regions, with rational planning and scientific positioning, to develop third-party logistics focused specialization of logistics. Railway departments should introduce advantage logistics technique and management to reform and improve logistics service. They should enter into the whole process of purchase, manufacture, sale, etc. to provide the consulting and planning of logistics plans, customer service, demand forecasting, storing controlling, goods reserving, purchasing, reduction, and so on. They should provide the connection of different mode of transportation based on the demand of the owner, to realize organic combination of transportation, storing, handling, distribution processing, distribution, information treatment, and so on, and provide integrated service, which organicly links the customers and the different process of logistics to realize seamless connection of market demands. Moreover, railway departments should strengthen the storing advantage of large and medium-sized stations opened to the public, encourage and introduce the enterprise to develop storing, distribution, etc. according to local situation. Based on the system of railway marketing, they can offer market analyzing, forecasting, and other value-added services to the customers.

6. Conclusions

The transform of traditional freight transportation into modern logistics is the objective needs of logistics development trend. Railway reform and development provides the impetus for railway developing modern logistics, and railway should increase the steps to change into modern logistics enterprises which take railway transportation as the main advantages, scale management, and system service, through integrating resource, developing railway trackless-stations, establishing railway logistics hubs, improving the functions of logistics, and so on, then develop it into the mainstay in China logistics market to maximize the benefit with good service.

7. Acknowledgements

We would like to thank Professor Kai Liu for his constructive suggestions. The work described in this paper was supported by The Foundation of Ministry of Railways (No. 2008X010-C), We would like to give our sincere thanks to them!

References

Nero, G. (1999). A note on the competitive advantage of large hub-and-spoke networks. Transportation Research Part E. 35(4): 225-239.

Li, H.Y.(2008) · Evaluation of Railway Transportation Productivity Layout Adjustment. Research Report. Beijing.

Du, H., Liu, K. (2007). Exploration on Supplying Fast Logistics Service By Railways. Railway Freight Transport . pp. 3-4.

Jiang, J.N., Liu, K., Ma, G.P. (2006). Relationship on Asset Utilization between Railway Diversified Logistics Enterprises and Railway Transport Enterprises. Railway Transport and Economy. pp. 66-68.

Liu, J.F. (2008). China's Railway Steadily Push Forward the Strategic Loading Site Construction. World Railway. pp. 33-35.

The State Department of Transportation of Ministry of Railways(2006). Adjust Overall Arrangement, Fully Release Railway Transportation Productivity. Chinese Railways. pp. 32-37.

Zapfel, G. and Wasner M. (2002), Planning and optimization of hub-and-spoke transportation networks of cooperative third-party logistics providers, International Journal of Production Economics. 78(2): 207-220.

Ministry of Railways's website: http://www.china-mor.gov.cn/tllwjs/tlwgh.html.

Airline emission charges: effects on airfares, service quality, and aircraft design

Jan K Brueckner[1] *and Anming Zhang*[2*]

[1] Department of Economics, University of California, Irvine, CA 92697, US.
E-mail: jkbrueck@uci.edu
[2] Sauder School of Business, University of British Columbia, Vancouver, BC, Canada, V6T 1Z2.
E-mail: anming.zhang@sauder.ubc.ca

Abstract

This paper explores the effect of airline emissions charges on airfares, airline service quality, aircraft design features, and network structure, using a detailed and realistic theoretical model of competing duopoly airlines. These impacts are derived by analyzing the effects of an increase in the effective price of fuel, which is the path by which emissions charges will alter airline choices. The results show that emission charges will raise fares, reduce flight frequency, increase load factors, and raise aircraft fuel efficiency, while having no effect on aircraft size. Given that these adjustments occur in response to the treatment of an emissions externality that is currently unaddressed, they represent efficient changes that move society closer to a social optimum.

1. Introduction

As concerns about global warming mount, policy makers have begun targeting emissions from aircraft as a means of reducing the production of greenhouse gases. Most notably, the European Union starting in 2012 will require airlines to hold emission permits in order to operate. Each airline must hold a number of permits commensurate with the CO_2 pollution generated by its fleet, with permits acquired through a trading process following an initial, partially free distribution among the carriers.[1] This approach follows existing "cap-and-trade" schemes applied to polluters in the industrial and energy sectors both in Europe and in the US (see Forsyth, 2008, for details and further discussion).

The planned emission trading system generates a permit price that becomes part of an airline's cost structure. With the carrier's required outlay on emissions permits varying in step with its total fuel consumption, the permit price is effectively added to the price of fuel, even though most of the permits will be freely distributed. Thus, the planned trading system can be viewed as equivalent to a carbon-tax scheme applied to aviation, which would explicitly raise the price of fuel. As a result, regardless of whether policy interventions to limit aviation emissions follow the EU's cap-and-trade approach or rely on taxation, they can all be depicted as policies that raise the fuel price paid by airlines. The effects of such policies thus operate in the same direction as the impacts of a secular fuel-price increase unrelated to government intervention, as occurred through mid-2008.

How will airlines respond to a policy-induced increase in the effective price of fuel? What will happen to airfares? How will service quality, as reflected in flight frequencies and load factors, change? How will aircraft design (fuel efficiency, seat capacity) evolve in response to changes in the derived demands of airlines as fuel prices rise? Will the structure of airline networks change? The purpose of the present paper is to answer these questions. The answers, which are derived from a simple theoretical model, are important since they predict the detailed response of a key, highly visible industry to a major government intervention in the fight against global warming.

---

[*] We thank Michael Levine for stimulating comments. Any errors or shortcomings in the paper, however, are our responsibility.

[1] As of late 2008, the EU's intention was to freely distribute 85 percent of the total emission permits, with the remaining 15 percent auctioned. Movement to a 100 percent auction system is envisioned in later years. The permit distribution would cap emissions at 97 percent of the 2004-2006 level (see Wall, 2008).

The paper builds on the approach used by Brueckner and Girvin (2008) in their analysis of the impact of airport noise regulation. In the model, passengers value airline flight frequency and dislike high load factors because of discomfort and the higher chance of being denied boarding (bumped from a flight).[2] Passengers choose between two competing airlines based on these factors along with the fares they charge. Each airline incurs fuel cost, which depends on aircraft fuel efficiency, as well as the capital cost of the aircraft in its fleet. This latter cost depends on aircraft size (seating capacity) as well as fuel efficiency. With higher efficiency requiring better engine technology as well as use of lighter materials, aircraft capital cost rises. Even though both design features (size and efficiency) are set by the manufacturer, the airline is portrayed as the ultimate decision-maker, with the manufacturer responding to its derived demand for aircraft characteristics.[3]

A key issue in the analysis is the form of the function relating aircraft capital cost to size and fuel efficiency, measured as fuel consumption per seat. While cost per seat is realistically assumed to fall with aircraft size holding fuel efficiency constant, the interaction between fuel efficiency and size in determining capital cost is less obvious a priori. The analysis imposes a plausible general restriction on this interaction, which is satisfied by the specific functional form that is adopted to facilitate the analysis. Although engineering data could, in principle, shed light on the appropriateness of the adopted functional form, available data appear not to be rich enough for this undertaking.[4]

The analysis in section 2 of the paper develops the model, deriving the airline profit function. The key parameter in this function is the price per unit of fuel. Section 3 analyzes the airline profit-maximization problem and carries out a comparative-static analysis. The model solution is given by a nonlinear simultaneous equation system, but despite the system's complexity, comparative-static analysis is feasible and yields determinate fuel-price impacts. Some of the comparative-static results are natural (for example, the fare rises with the fuel price), but some findings are not predictable a priori. As a group, however, the comparative-static results provide important information by showing how airline emission charges (either in the form of a price for emission permits or carbon tax) are likely to affect airline decisions. The derived fare and service-quality impacts also yield an expected negative impact on passengers, whose utility from direct consumption falls. This loss is ameliorated, however, by environmental improvements from reduced emissions, as seen in the brief efficiency analysis presented in section 4.

Section 5 introduces network considerations. Recognizing that an airline can operate either a hub-and-spoke (HS) network or a fully-connected, point-to-point (FC) network, the analysis investigates two questions. The first concerns the choice of fuel efficiency conditional on network type, asking how the chosen level of efficiency differs between HS and FC networks. The second question is whether an increase in the fuel price systematically favors one network type over the other. The

[2] While monetary compensation partly offsets the cost of denied boarding, some cost presumably remains for the typical bumped passenger.

[3] To a certain extent, airlines can also affect average aircraft characteristics through their decisions on retirement of the older, less fuel-efficient planes in their fleets, and through fleet deployment/operations decisions with respect to aircraft size.

[4] Data from the Boeing and Airbus websites (www.boeing.com and www.airbus.com) do indicate fuel-efficiency, measured as gallons consumed per seat mile, for various aircraft types. The numbers are 0.0171 for a 1988-vintage Airbus A320, 0.0139 for a 1998-vintage Boeing 737-800, 0.0166 for a 1998-vintage Boeing 777-300, 0.0180 for a 1995-vintage Airbus A340-600, and 0.0167 for a 2008-vintage Boeing 747-8. The values for the large aircraft (the last three) are similar, while the 737-800 value, which is from a smaller aircraft of a comparable vintage, is lower. Discounting the A320 value (which is similar to that for the larger aircraft) as reflecting the earlier 1988 vintage, the numbers suggest that the chosen level of fuel-efficiency is higher for smaller aircraft. While these numbers are revealing, exploring the relationship between capital cost and aircraft size and fuel efficiency would require additional data on aircraft selling prices. Since sales usually occur below list prices at values that are not publicly revealed, acquisition of price data is problematic. But even if such data were available, the small number of different existing aircraft types would probably not yield enough data points to estimate, using regression analysis, the desired relationship (this exercise, were it possible, would involve regressing price on fuel efficiency and size).

answer indicates whether imposition of airline emission charges will tend to alter current network structures, perhaps causing a shift away from the dominant HS structure. Section 6 offers some brief empirical evidence in support of the model's predictions, and section 7 presents conclusions.

2. The Model

Consider a travel market connecting two cities, which is served by two competing airlines. A main component of the cost of serving passengers is aircraft operating costs, which consist of fuel cost and the leasing (or ownership) cost of aircraft. The fuel cost depends on aircraft fuel efficiency, which is measured as fuel consumption per seat per flight hour, denoted by the variable e. The aircraft leasing cost per hour flown, denoted $g(e,s)$, depends on aircraft seating capacity s, as well as on fuel efficiency. Since a more fuel-efficient plane (with a lower e) is more expensive (requiring better engine technology, better design, and/or use of lighter materials), while a larger plane also costs more, g is decreasing in e and increasing in s. Economies of larger aircraft mean that g/s is decreasing in s, and since successive increases in fuel efficiency (decreases in e) are increasingly expensive, g is convex in e. Additionally, a negative cross partial derivative might make sense, implying that increases in fuel efficiency are harder to achieve in a larger plane. While the g function is most easily viewed as a pure capital cost, it could also include crew costs per hour, which depend on the size of the aircraft.

Letting r denote the price per unit of fuel, the flight cost per hour is given by $res + g(e,s)$, where the first term gives fuel cost. Letting k be hours per flight, the cost of one flight is then equal to $[res + g(e,s)]k$.[5] With f flights operated and k being an increasing function of distance d, the airline's total cost is thus given by

$$f[res + g(e,s)]k(d). \qquad (1)$$

Expression (1) indicates that the airline's cost depends on trip length and the number of seats per aircraft, two variables found by Swan and Adler (2006) to be the main factors affecting aircraft costs. Swan and Adler (2006) further argue $k(d)$ is a linear function with a positive intercept, implying economies of "stage length" (flight distance): the cost per kilometer flown declines as the number of kilometers flown increases.

The two airlines serving the market, referred to as firms 1 and 2, carry passenger volumes of q_1 and q_2, respectively.[6] Each consumer makes one trip on either airline 1 or airline 2, so that the total number of passengers in the market is fixed and normalized to 1 ($q_1 + q_2 = 1$). While consumers could in principle choose not to travel, making the passenger volume elastic, the model assumes that travel demand is strong enough to allow the no-travel option to be ignored. The quantities q_1 and q_2 depend on the fares charged by the airlines, denoted p_1 and p_2, and on the service qualities they provide.

---

[5] It should be noted that this cost formulation requires fuel consumption per flight hour to be independent of the duration of a flight. This independence is plausible given that high fuel consumption during the take-off phase of the flight is offset by low consumption during the landing phase, so that average fuel consumption per hour during these two phases may approximately equal to consumption per hour during the longer cruise phase of the flight. In this case, fuel consumption per hour may be roughly invariant to total flight duration, even though the landing and takeoff phases account for a smaller share of total flight hours on longer flights.

[6] Although the analysis focuses on a duopoly market, it extends immediately to the *n*-firm case. With an *n*-firm oligopoly, the ½ term in traffic *q* (see below) is replaced by 1/*n*. This change can be demonstrated by applying the approach of Brueckner (2008) to the current model. For a noncompetitive alternative in developing this type of model, see Girvin (2008).

One element of service quality is flight frequency, which determines the "frequency delay" experienced by a passenger (the difference between the passenger's preferred departure time and nearest flight time). When using a particular airline, a passenger's expected frequency delay is inversely proportional to the airline's flight frequency f, assuming the passenger's preferred departure times are random and uniformly distributed and that flights are equally spaced.[7] The passenger's cost of frequency delay on airline i can then be written γ/f_i, where γ is a cost parameter common to all passengers.

Douglas and Miller (1974) argue that another type of delay, denoted "stochastic delay," arises through excess demand, which may lead to denial of boarding on a flight that is oversold and thus an additional delay. Stochastic delay is affected by airline's load factor, which equals the percentage of its seats filled by passengers. Following Panzar (1979), the analysis assumes that a passenger's probability of being denied a seat, and hence stochastic delay, is proportional to an airline's load factor, denoted l_i for airline i. The cost of stochastic delay can then be written λl_i, where λ is a common cost parameter. Gathering all these elements, the cost of flying on airline 1 is then the sum of the fare and the two delay costs, equal to $p_1 + \gamma / f_1 + \lambda l_1$, with an analogous expression for airline 2.

An additional motivation for the appearance of the load factor in the passenger's cost expression is possible. While generating stochastic delay, a high load factor also imposes a cost on the passenger arising from aircraft crowdedness and the resulting discomfort. This interpretation of the load factor's impact will be useful in the welfare analysis presented below.

Although passengers compare the costs of flying in choosing between airlines 1 and 2, brand loyalty also affects the choice, as in Brueckner and Girvin (2008). Brand loyalty appears as an extra additive term in the cost expression for airline 1, which is negative for passengers preferring airline 1 and positive for passengers preferring airline 2. Assuming that this brand loyalty term is uniformly distributed over the interval $[-\alpha, \alpha]$, the number of passengers preferring airline 1 can be shown to equal

$$q_1 = \frac{1}{2} - \frac{1}{\alpha}\left(p_1 - p_2 + \frac{\gamma}{f_1} + \lambda l_1 - \frac{\gamma}{f_2} - \lambda l_2 \right), \qquad (2)$$

while the demand for airline 2 is given by the analogous expression with the 1 and 2 subscripts interchanged (see Brueckner and Girvin (2008)). Thus an increase in flight frequency by airline 1 will increase its demand while reducing airline 2's demand, and an increase in the load factor or the fare will have the opposite effect. Note that, when all three variables are equal across the two airlines, each airline faces a demand of ½, exactly splitting the total number of passengers with its competitor.

Analysis of this model using a general form for the capital-cost function $g(e,s)$ is inconclusive. To generate determinate results, the capital-cost function is assumed to take the following specific form:

$$g(e,s) = \frac{\beta + \varepsilon s}{e}, \qquad (3)$$

where β and ε are positive parameters. Note that, with fuel efficiency held constant, dividing (3) by s shows that cost per passenger is decreasing in aircraft size, as suggested above. In addition, cost increases as fuel efficiency rises (as e falls), and does so at an increasing rate.

Using (3), (2) and (1), the profit of airline 1 can be written as

[7] Assuming flights are equally spaced over the day, represented by a circle of circumference T, and that a passenger's preferred departure time is random and distributed uniformly across the circle, expected frequency delay equals $T/4f$.

$$\pi_1 = p_1 q_1 - f_1\left(re_1 s_1 + \frac{\beta + \varepsilon s_1}{e_1} \right) k(d)$$
$$= \left(p_1 - \frac{re_1 + (\varepsilon / e_1)}{l_1} k(d) \right) q_1 - f_1 \frac{\beta}{e_1} k(d)$$
$$= \left(p_1 - \frac{re_1 + (\varepsilon / e_1)}{l_1} k(d) \right) \left[\frac{1}{2} - \frac{1}{\alpha} \left(p_1 - p_2 + \frac{\gamma}{f_1} + \lambda l_1 - \frac{\gamma}{f_2} - \lambda l_2 \right) \right] - f_1 \frac{\beta}{e_1} k(d) . \quad (4)$$

Note that the equality $f_1 s_1 l_1 = q_1$, which equates total occupied seats to passengers, is used to eliminate s_1 in the second line of (4). Given this equality, only two of the variables f_1, s_1 and l_1 can be choice variables for the airline. As seen in (4), f_1 and l_1 from among this group are treated as the choice variables, leaving s_1 to be deduced from the relationship $s_1 = q_1 / f_1 l_1$. Note also from (4) that the carrier's total cost has two parts: one is proportional to the number of passengers carried q_1, while the other part, namely $(f_1 \beta / e_1) k$, does not vary with the number of passengers.[8]

With the model specification now clear, it is useful to explain how the imposition of airline emission charges can be represented by an increase in the fuel price. As indicated in the introduction, the impact of emissions charges is viewed as increasing the effective price of fuel regardless of whether an EU-style cap-and-trade system or a carbon tax is used (a similar point is made by Forsyth, 2008). To understand this claim, let x denote the number of permits required per unit of fuel consumed, so that the number of permits needed by an airline is given by $xfsek$. Letting the emission permit price be denoted z and the number of permits allocated to the airline be denoted m, the term $z(xfsek - m)$ would then be subtracted from profit. This expression gives the outlay for the purchase of permits when the airline requires more than its allocation (when $xfsek > m$), or the revenue from the sale of permits when the airline has more than it can use (when $xfsek < m$). As can be seen by inspecting the first profit expression in (4), the fuel price r would then be replaced by $r + zx$, with the constant zm also subtracted from the profit expression. Similarly, with a carbon tax t per unit of fuel, the fuel price would be replaced by $r + t$. In either case, therefore, the impact of airline emission charges may be assessed simply by analyzing the impact of an increase in the fuel price.

A final point regarding the model setup concerns the assumption of a fixed total passenger volume ($q_1 + q_2 = 1$). Since the imposition of airline emission charges would be expected to reduce the total volume of air travel by increasing its cost relative to other goods, reliance on a model where this volume is fixed is not ideal. However, when the current approach is imbedded in a model with an elastic travel demand, the complexity of the resulting framework makes it amenable only to numerical analysis. Such analysis, which could adopt the elastic-demand setup of Brueckner (2008), might be an undertaking for future research.[9] But the first priority is to explore a model capable of generating analytical results, and the present paper reflects this priority despite the limitation of a fixed passenger volume.[10]

[8] It should be noted that the above formulation assumes that aircraft size and frequency can be smoothly adjusted to suit the size of the market. In reality, such decisions involve indivisibilities such as minimum aircraft sizes and minimum viable flight frequencies, which may constrain actual choices.

[9] Brueckner and Flores-Fillol (2007) analyze a related model where the passenger volume is elastic, but Brueckner's (2008) framework is more tractable.

[10] This limitation was also present in the Brueckner and Girvin's (2008) analysis of noise mitigation policies, which raise the fare but leave total passenger volume unchanged in their model.

3. Profit Maximization and Comparative Statics

This section analyzes the profit-maximization problem and carries out comparative-static analysis. Consider the decisions faced by airline 1, which chooses p_1, e_1, f_1 and l_1 to maximize profit. Using the profit function (4), airline 1's first-order conditions can be written (suppressing, hereinafter, the *d* argument of *k*(*d*)):

$$\frac{\partial \pi_1}{\partial p_1} = q_1 + \left(p_1 - \frac{re_1 + (\varepsilon / e_1)}{l_1} k \right)\left(-\frac{1}{\alpha} \right) = 0 \tag{5}$$

$$\frac{\partial \pi_1}{\partial e_1} = -\frac{r - (\varepsilon / e_1^2)}{l_1} kq_1 + f_1 \frac{\beta k}{e_1^2} = 0 \tag{6}$$

$$\frac{\partial \pi_1}{\partial f_1} = \left(p_1 - \frac{re_1 + (\varepsilon / e_1)}{l_1} k \right)\frac{\gamma}{\alpha f_1^2} - \frac{\beta k}{e_1} = 0 \tag{7}$$

$$\frac{\partial \pi_1}{\partial l_1} = q_1 \frac{re_1 + (\varepsilon / e_1)}{l_1^2} k + \left(p_1 - \frac{re_1 + (\varepsilon / e_1)}{l_1} k \right)\left(-\frac{\lambda}{\alpha} \right) = 0. \tag{8}$$

The first-order conditions for airline 2 are symmetric, and the second-order condition (positive definiteness of the Hessian matrix of π_1) is assumed to hold.

The conditions (5)-(8) have standard interpretations. For example, (5) gives marginal revenue from an increase in p_1, with the bracketed terms capturing the effect of the induced decline in q_1 (note that $-1/\alpha$ is q_1's derivative with respect to p_1). Similarly, the first part of (7) equals marginal revenue from an increase in frequency, with $\gamma / \alpha f_1^2$ giving the impact of the higher f_1 on q_1, while the last term gives the effect on cost. The remaining equations have similar interpretations.

Given the symmetry of the model, the equilibrium values of p_i, e_i, f_i and l_i, i = 1,2, will be symmetric across carriers, and each airline's equilibrium traffic will equal ½. Imposing symmetry in (5)-(8), substituting and rearranging, the following equations are obtained:

$$p = \frac{\alpha}{2} + \frac{re + (\varepsilon / e)}{l} k \tag{9}$$

$$e^2 = \frac{\varepsilon + 2\beta fl}{r} \tag{10}$$

$$f^2 = \frac{\gamma e}{2\beta k} \tag{11}$$

$$l^2 = \frac{re^2 + \varepsilon}{\lambda e} k. \tag{12}$$

The equilibrium, denoted (p^*, e^*, f^*, l^*), is the solution to equations (9)-(12).

To start the comparative-static analysis, let (10) be rewritten as $re^2 - \varepsilon = 2\beta fl$. Squaring both sides, substituting (11) for the resulting f^2 and (12) for l^2, and rearranging, the equation can be rewritten as

$$r^2 e^4 - r \cdot (2\varepsilon + \mathrm{K}) e^2 + \varepsilon \cdot (\varepsilon - \mathrm{K}) = 0; \qquad \mathrm{K} \equiv 2\gamma\beta / \lambda, \tag{13}$$

where K is a positive parameter. Applying the quadratic formula (viewing e^2 as the unknown), the solution for e is given by

$$e^* = \left[\left(r(2\varepsilon + \mathrm{K}) \pm \sqrt{r^2(2\varepsilon + \mathrm{K})^2 - 4r^2\varepsilon(\varepsilon - \mathrm{K})}\right)/ 2r^2\right]^{1/2} = \left[\left(2\varepsilon + \mathrm{K} \pm \sqrt{(8\varepsilon + \mathrm{K})\mathrm{K}}\right)/ 2r\right]^{1/2}. \quad (14)$$

Equation (14) yields two positive solutions when $\varepsilon > \mathrm{K}$, but one positive and one complex solution when $\varepsilon < \mathrm{K}$. In the latter case, the square root expression in the first part of (14) is larger than the expression prior to the $\pm$ sign, making the entire expression (which is raised to the ½ power) negative. To rule out the first multiple-solution case, $\varepsilon < \mathrm{K}$ is assumed to hold, so that the unique (real) solution is given by the last expression in (14) with a plus sign prior to the square root. Interpretation of this condition is provided in section 5 below.

Equation (14) determines equilibrium e^* as a function of the fuel price r and other parameters ($\varepsilon, \gamma, \beta, \lambda$), written as $e^*(r)$ after suppressing these latter arguments. Differentiating (14) to find the effect of r yields

$$\frac{de^*}{dr} = -\frac{e^*}{2r} < 0. \quad (15)$$

Equation (15) indicates that an increase in the fuel price leads to a lower e^* or, equivalently, more fuel-efficient planes.

The comparative-static effects of the fuel price on the equilibrium values of the remaining choice variables (p^*, f^*, l^*) are examined next. Once $e^*(r)$ is determined, flight frequency, by (11) equal to $f^* = \sqrt{\gamma e^* / 2\beta k},$ is simply a function of parameters. Since f^* increases in e^*, which by (15) decreases in r, it follows that

$$\frac{df^*}{dr} < 0, \quad (16)$$

so that an increase in the fuel price leads to a lower flight frequency.

For the effect on the load factor, rearranging (10) yields $f^*l^* = [r(e^*)^2 - \varepsilon]/2\beta$. Differentiating this equation yields

$$\frac{d(f^*l^*)}{dr} = \frac{1}{2\beta}\left(2r\frac{de^*}{dr} + e^*\right) = 0, \quad (17)$$

where the second equality follows from (15). Expanding the derivative on the left-hand side of (16) and rearranging then yields

$$\frac{dl^*}{dr} = -\frac{l^*}{f^*}\frac{df^*}{dr} > 0, \quad (18)$$

using (16). That is, an increase in the fuel price leads to a higher load factor. Furthermore, since $s^* = 1/2f^*l^*$, (17) implies

$$\frac{ds^*}{dr} = 0, \quad (19)$$

so that a change in the fuel price does not affect aircraft size.

The fare in (9) depends on the fuel price both directly and indirectly via fuel efficiency and the load factor. Capturing these channels, the impact of a higher fuel price can be decomposed into four parts: (i) a positive direct effect (captured by the $e^*k/l^* > 0$ expression multiplying r^*); (ii) a negative indirect effect via the use of more fuel-efficient planes, which which leads to fuel savings (captured by rk/l^* times de^*/dr); (iii) a positive indirect effect via the use of more fuel-efficient planes, which results in higher aircraft leasing cost (this effect operates through e^* in the denominator of the upper ratio term in (9)); and (iv) a negative indirect effect via a higher load factor, which leads to downward pressure on per-passenger cost (this effect operates through l^* in the large ratio expression in (9)). Using (12) in (9), p^* can be rewritten as $p^* = \alpha/2 + \lambda l^*$, which allows the net effect of these four impacts to be captured simply through the effect of the higher fuel price on l^*. Thus, using (18),

$$\frac{dp^*}{dr} > 0, \quad (20)$$

so that the two positive effects above dominate the negative effects, leading to a higher fare in response to an increase in the fuel price. Summarizing the above results yields

Proposition 1. *An increase in the fuel price, or an equivalent imposition of airline emissions charges, will lead to a higher fare, lower flight frequency, a higher load factor, more fuel-efficient aircraft, and an unchanged aircraft size.*

Given that both elements of airline service quality worsen, while the fare increases, passengers are unambiguously worse off following the imposition of emission charges. In addition, application of the envelope theorem to the profit expression in (4) shows that airline profit falls in response to the imposition of charges (to the rise in r). Environmental benefits from reduced emissions have not yet been considered, however, and this element will be added in the next section.

A final comparative-static question concerns the effect of flight distance on the various choice variables, and the answers are immediate. From (14), since e^* is independent of k, fuel efficiency is unaffected by flight distance. Equations (9), (11), and (12) then show that p^*, f^*, and l^* are respectively increasing, decreasing and increasing in k, so that the fare and load factor rise, while frequency falls, when flight distance increases. Finally, from (10), $f^*l^* = [r(e^*)^2 - \varepsilon]/2\beta$, and so f^*l^* is unaffected by flight distance. Consequently, aircraft size $s^*(=1/2f^*l^*)$ is unaffected by flight distance.

4. Welfare Analysis

To carry out a welfare analysis, the damage from emissions must be considered along with the interests of passengers and airlines. The treatment of passenger interests in the welfare analysis can be greatly simplified if the costs associated with a higher load factor are attributed entirely to aircraft crowding and discomfort rather than to stochastic delay. The reason is that incorporation of the stochastic-delay element would require a more sophisticated analysis involving random travel decisions. Under this restriction, total consumer utility can be represented simply by consumption expenditure, equal to income minus travel costs, which are in turn given by the airfare plus the costs of schedule delay and crowding. Brand loyalties also affect utility, but they aggregate to a constant.

To compute social welfare W, consumer utility is added to airline profit, with emissions damage then subtracted. Given symmetry and the unit mass of passengers, welfare can then be written as

$$W = 2\pi + y - p - \gamma / f - \lambda l - \mu(2fesk), \qquad (21)$$

where y is passenger income and the last term represents emissions damage. Note that $2fesk$ represents total fuel usage by the two airlines, while μ is a parameter that gives emissions damage per unit of fuel burned (a linear relationship is assumed for simplicity). In the analysis of (21), the fare cancels since it represents a transfer between passengers and the airlines.

Using a related model without any environmental components, Brueckner and Flores-Fillol (2007) show that equilibrium is efficient in a situation when all potential passengers travel, as in the current analysis. Adapting this result to the present context, the implication is that if μ were equal to zero (eliminating any environmental concerns), then the equilibrium values of p, f, e, and l would maximize the welfare expression in (21). The efficiency result in Brueckner and Flores-Fillol's model disappears, however, with the introduction of a travel/no-travel margin, which makes the total quantity of passengers dependent on airline choices rather than fixed. Efficiency would similarly disappear in the present model if it were modified to allow an elastic travel demand, although (as explained above) this modification makes much of the analysis intractable.

While equilibrium in the current fixed-passenger-volume framework is efficient when when $\mu = 0$, the equilibrium is inefficient when $\mu > 0$. Emissions damage then arises, which airlines ignore in their decisions, and achievement of efficiency requires charging for this damage. Given the linearity of the damage function, this charge should be set equal to μ per unit of fuel. Thus, μ is the proper value for the carbon tax t. In addition, it is easily shown that the endogenous emissions-permit cost per unit of fuel emerging from the trading process (zx from above) will also equal the damage parameter μ, provided that the total number of permits distributed to the airlines is equal to the socially optimal emissions volume from (21).

The current situation, where airline emissions charges are absent, thus involves an inefficient equilibrium. Since the effective fuel price is too low, the preceding analysis indicates the directions of the divergence from efficiency. In particular, aircraft fuel efficiency is currently too low, flight frequency is too high, and the load factor is too low, although aircraft size is efficient. Emissions charges, by increasing the effective fuel price, would correct these inefficiencies by raising aircraft fuel efficiency, reducing flight frequency, and raising the load factor.

5. Network Considerations

While the previous analysis derived the impact of higher fuel prices by focusing on a single route, actual airline service is provided in a network context. This section of the paper imbeds the model in a network setting, where carriers can either operate a hub-and-spoke (HS) network or a fully connected, point-to-point (FC) network. The analysis focuses on two questions. First, how does chosen level of fuel efficiency differ between the two network structures? In other words, does an airline choose more fuel-efficient aircraft when its passengers connect at a hub airport or when it provides point-to-point service? Second, how is the profit-maximizing network structure (HS vs. FC) affected by imposition of emission charges (by an increase in r)? Industry observers, policy makers and researchers have all speculated about the likely network impacts of emissions charges, making the latter inquiry useful.[11]

5.1. The setup and the effect of network structure on fuel efficiency

The analysis is conducted using a three-node symmetric city layout, with all the links equidistant and

[11] For example, Albers, et al. (2009) argue that airlines may add an intermediate stop outside the EU on long international flights, so that the EU emissions charge only applies to the final short leg.

the three city-pair markets assumed to have the same demand.[12] The three-city system is serviced by the two competing airlines, which both use a FC network to serve the cities or both use a HS network (asymmetric network choices are ruled out). Under the FC network, passengers in the three city-pair markets are carried by direct (nonstop) flights on three routes. For a HS network, with one city serving as the hub, there are just two "spoke" routes, which connect the two non-hub cities to the hub. While spoke passengers continue to take direct flights, passengers traveling between the two non-hub cities must take two flights and connect at the hub. As a result, on a given spoke route, both local traffic and connecting traffic is carried. While this higher traffic volume allows airlines to capture economies of aircraft size, thereby reducing costs, connecting passengers fly a longer distance than under the FC network, which tends to raise costs.[13] For simplicity and without loss of generality, connecting passengers are assumed to pay the same fare as nonstop passengers.[14]

Given this setup, FC profit is just three times the single-route profit expression from (4). The FC equilibrium, denoted $(p_{FC}, e_{FC}, f_{FC}, l_{FC})$, is then still the solution to equations (6)-(9). In particular, e_{FC} is determined by equation (14), and f_{FC} is then determined by (11), l_{FC} by (12), p_{FC} by (9), and $s_{FC} = 1/2 f_{FC} l_{FC}$.

In the HS case, however, revenue is $3pq$, but costs are two times (1), or $2f[res + g(e,s)]k$, given that there are only two routes under the HS network. Furthermore, while the previous equality $f_1 s_1 l_1 = q_1$ applied in the FC case for airline 1, $f_1 s_1 l_1 = 2q_1$ holds in the HS case given that each spoke route carries passengers in two markets. As a result, HS profit for airline 1 is

$$\pi_1^{HS} = 3\left\{\left[p_1 - \frac{4}{3}\frac{re_1 + (\varepsilon / e_1)}{l_1}k\right]q_1 - \frac{2}{3}f_1\frac{\beta k}{e_1}\right\}. \tag{22}$$

That is, the HS profit equals three times a modified form of profit expression (4), where the ratio term involving l_1 is multiplied by 4/3 and the last term is multiplied by 2/3.

In the HS case, airline 1 maximizes profit (22) with respect to p_1, e_1, f_1 and l_1, with aircraft size s_1 again the residual variable. Imposing symmetry in the resulting first-order conditions and rearranging, the HS equilibrium, denoted $(p_{HS}, e_{HS}, f_{HS}, l_{HS})$, is the solution to the following equations:

$$p_{HS} = \frac{\alpha}{2} + \frac{4}{3}\frac{re_{HS} + (\varepsilon / e_{HS})}{l_{HS}}k \tag{23}$$

---

[12] This three-city framework has been used by, among others, Brueckner (2004), Oum, et al. (1995), Zhang (1996) and Pels, et al. (2000).

[13] The ability to exploit the economies from larger aircraft when passenger volumes increase is an important force behind economies of traffic density. But, by allowing an increase in service quality, concentration of traffic in an HS network also affects demands in individual markets served by the network. Specifically, higher traffic allows an airline to increase flight frequency, and the improved convenience raises demand. On the other hand, higher traffic may allow the airline to raise its load factor, lowering per-passenger costs while at the same time reducing service quality, which tends to reverse the frequency-related demand increase. All of these forces are accounted for in the ensuing analysis. It should be noted that, while Oum, et al. (1995) explore the effects of these network complementarities on airlines' competitive strategies in network choice, the present paper abstracts from network rivalry considerations. By assuming that the two airlines either both use a FC network or both use a HS network, the focus is instead on the effect of emissions charges on network choice at the industry level.

[14] Without this restriction, the airlines will choose different fares for connecting and non-stop passengers. However, the equilibrium solutions for the remaining choice variables are nevertheless the same as when fares are constrained to be equal. Therefore, for expositional simplicity, the equal-fare assumption is imposed.

$$e_{HS}^2 = \frac{\varepsilon + \beta f_{HS} l_{HS}}{r} \quad (24)$$

$$f_{HS}^2 = \frac{3}{4}\frac{\gamma e_{HS}}{\beta k} \quad (25)$$

$$l_{HS}^2 = \frac{4}{3}\frac{re_{HS}^2 + \varepsilon}{\lambda e_{HS}} k. \quad (26)$$

These equations differ from (9)-(12) by the appearance of the new multiplicative factors from (22) in the appropriate positions. Tracing through the derivations, the new equation determining e_{HS} is

$$r^2 e^4 - r \cdot [2\varepsilon + (\mathrm{K}/2)]e^2 + \varepsilon \cdot [\varepsilon - (\mathrm{K}/2)] = 0. \quad (27)$$

Note that (27) is the same as (13), which determines e_{FC}, except that the K term in (14) is now replaced by $\mathrm{K}/2$. To again ensure a unique solution to (27), $\varepsilon < \mathrm{K}/2$ is assumed to hold, leading to the solution[15]

$$e_{HS} = \left[\left(2\varepsilon + (\mathrm{K}/2) + \sqrt{[8\varepsilon + (\mathrm{K}/2)](\mathrm{K}/2)}\right)/2r\right]^{1/2}. \quad (28)$$

Given e_{HS}, f_{HS} is then determined by (25), l_{HS} by (26), p_{HS} by (23) and $s_{HS} = 1/f_{HS} l_{HS}$.

From (28) and (14), it is easily seen that $e_{HS} < e_{FC}$, so that more fuel-efficient planes are used under the HS network than under the FC network. The apparent intuition is that, since connecting passengers under the HS network take two flights rather than one (thus consuming more fuel, holding e constant), the airline has a greater incentive to economize through higher aircraft fuel efficiency. With K rising by a factor of 2 going from the HS to FC networks, it is easily seen that the e solution rises by less than a factor of $\sqrt{2}$, so that $e_{FC}/e_{HS} < \sqrt{2}$.[16] Thus,

$$e_{HS} < e_{FC} < \sqrt{2} e_{HS}. \quad (29)$$

Flight frequencies can also be compared between network types. Combining (11), (25) and (29) yields

$$\frac{f_{FC}^2}{f_{HS}^2} = \frac{2}{3}\frac{e_{FC}}{e_{HS}} < \frac{2}{3}\sqrt{2} < 1. \quad (30)$$

Thus $f_{FC} < f_{HS}$, so that flight frequency is higher under the HS network than under the FC network, a result also derived by Brueckner (2004) in a related monopoly model without fuel-efficiency and load-factor choices.

To compare aircraft sizes, note that (9) and (24) yield

$$f_{FC} l_{FC} = (re_{FC}^2 - \varepsilon)/2\beta, \qquad f_{HS} l_{HS} = (re_{HS}^2 - \varepsilon)/\beta.$$

Dividing the two equations and using the equalities $f_{FC} l_{FC} = 1/2s_{FC}$ and $f_{HS} l_{HS} = 1/s_{HS}$ yields

[15] Note that this condition implies the previous condition $\varepsilon < K$.

[16] Note that $e_{FC}/e_{HS} = \sqrt{2}$ if and only if $\varepsilon = 0$.

$$\frac{s_{HS}}{s_{FC}} = \frac{re_{FC}^2 - \varepsilon}{re_{HS}^2 - \varepsilon} > 1, \qquad (31)$$

where the inequality follows from (29). Thus $s_{HS} > s_{FC}$ holds, so that larger aircraft are employed under the HS network than under the FC network. This is a sensible result, also derived by Brueckner (2004): by concentrating passengers on fewer routes, a HS network allows better exploitation of the economies from larger aircraft. Finally, the load factor comparison between the two network types is ambiguous, an issue that is explored further in the appendix. Summarizing yields

Proposition 2. *Under the assumed network structure, the HS network has more fuel-efficient aircraft than the FC network. In addition, aircraft are larger and flight frequency is higher in the HS network than in the FC network.*

5.2. The effect of emission charges on network structure

The discussion now turns to the impact of emissions charges on airline network structure, recognizing that an airline will choose the network type (HS or FC) that yields the higher profit level. Although the analysis cannot explicitly identify which network type is optimal, it is possible to investigate how emissions charges (an increase in *r*) affect the *relative* profitability of HS and FC networks, which reveals the direction of the incentive to switch configurations as *r* rises.[17]

The HS-FC profit differential, $\pi_1^{HS} - \pi_1^{FC}$, is given by the difference between (22) and (4). Applying the envelope theorem, the derivative of $\pi_1^{HS} - \pi_1^{FC}$ with respect to r, evaluated at the symmetric equilibrium, is given by

$$\left(-4\frac{e_{HS}}{l_{HS}} + 3\frac{e_{FC}}{l_{FC}}\right)\frac{k}{2}. \qquad (32)$$

An increase in *r* will favor a HS (FC) network, raising (lowering) the HS–FC profit differential, when (32) is positive (negative). However, the sign of this expression is ambiguous. On the one hand, more fuel efficient planes are used under the HS network ($e_{HS} < e_{FC}$), which tends to make the expression positive. On the other hand, greater trip circuity under the HS network raises costs relative to the FC network, as reflected in the difference between the 4 factor in the first term of (32) and the 3 factor in the second term. This difference tends to make (32) negative. In addition, as shown in the appendix, the load factors l_{HS} and l_{FC} in general take different values.[18]

To resolve this ambiguity, let $F_{HS} \equiv 2ke_{HS}/l_{HS}$ and $F_{FC} \equiv 3ke_{FC}/2l_{FC}$, and note that (32) is proportional to $F_{FC} - F_{HS}$. However, it is easier to work with the ratio F_{HS}^2/F_{FC}^2, which is less (greater) than 1 as $F_{FC} - F_{HS}$ is positive (negative). From (a1) in the appendix, this ratio equals

---

[17] Although the impact of a higher fuel price on the HS–FC profit differential is the focus of the present analysis, a straightforward extension shows that the same results apply to the HS–FC welfare differential.

[18] Even if the two load factors were the same, the sign of (32) would be ambiguous. For instance, using $e_{FC} < \sqrt{2}e_{HS}$ from (29), the sign of (32) is the same as the sign of

$$-2e_{HS} + (3/2)e_{FC} = e_{HS}[-2 + (3/2)e_{FC}/e_{HS}] < e_{HS}[-2 + (3/2)\sqrt{2}].$$

But since the last expression is positive, the inequality does not give the sign of the first expression and hence (32).

$$\frac{F_{HS}^2}{F_{FC}^2} = \frac{4}{3}\frac{e_{HS}^3}{e_{FC}^3}\frac{re_{FC}^2+\varepsilon}{re_{HS}^2+\varepsilon}. \qquad (33)$$

Both the first and last ratio terms on the right-hand side of (33) are greater than 1, but the second term is less than 1. As a consequence, the relation between F_{HS}^2 and F_{FC}^2 is unclear a priori. However, substituting (14) and (28) into (33) yields

$$\frac{F_{HS}^2}{F_{FC}^2} = \frac{2\sqrt{2}}{3}\frac{4\theta+1+\sqrt{8\theta+1}}{8\theta+1+\sqrt{16\theta+1}}\left(\frac{4\theta+1+\sqrt{16\theta+1}}{2\theta+1+\sqrt{8\theta+1}}\right)^{3/2}; \qquad 0<\theta<0.5, \qquad (34)$$

where $\theta \equiv \varepsilon / \mathrm{K}$. Setting the RHS of (34) to 1 results in a unique solution of $\theta_F = 0.2025$. Further, as depicted in Figure 1, $F_{HS}^2 / F_{FC}^2 < (>) 1$, as $\theta < (>) 0.2025$. Combining this result with the above analysis, therefore, an increase in r will favor the HS (FC) network if $\theta < (>) 0.2025$.

To interpret this condition as well as the condition in (34), substitute the K expression from (13) into $\theta \equiv \varepsilon / \mathrm{K}$. Rearrangement then shows that, for θ to be less than 0.5, ensuring a unique solution to (27), the ratio ε/β of cost parameters must be less than the ratio γ/λ of demand parameters. Regarding network choice, if the cost ratio ε/β is close to the demand ratio γ/λ (if $0.2025 < \theta < 0.5$), then an increase in r favors the FC network. However, if ε/β is well below γ/λ (if $0 < \theta < 0.2025$), then an increase in r favors the HS network. Next, note that the inverse of the cost ratio is a measure of the economies of aircraft size holding e fixed, being equal to the fixed cost β/e divided by the marginal seat cost, ε/e. By contrast, the demand ratio γ/λ is the ratio of the cost of frequency delay and the cost associated with a higher load factor. Holding the demand ratio fixed, the analysis thus shows that when size economies, as measured by β/ε, are sufficiently strong (when ε/β is sufficiently small), an increase in the fuel price favors the HS network. When size economies are weaker, an increase in the fuel price favors the FC network.[19] Summarizing yields[20]

Proposition 3. *When the cost ratio* ε/β (*which is inversely related to aircraft size economies*) *is sufficiently large relative to the demand ratio* γ/λ, *an increase in the fuel price* (*or an equivalent imposition of emission charges*) *favors the FC network, yielding a possible movement away from current HS structures, Otherwise, emission charges strengthen any existing preference for the HS network.*

6. Empirical Evidence

Although a full empirical test of the model's predictions is beyond the scope of the paper, Figure 2 provides some suggestive evidence supporting some of the implications of the analysis. The Figure shows four series over the 1993-2008 period: the aviation fuel price, revenue passenger kilometers (RPK) per barrel of fuel (a measure of fuel productivity), non-fuel unit cost (total non-fuel operating expenses divided by RPK), and the share of fuel cost in operating expenses.[21] The first three series are normalized by setting 1993 values equal to 100. The numbers are based on aggregate data for all the world's airlines offering scheduled services, provided by the International Civil Aviation Organization (ICAO).

---

[19] Size economies must also not be so large as to violate the condition in (34), yielding the assumed unique solution to (27).

[20] Berry and Jia (2009), using a detailed empirical model, show that the cost of a connecting HS trip has risen relative to the cost of a direct, point-to-point trip since 2000. They attribute this effect to the greater fuel consumption of connecting passengers and argue that the relative cost change partly explains a documented shift toward use of direct flights. In the model, this outcome would require weak aircraft size economies.

[21] The data used in generating the four series are not available for the pre-1993 years.

The aviation fuel price series shows dramatic escalation, rising almost seven-fold over the period (from \$16.79 per barrel in 1993 to \$107.25 in 2008). This rise contrasts with the slight, gradual decline in non-fuel expenses, measured on an RPK basis. The airlines' response to escalating fuel prices is documented in the remaining two series. Fuel productivity (RPK per barrel of fuel) increases in step with the rising the fuel price, showing airline attempts to conserve fuel as it becomes more expensive. This productivity increase presumably arises from two separate adjustments portrayed in the model: (i) the improving fuel efficiency of new aircraft, along with a shift in fleet composition toward such aircraft, and (ii) higher load factors. Both adjustments lead to lower fuel usage per passenger kilometer, and together, they would serve to moderate the price-driven increase in fuel expenses as a share of operating cost. As can be seen in the fourth series, the rate of increase in this share declines after 2005, with the share actually falling in 2008 (from 26 percent in 2007 to 24 percent in 2008) despite the large fuel-cost spike in that year.

It should be noted that imposition of emissions charges would lead to a less dramatic increase in the effective price of fuel than the secular increase portrayed in Figure 2. A sense of the relevant magnitude can be gained using data and calculations presented by Scheelhasse and Grimme (2007). Consider their numbers for the low-cost carrier Ryanair, all of whose operations are within the EU, making for an easy appraisal of the impact of EU-level charges. Assuming that the price of a pollution permit (allowing the emission of one ton of CO_2) equals 30€, Scheelhasse and Grimme (2007) 's computations show that the value of Ryanair's required permits would equal 2.65€ per passenger in 2008. Ryanair's average fare is 44€, and assuming zero profit and a 25 percent fuel share in costs (using the end-of-period value from Figure 2), the implied fuel cost is 11€ per passenger. Since the permit cost per passenger is 24 percent of this cost, emission charges can then be viewed as leading to a 24 percent increase in the effective price of fuel. This increase is much smaller than the seven-fold rise over the 1993-2008 period but appreciable nevertheless.

Note that, for a higher-cost carrier, this calculation would involve a higher fare but (given such a carrier's higher labor costs relative to Ryanair's) a lower fuel share in total cost. The resulting effective fuel-price increase would then be similar in magnitude to the 24 percent value from above. Observe also that use of a fuel cost share smaller than the assumed 25 percent value would raise the percentage increase in the effective fuel price associated with emissions charges. Such a lower fuel cost share would be appropriate if the currently low fuel price persists.

7. Conclusion

This paper has explored the effect of airline emissions charges on airfares, airline service quality, aircraft design features, and network structure, using a detailed and realistic theoretical model of competing duopoly airlines. These impacts are derived by analyzing the effects of an increase in the effective price of fuel, which is the path by which emissions charges will alter airline choices. The results show that emission charges will raise fares, reduce flight frequency, increase load factors, and raise aircraft fuel efficiency, while having no effect on aircraft size. Given that these adjustments occur in response to the treatment of an emissions externality that is currently unaddressed, they represent efficient changes that move society closer to a social optimum.

Although these impacts are clear-cut, the effect of emission charges on the optimal structure of airline networks is ambiguous. Under some parameter values, emission charges may generate a shift away from current hub-and-spoke networks toward fully-connected, point-to-point networks. But the profitability of HS networks could be reinforced by emission charges under other parameter values.

The analysis has several limitations that could be addressed in future work. Most importantly, the model assumes that the total volume of airline passengers is fixed and thus unaffected by fuel prices and hence emission charges. More-realistic models that use the main elements of the present approach but incorporate an elastic demand for travel could be analyzed, but the resulting increase in complexity would necessitate use of numerical methods. The analysis is also based on a special, though realistic, form for the key function relating aircraft capital cost to seat capacity and fuel efficiency. Since some of the results (for example, the invariance of aircraft size to fuel prices) may depend on use of this functional form, the effect of adopting other specifications should be explored, perhaps numerically.

Airline emission charges are an important potential policy tool in the growing movement to address global warming, and they affect a highly visible industry that serves an important, affluent clientele. As a result, analysis of the impact of emission charges on airline decisions is a high-priority undertaking, and this paper has offered a first step toward filling this need.

References

Albers, S., Buhne, J.A. and Peters, H. (2009), Will the EU-ETS instigate airline network reconfigurations? Journal of Air Transport Management 15, 1-6.

Berry, S. and Jia, P. (2009), Tracing the woes: An empirical analysis of the airline industry. Unpublished paper.

Brueckner, J.K. and R. Flores-Fillol (2007), Airline schedule competition, Review of Industrial Organization 30, 161-177.

Brueckner, J.K. (2004), Network structure and airline scheduling, Journal of Industrial Economics 52, 291-312.

Brueckner, J.K. (2008), Schedule competition revisited, Unpublished paper, Department of Economics, UC Irvine.

Brueckner, J.K. and R. Girvin (2008), Airport noise regulation, airline service quality, and social welfare, Transportation Research Part B 42, 19-37.

Douglas, G.W. and Miller, J.C. (1974), Quality competition industry equilibrium, and efficiency in the price constrained airline market, American Economic Review 64, 657-669.

Forsyth, P. (2008), The impact of climate change policy on competition in the air transport industry, Discussion paper 2008-18, OECD/ITF, Paris.

Girvin, R. (2008), Airport noise regulation, airline service quality, and social welfare: The monopoly case, Journal of Transport Economics and Policy, forthcoming.

Oum, T.H., Zhang, A. and Zhang, Y. (1995), Airline network rivalry, Canadian Journal of Economics 28, 836-857.

Panzar, J.C. (1979), Equilibrium and welfare in unregulated airline markets, American Economic Review 69, 92-95.

Pels, E., Nijkamp, P. and Rietveld, P. (2000), A note on the optimality of airline networks, Economics Letters 69, 429-434.

Scheelhasse, J.D. and Grimme, W.G. (2007), Emissions trading for international aviation—an estimation of the economic impact on selected European airlines, Journal of Air Transport Management 13, 253-263.

Swan, W.M. and Adler, N. (2006), Aircraft trip cost parameters: A function of stage length and seat capacity, Transportation Research Part E 42, 105-115.

Wall, R. (2008), Terms of trading, Aviation Week and Space Technology, October 6, p. 64.

Zhang, A. (1996), An analysis of fortress hubs in network-based markets, Journal of Transport Economics and Policy 30, 293-308.

Appendices

To compare the load factors under the two networks, (26) and (12) are used to yield

$$\frac{l_{HS}^2}{l_{FC}^2} = \frac{4}{3}\frac{re_{HS}^2 + \varepsilon}{re_{FC}^2 + \varepsilon}\frac{e_{FC}}{e_{HS}} \qquad \text{(a1)}$$

While both the first and last ratio terms on the right-hand side of (a1) are greater than 1, the second term is less than 1 and so the relation between the load factors is unclear, *a priori*. However, substituting (22) and (27) into (a1) and letting $\theta \equiv \varepsilon / \mathrm{K}$ yields

$$\frac{l_{HS}^2}{l_{FC}^2} = \frac{2\sqrt{2}}{3}\frac{8\theta+1+\sqrt{16\theta+1}}{4\theta+1+\sqrt{8\theta+1}}\left(\frac{2\theta+1+\sqrt{8\theta+1}}{4\theta+1+\sqrt{16\theta+1}}\right)^{1/2}, \qquad 0 < \theta < 0.5 \qquad \text{(a2)}$$

Note that $0 < \theta < 0.5$ is equivalent to the earlier condition of $\varepsilon < \mathrm{K}/2$. The function in (a2) is depicted in Figure 3, where $l_{HS}^2 / l_{FC}^2 = 1$ for $\theta = 0.0030$. As can be seen, $l_{HS}^2 / l_{FC}^2 < 1$ and > 1 for $\theta < 0.0030$ and > 0.0030 respectively. Thus, only when ε is sufficiently small relative to $\mathrm{K} \equiv 2\gamma\beta/\lambda$ will the HS load factor be smaller than the FC load factor. For most of the allowable range of θ, the load factor is greater under the HS network than under the FC network.

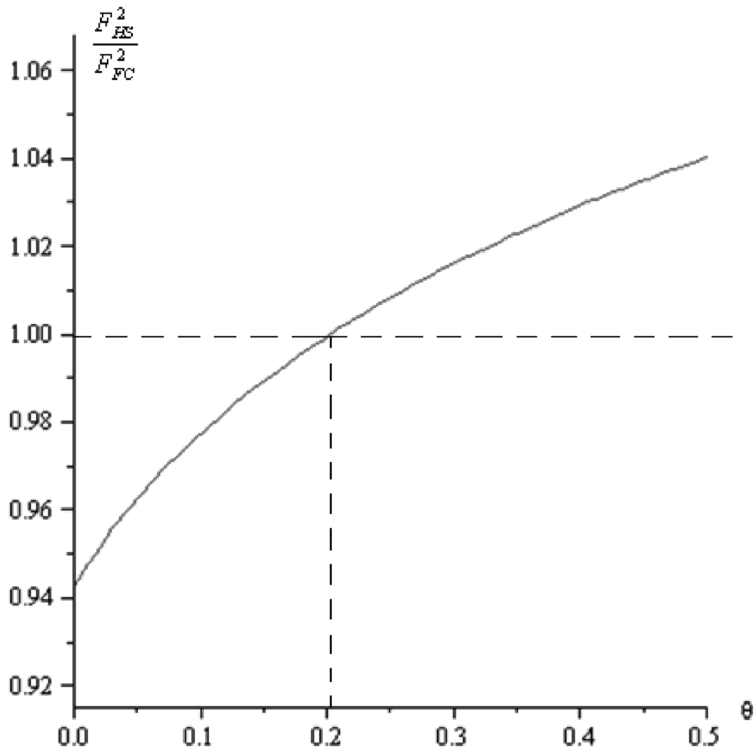


Figure 1: Effect of emissions charge on airline networks, $\theta \equiv \varepsilon / \mathrm{K}$

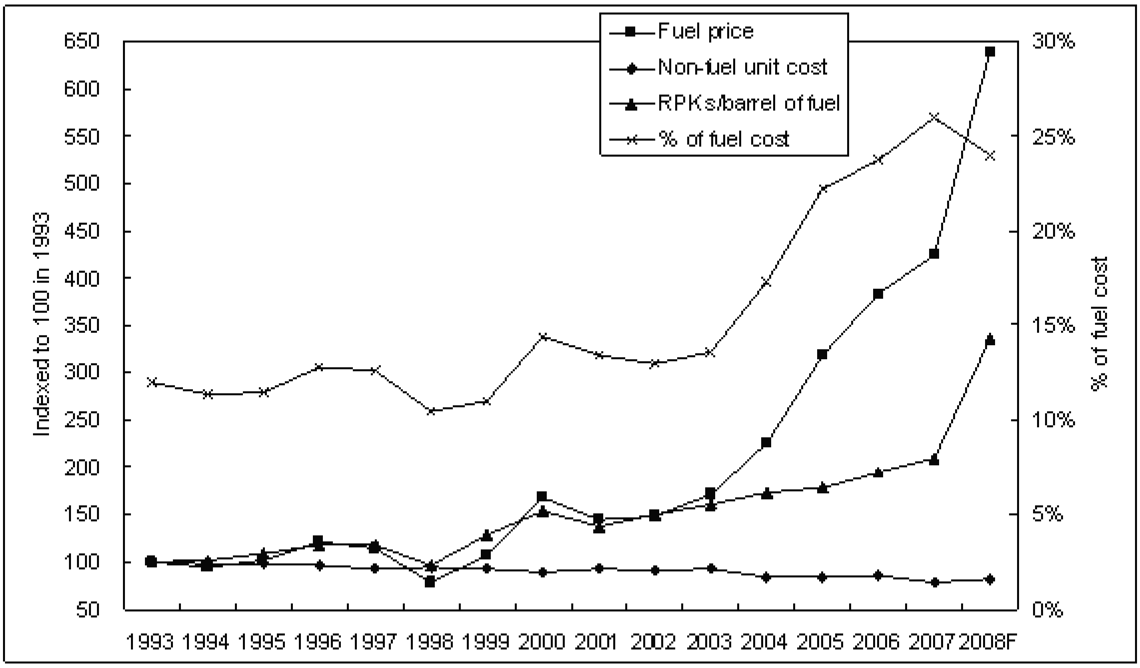


Figure 2: Fuel price, fuel productivity, share of fuel cost and non-fuel unit cost, 1993-2008

Notes: RPK = Revenue passenger kilometers; Fuel productivity = World RPK / barrel of fuel; Share of fuel cost = % of fuel cost in total operating expenses; Non-fuel unit cost = Total non-fuel operating expenses / RPK; 2008F = forecast figures for 2008; Fuel price, fuel productivity and non-fuel unit cost are all indexed to 100 in 1993.

Source: Authors' calculation based on the ICAO (International Civil Aviation Organization) database, http://icaodata.com/.

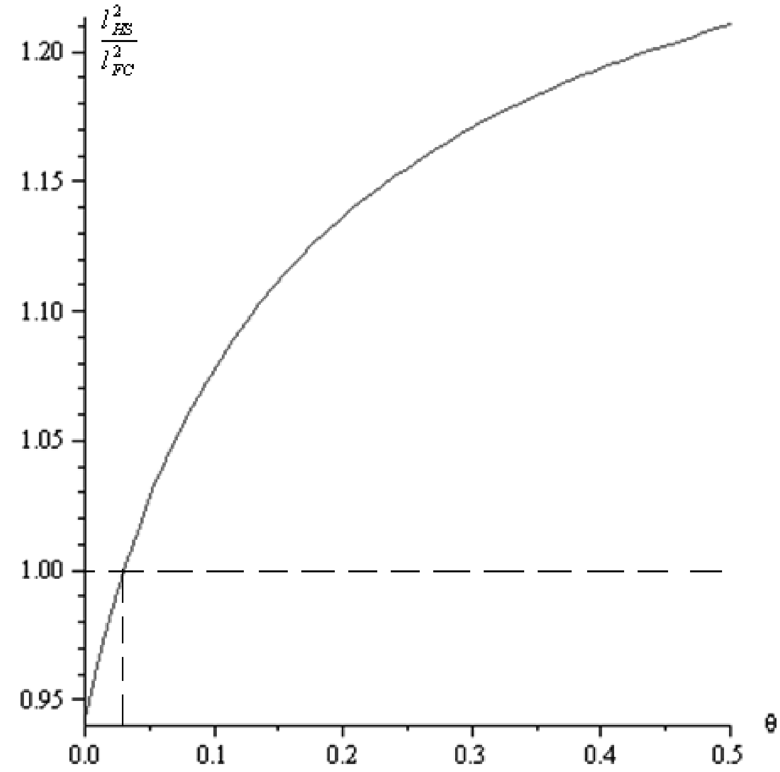


Figure 3: Comparison of load factors under two networks, $\theta \equiv \varepsilon / \mathrm{K}$

Finding reliable shortest paths in stochastic time-dependent road networks for online ATIS applications

Bi Yu Chen, William H K Lam, Mei Lam Tam and Agachai Sumalee

Department of Civil and Structural Engineering, The Hong Kong Polytechnic University, Hong Kong.

Abstract

In urban road networks, travel times should be considered as stochastic and time-dependent (STD) variables under uncertain conditions due to recurrent and non-recurrent congestions. Finding reliable paths in STD networks is vital to logistic applications especially for the 'just-in-time' goods distribution systems. This paper investigates the reliable shortest path problem (RSPP) in large-scale STD networks. A new model is proposed to determine the optimal path to arrive at destination within a certain level of confidence while minimizing the path travel time, toll charge and travel distance. An efficient heuristic solution algorithm is developed to solve the RSPP. Numerical results based on real-world advanced traveller information system are presented to demonstrate the features of the proposed model and solution algorithm for online routing applications.

Keywords: Reliable shortest path problem, stochastic time-dependent network, advanced traveller information system

1. Introduction

Road networks are exposed to uncertainties on both demand and supply sides of the transportation system. The uncertainties can come from various sources including fluctuations of traffic demand, accidents, breakdowns, bad weather, etc. Due to these uncertainties and temporal changes of demand and supply in the network, travel times on links are stochastic and time-dependent (STD) in which the statistical distribution of travel time may change with time of day. Finding a reliable shortest path in the STD network is vital to logistic applications with time-window constraints particularly for the 'just-in-time' goods distribution systems. In order to compensate for the uncertainties in travel time, freight operators are likely to choose reliable paths which provide sufficient "buffer times" to ensure on-time deliveries.

With the advanced traveler information system (ATIS), an opportunity exists for improving the efficiency and reliability of systems used in fleet operations. The ATIS utilizes technological infrastructures to collect real-time traffic data. There are numerous techniques for real-time travel time data collection including inductive loop detectors, automatic vehicle identification (AVI), global positioning system (GPS), cellular phone tracking and image processing techniques. The collected data are then transmitted to a central traffic management center to generate real-time traffic information. The traffic information is subsequently disseminated to users through various media, or applied with other intelligent transportation system applications, e.g. a route guidance system.

Several studies on the stochastic time-dependent shortest path problem (STDSPP) have already been carried out. Hall (1986) first studied the STDSPP in a transportation network. He found that the STDSPP cannot be solved by a standard shortest path algorithm due to the violation of the Bellman's principle of optimality. Fu and Rilett (1998) used Taylor series expansions to approximate the mean and variance of the stochastic path travel time. They proposed a heuristic solution algorithm for finding an optimal path with the least expected time in the STD network. Miller-Hooks and Mahmassani (2000) established a dominance rule for determining paths with the least expected travel time in STD networks, and designed a label correcting algorithm to find these paths. However, most of these existing methods in the STDSPP literature focus mainly on finding paths with the least expected travel time (i.e. mean travel time), and ignore the consideration of travel time reliability of each path.

The objective of this paper is to develop a new model and solution algorithm to solve the reliable shortest path problem (RSPP) in the STD network. The RSPP problem involves finding a path to arrive at a destination within a certain level of confidence while minimizing the path travel time, toll charge, and travel distance. The rest of paper is organized as follows. The next section presents the definition of the RSPP. The estimation of path travel time distribution in STD network is then introduced in Section 3. We then propose a solution algorithm for solving the RSPP in Section 4. A case study based on a real-world ATIS case is also conducted in Section 5 to demonstrate the proposed model and the performance of the solution algorithm. Finally, conclusions are given together with recommendations for further study.

2. Problem Definitions and Model Formulations

Let G = (N, A) be a STD network, where N = $\{n_1, n_2, \ldots, n_m\}$ is a set of nodes and A = $\{a_1, a_2, \ldots, a_n\}$ is a set of links A⊆N×N. Each link a_{ij} , connecting two adjacent nodes i and j, is associated with an attribute vector ($X_{ij}(t)$, τ_{ij}, d_{ij}), where $X_{ij}(t)$ is the predicted link travel time, τ_{ij} is the toll charge on the link, and d_{ij} is the link length. Without loss of generality, the link toll rate and length are assumed to be deterministic and static; while the link travel time, $X_{ij}(t)$, is assumed to be a STD variable whose probability distribution depends on the time of day. In the context of ATIS, only the observed mean and variance of link travel times are available for discrete time periods (eg. 5 min). The mean and variance of link travel time at discrete time periods are denoted as $u_{x_{ij}}(t)$ and $v_{x_{ij}}(t)$. Therefore, the general assumptions are that the link travel time during each discrete time period follows the normal distribution, and all link travel times are statistically independent. This assumption is supported by some empirical studies which found the vehicular travel times on congested links to follow the normal distribution (Richardson and Taylor, 1978; Taylor, 1982). Nevertheless, it is noteworthy that the proposed model and algorithm in this paper can also be applied to other two-parameter distributions (eg. Gamma distribution).

Let P^{rs} = $\{p_1, \ldots, p_i\}$ be a set of non-cyclic paths from origin r to destination s. Let t_r denotes the departure time from origin r. Let δ_{ij} be a binary decision variable, where if the link a_{ij} is on the optimal path then $\delta_{ij} = 1$, otherwise $\delta_{ij} = 0$. Denote the travel time, toll charge and travel distance of the path p as $X_p(t_r)$, τ_p and d_p. Then, the travel time, toll charge and travel distance of the path p can be formulated as the summation of link attributes as follow

$$X_p(t_r) = \sum_{(i,j)\in A} X_{a_{ij}}(t)\delta_{ij} \tag{1}$$

$$\tau_p = \sum_{(i,j)\in A} \tau_{ij}\delta_{ij} \tag{2}$$

$$d_p = \sum_{(i,j)\in A} d_{ij}\delta_{ij} \tag{3}$$

It is clear that the toll charge and travel distance of the journey are deterministic and static, while the path travel time of the trip, $X_p(t_r)$ is a STD variable.

Under the travel time uncertainty, freight operators may be willing to pay a reliability premium to avoid unreliable paths. In order to take into account path travel time reliability in the shortest-path problem, a dependent-chance model (DCM) and a chance-constrained model (CCM) were proposed in the literature. As shown in Figure 1, the DCM aims to choose a reliable path with the maximum probability of travelers to arrive at a destination on time within a given travel time budget (Frank, 1969; Mirchandani, 1976; Chen and Ji, 2005). On the other hand, as shown in Figure 2, the CCM focuses on choosing a reliable path with the minimum travel time budget at a certain confidence level of arriving on time (Chen and Ji, 2005; Chen et al., 2008). For the DCM, an underlying assumption is that the travelers can determine their travel time budgets prior to their trips. However, the travel time budget depends largely on the distance between each origin and destination (O-D) pair and the congestion level in the network. The travelers under the DCM may thus unable to define an appropriate travel time budget prior to their trips. The CCM, however, requires the travelers to express their desired confidence level of arriving at the destination on time which can be related to the importance of their trips. This paper postulates that the travelers can better identify their desired confidence levels of arriving on time (based on their trip purposes) rather than defining the travel time budget. Therefore, we adopt the CCM for determining a reliable path in the STD network.

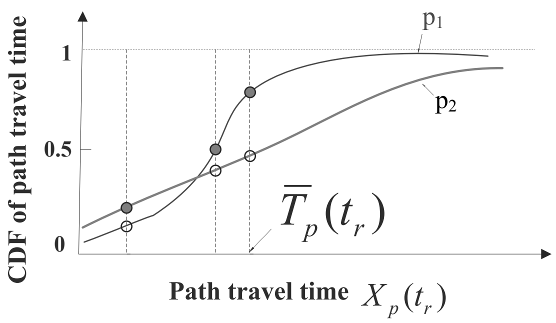


Figure 1: Dependent-chance model

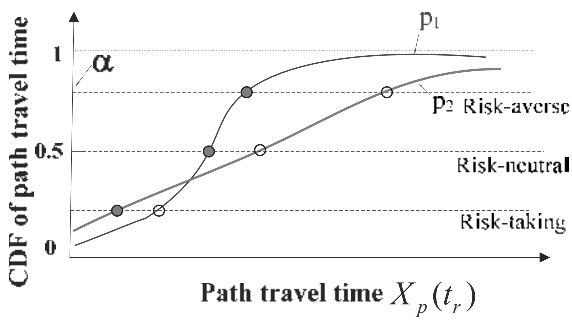


Figure 2: Chance-constraint model

Given an O-D pair and a departure time t_r, the RSPP involves finding a reliable shortest path p^* that minimizes an objective function, which is a weighted sum of the travel time budget (denoted by $\overline{T}_p(t_r)$), toll charge (denoted by τ_p), and travel distance (denoted by d_p), while arriving at a destination within a certain confidence level of . The objective function of the RSPP can defined as follows:

$$\overline{T}_p(t_r) + (\tau_p + d_p VOD) / VOT \qquad (4)$$

Eq. 4 defines the utility cost of a journey that travelers want to minimize. A value of distance denoted as *VOD* is introduced for converting travel distance into the petrol cost. A value of time *VOT* are defined as weighting factors for converting the toll charge and petrol cost into the travel time. The larger value of *VOT* indicates that travelers are more concerns to the travel cost saving in their route choice decision. The travel time budget ($\overline{T}_p(t_r)$) is determined by the chance constraint:

$$\Pr(X_p(t_r) \le \overline{T}_p(t_r)) \ge \alpha \qquad (5)$$

The travel time budget is determined such that the cumulative probability of path travel time under this level is greater than or equal to a confidence level of α. The confidence level is a predetermined threshold representing the travelers' attitude towards the travel time uncertainty. The larger the value of confidence level, the more concerns of the travelers to the reliability of the path travel time. By inverting the cumulative distribution function (CDF) of path travel time in Eq. 5, the chance constraint can be reformulated as

$$\overline{T}_p(t_r) \geq \Phi_p^{-1}(\alpha) \tag{6}$$

where $\Phi_p^{-1}(\alpha)$ is the inverse of the CDF of the path travel time as a function of α. Substituting Eq. 6 into Eq. 4, the objective of the RSPP can be redefined and the RSPP can be formulated as:

$$\min_{\delta_{ij}} \left[\Phi_p^{-1}(\alpha) + (\tau_p + d_p VOD) / VOT \right] \tag{7}$$

s.t

$$X_p(t_r) = \sum_{(i,j)\in A} X_{a_{ij}}(t)\delta_{ij} \tag{8}$$

$$\tau_p = \sum_{(i,j)\in A} \tau_{ij}\delta_{ij} \tag{9}$$

$$d_p = \sum_{(i,j)\in A} d_{ij}\delta_{ij} \tag{10}$$

$$\sum_{j\in SCS(i)} \delta_{ij} - \sum_{k\in PDS(i)} \delta_{ki} = \begin{cases} 1 & \forall\, i = r \\ 0, & \forall\, i \neq r;\, i \neq s \\ -1 & \forall\, i = s \end{cases} \tag{11}$$

Eqs. 8-10 are based on the Eqs. 1-3 explained earlier. Eq. 11 ensures the decision variables are binary.

3. Constructing an Approximation of Probability Distributions of Path Travel Time

In this paper, we assume that the stochastic travel times across different links are independent in which the distribution of travel time is time-dependent. Therefore, the distribution of the travel time on a certain link on a path is determined by the arrival time of the traveler to that link which in turn depends on the travel times on all previous links of the traveler on that path. Consider a vehicle arriving at node i at time Y_i, then traversing link a_{ij} and arriving at node j at time Y_j. The arrival time at node j, Y_j, is equal to the sum of the arrival time at upstream node i and the travel time of link a_{ij}:

$$Y_j = Y_i + X_{ij}(t \mid Y_i) \tag{12}$$

Notice that $X_{ij}(t \mid Y_i)$ is conditional on Y_i which represents the time-dependent property of the link travel time. It is possible to evaluate the probability density function (PDF) of the arrival time at node j if the joint continuous PDF of X_{ij} over a continuous arrival time is available. However, this may result in a very complex form of the PDF, and in practice the joint PDF of the link travel time may not be available (Fu and Rilett, 1998; Pattanamekar et al., 2003; Chang et al., 2005).

Similar to Miller-Hooks and Mahmassani (1998), we propose a simple and effective discretization scheme that overcomes the above difficulties. The CDF of the link travel time $X_{ij}(t)$ at each discrete time interval is first discretized in to λ intervals with an equal cumulative probability denoted as ($q\omega$-ω, $q\omega$), q=1,2,…, λ, where ω is the cumulative probability of an interval, $0<\omega<1$ and $\lambda\omega=1$. According to the Mean-Value Theorem, within each interval ($q\omega$-ω, $q\omega$) there is at least one point x_{ijq} satisfies:

$$f(x_{ijq}) = \omega / \{\Phi_{x_{ij}}^{-1}(q\omega) - \Phi_{x_{ij}}^{-1}(q\omega - \omega)\} \tag{13}$$

where $f(x_{ijq})$ is the PDF of link travel time at point x_{ijq}, and $\Phi_{x_{ij}}^{-1}(q\omega-\omega)$ and $\Phi_{x_{ij}}^{-1}(q\omega)$ are inverse of CDF at the boundary points of the interval. Under normality of link travel time, Eq. 13 can be normalized as

$$f(z_{ijq}) = \omega / \{\Phi_z^{-1}(q\omega) - \Phi_z^{-1}(q\omega-\omega)\} \tag{14}$$

where z_{ijq} is the normalized value of x_{ijq}, and $\Phi_z^{-1}()$ is the inverse of the CDF of standard normal distribution (SND) as function of confidence level. Substituting Eq. 14 into the definition of PDF of SND, z_{ijq} can be obtained as:

$$z_{ijq} = \pm\sqrt{-2\ln(f(z_{ijq})\sqrt{2\pi})} \tag{15}$$

The Eq. 14 gives two solutions and we can discard an infeasible solution by:

$$\Phi_z^{-1}(q\omega-\omega) < z_{ijq} < \Phi_z^{-1}(q\omega) \tag{16}$$

Then x_{ijq} can be calculated using:

$$x_{ijq} = z_{ijq}\sqrt{v_{x_{ij}}(t)} + u_{x_{ij}}(t) \tag{17}$$

Therefore, the probability mass function (PMF) of the discretization is given as:

$$P(X_{ij}(t) = x_{ijq}) = \omega, q = 1,2,...,\lambda \tag{18}$$

An illustration of the discretization is shown in Figure 3.

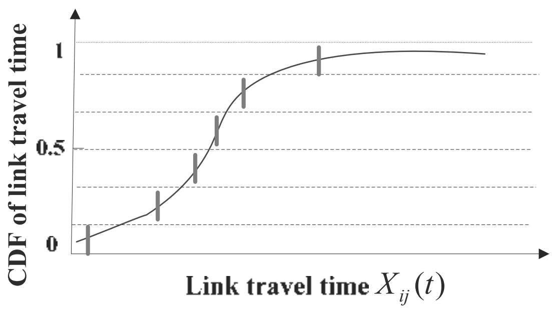


Figure 3: Discretization of link travel time

Similarly, the CDF of arrival times at node *i* can be discretized into η equal cumulative probability intervals denoted as (qε−ε, qε), q=1,2,…,η, where ε is the cumulative probability of an interval, 0<ε<1 and λη=1. The PMF of arrival travel time Y_i is given:

$$P(Y_i = y_{iq}) = \varepsilon, q = 1,2,...,\eta \tag{19}$$

Accordingly, the arrival time Y_j which depends on arrival time at node *i* can be formulated as:

$$y_{jq_0} = y_{iq_1} + x_{ijq_2}(y_{iq_1}), \quad q_0 = q_1q_2, \quad q_1 = 1,2,...,\eta, \quad q_2 = 1,2,...,\lambda \tag{20}$$

$$P(Y_j = y_{jq_0}) = \varepsilon\omega, q_0 = 1,2,...,\eta\lambda \tag{21}$$

Then by sorting the $y_{jq_0}, q_0 = 1,2,...,\eta\lambda$ in ascendant order, we can generate the CDF of arrival time at node *j*. To prevent the number of elements in the CDF of arrival time from growing exponentially with the number of constituent links, the PMF of arrival time shown in Eqs. 20-21 have to be aggregated:

$$y_{jq} = (\sum_{q_0=(q-1)\lambda+1}^{q\lambda} y_{jq_0}) / \lambda, \; q = 1,2,...,\eta \tag{22}$$

$$P(Y_j = y_{jk}) = \varepsilon, k = 1,2,...,\eta \qquad (23)$$

Therefore, by propagation of the CDF of the arrival time at a node along the links of a path, Eqs. 13-23 can provide a way to calculate the distribution of arrival time at destination.

4. Solution Algorithm

The RSPP is an NP-Complete problem. The only way to find the optimal solution is to enumerate all feasible paths (Hall, 1986). However, it is impractical and computationally intensive to enumerate all feasible paths in any large-scale network. The heuristic algorithm proposed in this paper instead tries to determine potential better paths. The algorithm proceeds as follows:

Step 1. Find K reliable shortest path p_k (k=1…K) from origin to destination.

Step 2. Set k=1, take the kth shortest path p_k. Calculate the CDF of the path travel time using Eqs. 13-23, and the toll charge and travel distance using Eqs. 2-3. Calculate the utility cost denoted as $U(p_k)$ using Eq. 7.

Step 3. if k>K then p^* is the 'optimal' path; stop. Otherwise, go to Step 4

Step 4. Set k=k+1, take the kth shortest path p_k. Calculate the CDF of the path travel time using Eqs. 13-23, and the toll charge and travel distance using Eqs. 2-3. Calculate the utility cost $U(p_k)$ using Eq. 7. If $U(p_k) < U(p^*)$ then denote $U(p_k)$ as $U(p^*)$. Go to Step 3.

In the first step, a set of K reliable shortest paths (KRSP) are identified. The KRSP can be solved using the K reliable shortest path algorithm which is similar to RSPRA (Chen et al., 2008). The last three steps determine the 'optimal' path with minimum utility cost.

5. Numerical Examples

In this section, a numerical example based on the real-world ATIS is conducted to test the proposed model. The proposed algorithm is coded in the Visual Basic 6 programming language and further integrated with the ArcGIS software for numerical experiments. All experiments are conducted in AMD Athlon computer with a duo-core 1.9GHz CPU and run on the Windows XP operation system.

In Hong Kong, real-time traffic information for major urban roads is provided by a Real-time Travel Information System (RTIS) (Tam and Lam, 2008). In the RTIS, area-wide link travel times are generated at 5-min intervals. A GIS-based website portal has been developed for disseminating the RTIS results to road users (http://rtis.td.gov.hk/rtis). Based on the RTIS, a solution algorithm for the short-term travel time prediction in urban roads of Hong Kong has also been proposed (Tam and Lam, 2009). As shown in Figure 4, the road network of the whole territory of Hong Kong with 1367 nodes and 3655 links is adopted in the RTIS. In this study, three month's RTIS travel time estimates are collected from April to June in 2008. The estimated travel times on thirteen Wednesdays are used to generate the mean and variance of link travel times for the whole network.

The toll rates of tunnels/bridges in Hong Kong in 2008 are collected from the Transport Department of Hong Kong Government (TD) in this study. According to the information from the TD, there are totally 11 toll tunnels/bridges in Hong Kong in 2008, in which three of them are cross-harbor tunnels. These three cross-harbor tunnels are significantly different in toll rates, mean travel times and travel time variances. Cross Harbor Tunnel (CHT) is the cheapest but with the largest travel time variation. Western Harbor Crossing (WHC) is the most reliable one and the expected travel time is also the least. However, the toll charge is the most expensive. The expected travel time for Eastern Harbor Crossing (EHC) is the largest but its travel time variation is similar to that of WHC.

In the experiments, we consider a residential zone at Richland Gardens (RG) as the origin and the Central Business District (CBD) in Central as the destination with the departure time at 8:00AM to examine the effect of different confidence level, value of time, and value of distance on travelers' route choice. Table 1 provides the numerical results under different confidence levels, value of times and value of distance. The reliable shortest paths from origin to destination are also illustrated in Figure 4. As shown in Figure 4, three paths are found under each of the confidence levels and value of times. Path2 passing through CHT is the cheapest route which likely to be a preferred path of the travel cost sensitive travelers. In contrast, travelers with higher value of time would prefer Path1 and Path3 which are faster and more reliable.

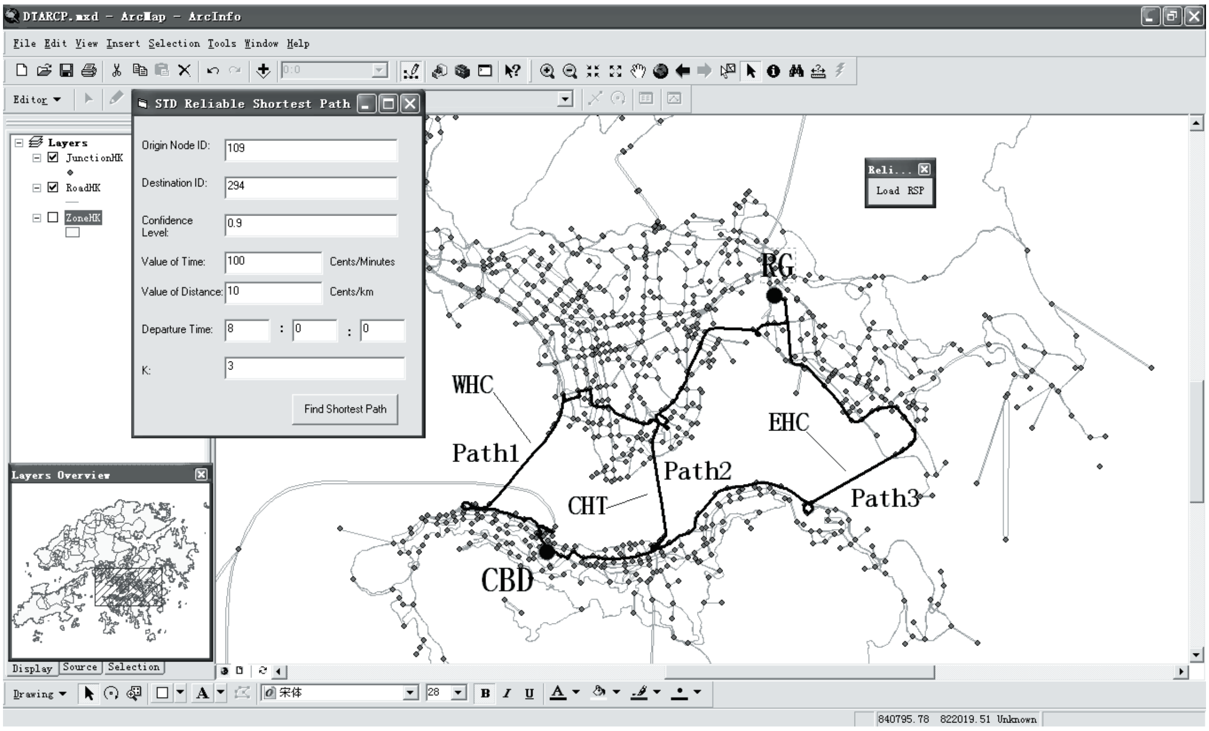


Figure 4: Reliable shortest paths from RG (origin) to CBD (destination)

Table 1: Numerical results from RG (origin) to CBD (destination)

Confidence level	VOT (cents / min)	VOD (cents / km)	Path	Arrival Time	Utility cost (min)	Travel cost (HKD)	Expected travel time (min)	Std. dev. of travel time (min)
0.5	100	10	Path2	**8:29:28**	50.61	21.14	29.48	5.35
0.6	100	10	Path2	**8:30:03**	51.19	21.14	29.48	5.35
0.7	100	10	Path2	**8:30:40**	51.81	21.14	29.48	5.35
0.8	100	10	Path2	**8:31:25**	52.56	21.14	29.48	5.35
0.9	100	10	Path2	**8:32:26**	53.57	21.14	29.48	5.35
0.99	**100**	10	Path2	**8:34:52**	56.01	**21.14**	29.48	5.35
0.99	**300**	10	Path3	8:30:25	39.30	**26.64**	27.73	1.33
0.99	**500**	10	Path3	8:28:26	33.76	**26.64**	27.73	1.33
0.99	**1000**	**10**	Path1	8:28:26	33.09	**46.60**	25.89	1.20
0.99	1000	**20**	Path1	8:28:26	33.25	**48.20**	25.89	1.20
0.99	1000	**30**	Path1	8:28:26	33.41	**49.80**	25.89	1.20
0.99	1000	**50**	Path1	8:28:26	33.73	**53.00**	25.89	1.20

As shown in Table 1, with an increase of the confidence level, the utility cost is increasing. In this case, travelers may become more concern in the travel time variability, and assign a higher travel time budget, i.e. departing from the origin earlier to ensure an on-time arrival. By keeping the confidence level at 0.99 and varying the value of time and value of distance, we can observe from Table 1 that with the higher value of times and value of distance travelers are more willing to pay a premium to use the path with less mean and variance of travel time to ensure their on-time arrival. By varying the values of time, value of distance and levels of confidence, the proposed model is robust to take account traveler' preference towards travel time, travel cost and travel time variability.

The computational time of the proposed algorithm with respect to the number of K reliable shortest paths is shown in Figure 5. The experiments are conducted based on the road network in Hong Kong with 244 O-D pairs from various zones to the CBD. As a result, the averages of computational time required are presented. It can be observed that the computational time is within 1 second when the K value less than 5. However, the computational time increases significantly with the K value.

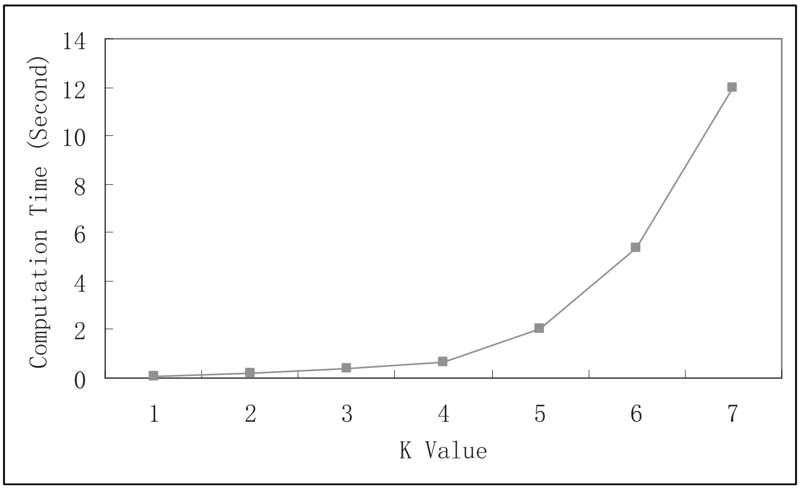


Figure 5: Computational time of proposed algorithm

6. Conclusions and Further Research

In this paper, a new model was proposed for modeling the reliable shortest path problem (RSPP) in stochastic and time dependent networks for online ATIS applications. A heuristic solution algorithm has been developed to solve the RSPP in the proposed model. Numerical results showed that the proposed model was robust to take account travelers' various preference towards the travel cost, travel time and travel time variability of their journeys. The computational results indicated that the proposed solution algorithm was applicable for solving the RSPP efficiently in large-scale road networks when the number of reliable shortest paths for each O-D pair was less than 5. It was found that the proposed model and solution algorithm can be used for development of online routing applications in large-scale road networks. Further study is required to validate the model results with empirical results and to compare the solution of the proposed algorithm against the optimal solution in networks under various congested conditions.

7. Acknowledgements

The work described in this paper was jointly supported by a research grant from the Research Grant Council of the Hong Kong Special Administration Region to the Hong Kong Polytechnic University (Project No. PolyU 5195/07E) and internal research grants J-BB7Q and 1-BBZG from the Research Committee of the Hong Kong Polytechnic University.

References

Chang, T.S., Nozick, L.K. and Turnquist, M.A. (2005), Multiobjective path finding in stochastic dynamic networks, with application to routing hazardous materials shipments, Transportation Science 39(3): 383-399.

Chen, A. and Ji, Z.W. (2005), Path finding under uncertainty, Journal of Advanced Transportation 39(1): 19-37.

Chen, B.Y., Lam, W.H.K. and Tam, M.L. (2008), Using a multi-criteria approach to solve path finding problem in road network with uncertainty, In: H. P. Lo, S. C. H. Leung and S. M. L. Tam, (eds.), the 13th International Conference of Hong Kong Society for Transportation Studies: Hong Kong, pp. 229-238.

Frank, H. (1969), Shortest paths in probabilistic graphs, Operations Research 17(4): 583–599.

Fu, L. and Rilett, L.R. (1998), Expected shortest paths in dynamic and stochastic traffic networks, Transportation Research Part B 32(7): 499-516.

Hall, R.W. (1986), The fastest path through a network with random time-dependent travel-times, Transportation Science 20(3): 182-188.

Mirchandani, P.B. (1976), Shortest distance and reliability of probabilistic networks, Computers and Operations Research 3(4): 347–355.

Miller-Hooks, E.D. and Mahmassani, H.S. (1998), Optimal routing of hazardous materials in stochastic, time-varying transportation networks, Transportation Research Record 1645: 143-151.

Miller-Hooks, E.D. and Mahmassani, H.S. (2000), Least expected time paths in stochastic, time-varying transportation networks, Transportation Science 34(2): 198-215.

Pattanamekar, P., Park, D., Rilett, L.R., Lee, J. and Lee, C. (2003), Dynamic and stochastic shortest path in transportation networks with two components of travel time uncertainty, Transportation Research Part C 11(5): 331-354.

Richardson, A.J. and Taylor, M.A.P. (1978), A study of travel time variability on commuter journeys, High Speed Gound Transportation Journal 12(1): 77-99.

Tam, M.L. and Lam, W.H.K. (2008), Using automatic vehicle identification data for travel time estimation in Hong Kong, Transportmetrica 4(3): 179-194.

Tam, M.L. and Lam, W.H.K. (2009), Short-term Travel Time Prediction for Congested Urban Road Networks, In: 88th Transportation Research Broad Annual Meeting (CD-ROM), Washington, D.C.

Taylor, M.A.P. (1982), Travel time variability- the case of two public modes, Transportation Science 16(4): 517-521.

Website of Transport Department of the Government of the Hong Kong Special Administrative Region: http://rtis.td.gov.hk/rtis/

A six sigma based quality improvement framework for Taiwan Taoyuan International Airport

Chih-Hsien Chen

Department of Management, Fo-Guang University, Yilan, Taiwan.

Abstract

The air transportation includes the complicated international interface, once the process error will cause serious international influence and significant casualties. In recent years, the operation results and flight safeties of Taiwan Taoyuan International Airport all express not well, and the process quality causes the important influence to the operation results and flight securities of Airports exactly. Therefore, the purpose of this study is establishing a DMAIC process improvement framework via an integrated Six Sigma model in order to improve the operation quality and reduce the flight safety accident. However, the Six Sigma that based on the customer needs to reach the quality results perfectly—namely no more than 3.4 failures during one million operations—is the most suitable to match the improvement tool of the air transportation industry. From the study results, the process quality improvement framework achieving zero defects is established by combining the Balance Scorecard, Theory of Constraint, Quality Function Deployment, and benchmarking management. In addition, this framework not only can improve the operation results of air terminals but also can provide the authorities to make decisions for process quality improvement of air transportation.

Keywords: Airports, Six Sigma, DMAIC, Quality improvement

1. Introduction

The air transport industries include Airlines, airports, ground handling agents, custom brokerage firms, shipping companies, cargo agents, consignors, travel agencies, travelers, and the international express industries. The air transport can be divided into passenger and the cargo according to its target client. Cargo tended to be seen as an attached business of passenger business in the past. However, even if of the turnover of cargo has already surpassed turnover of passenger; the quantity of passengers' flights is still more than cargo due to most airline companies use cargo on passenger flights.

A typical airport has six main users: passengers, airline companies, airport staff, customs, cargo business agents as well as brokers. The main revenue of airport is from passenger, and the revenue of air cargo is mainly from consignor. According to Heathington and Jones (1975), when airports consider renewal design or renovation, should give priority to consignors and passengers. In other words, the airports should reinforce the efficiency of operation and management to provide customers better service.

Reviews the aviation security record of Taiwan, according to statistical data of IATA and CAA, the accident rate was about 1.74 per one million flight hours, approximately three times more than global average 0.56 from 1993 to 2006. From 1998 to 2006, there were132 violations of Civil Aviation Act, of which China Airlines (CI) had 39 violations and EVA Group had about 18 times. In recent 15 years, Taiwan had three aviation accidents and 700 people died, those were on April 26, 1994, while the CI aircraft was making an approach to runway of Nagoya Airport, the aircraft stalled and crashed, 264 people died. On February 16, 1998 CI A300-600R crashes outside the runway of CKS International Airport, and 202 people died. Again, on May 25, 2002 when CI flew to Hong Kong, vanished from radar screen on the way to Makung Northwest (Penghu), 225 people died.

Accidents happen every year, and news spread out fast around the world. Why was there no way to stop the aviation accident tragedy in technical highly developed today, what means used to prevent this kind of accident's occurrence? According to the related statistics report, the heavy accident which artificial creates approximately composes the accident total 65%~80%. Therefore, this research applied the Six Sigma, that is no more than 3.4 defects per million operations, inducts the procedure and quality control of aviation transport supply chain, to provide a good reference for the performance improvement.

2. Paper Review

The operation performance of air transport in Taiwan can be explained in two aspects: the aviation passenger transportation and the air freight. The CKS international airport started operation in 1979; the number of passengers grew tremendously. Again in 1988, the government implemented the Open Sky Policy as well as release Taiwanese visit relatives China, the number of passengers increased 25%. Moreover, the number of passengers in 1991 reached ten million of people for the first time, 2004 has surpassed twenty million travelers. However, the first appearance negative growth (- 1.7%) in 1998, and 2003 decreased 19.3% because of the SARS. The Taiwan Taoyuan International Airport (TTIA) had worse business performance in the recent years; the world ranking and the passenger/cargo traffic were declining in recent years. Detailed material is as shown in Table 1.

Table 1: Operation results of Taiwan Taoyuan International Airport

Year	Movement (thousand flights)	Growth rate	Passenger (thousand pax)	Growth rate	Cargo (thousand tons)	Growth rate
1988	43	7.5%	7,129	25.5%	485	5.7%
1989	48	11.6%	8,297	16.4%	551	13.6%
1990	57	18.8%	9,680	16.7%	595	8.0%
1991	62	8.8%	10,364	7.1%	634	6.6%
1992	69	11.3%	12,072	16.5%	723	14.0%
1993	74	7.2%	12,494	3.5%	743	2.8%
1994	83	12.2%	13,349	6.8%	747	0.5%
1995	92	10.8%	14,478	8.5%	754	0.9%
1996	101	9.8%	15,614	7.8%	796	5.6%
1997	108	6.9%	15,969	2.3%	914	14.8%
1998	109	0.9%	15,699	-1.7%	932	2.0%
1999	110	0.9%	17,044	8.6%	1,057	13.4%
2000	116	5.5%	18,681	9.6%	1,209	14.4%
2001	124	6.9%	18,461	-1.2%	1,190	-1.6%
2002	132	6.9%	19,228	4.2%	1,381	16.1%
2003	126	-5.1%	15,514	-19.3%	1,500	8.6%
2004	149	18.5%	20,084	29.5%	1,701	13.4%
2005	153	2.5%	21,111	8.1%	1,705	0.2%
2006	158	3.3%	22,857	8.3%	1,699	-0.4%
2007	160	1.5%	23,426	2.5%	1,593	-6.2%
Average		7.3%	Average	7.9%	Average	6.6%

Source: Civil Aeronautics Administration, Taiwan (2008)

Seneviratne and Martel (1991) proposed a set criterion can be used for the performance index: 1, it may respond the explicit management goal; 2, it may simply define and the quantification; 3, it does not need the depth and expensive materials; 4. it may presents the reasonable sensitive change because of the management activity or the improvement activity. However, the process quality is the primary factor that affects the operation performance of air transport industry, and has the influence to airline's flight safety. DMAIC—Design, Measure, Analysis, Improve, and Control—is one type of tool of Six Sigma project for continuous process improvement. Generally speaking, Six Sigma has two dimensions: one is the statistical, the related statistical variation is the primitive definitions of Six Sigma, i.e. no more than 3.4 ppm. The other dimension is Six Sigma may regard as one kind of desire improvement for the enterprise profit, the lifting efficiency/quality, the strengthening competitive advantage, reaches the customer demand or surmounted one kind of business strategy which the customer expected (Sokovic *et al*., 2005).

Wilson (2003) designed a BSC framework for the quality performance, to study the customers' satisfaction in five 9 main quality perspectives, the result shows transportation operator should strike a balance in all quality perspectives and prioritize the requirements of key customers. Rouse et al. (2002) spent four years to study the performance monitoring and control system of international airlines' maintenance operation. The monitoring and control system is the derivative mathematical model of BSC; it eventually became an integrated performance evaluation system. The results of the study lead to the development of an accurate and effective performance feedback system. Veen-Dirks and Wijn (2002) indicated that enterprises should focus more in the non-financial related performance indicators given the rapid change of circumstances.

Blackstone (2001) and Spencer (2000) mentioned TOC has been applied to the manufacturing sector widely, besides, it also been applied to non-manufacturing areas in recent years. For example, TOC specified why the organizations' performance will be affected by these constraints for services sector, therefore, it is very important organizations should overcome these constraints through continuous improvement.

Wasserman (1993) presented a mathematical justification for prioritizing design resources by ranking the index of technical importance to cost during the QFD planning process. Lee and Ko (2000) established a framework for developing and implementing a corporate business strategic plan. The framework has two steps. The first step is to conjoin the SWOT matrix with the BSC to construct a systematic and holistic strategic management system, and the second step uses the QFD method with the BSC and the main strategies of Sun Tzu. Lee *et al.* (2000) presented a framework for formulating a vocational education strategy by linking SWOT, BSC, QFD and MBNQA (Malcolm Baldrige National Quality Award). Their strategy formulation framework has four parts—SWOT analysis, linking SWOT with BSC, deploying all indices of MBNQA, and merging BSC with MBNQA using the QFD method. Chen *et al.* (2007) proposed a hybrid framework for service quality measurement, the framework is a combination of QFD and Kano model, identifying the key quality factors and weights to develop the House of Quality matrix.

For the applications of Benchmarking, Lobo and Zairi (1999a) presented Lufthansa Cargo AG (LGAC) is the best practice performance among nine airlines by quantitative Benchmarking analysis. In qualitative benchmarking analysis, LCAG recognized that customers can only decide everything, so the company must establish a mechanism to gather customer information. Besides, British Airway World Cargo (BAWC) has a regular customer listening forum and customer call center that deal with customer feedback on a daily basis; .Emirates Sky Cargo (EK) held professional staffs-ensure system that the customers are given the right advice at all time; Singapore Airline (SQ) emphasis that enterprises should work together with customers to meet customer demand; KLM divided the customer service department from sales and operation department; Cargolux recognized that only motivation, commitment, and service quality can meet the customer's needs; Cargo Express of Cathay Pacific is known for its efforts in easing the process for its customers; DHL sought to keep the leadership for the global customer service commitment in its staff statement; FedEx realized that airlines should continue to provide high quality and innovative services to customers (Lobo and Zairi, 1999b).

3. Methodology and Framework

Crosby (1979) proposed the concept of Zero Defects (ZD), he considered that all of the defects can be in advance to prevent. Motorola formally announced the plan of Six Sigma project in 1987, two years later, it obtained the highest honor of quality awards—MBNQA—in the United States. Six Sigma has been one of the emerging management methods to pursue the excellence of quality results for enterprises, it was stressed that the improvement capability of process variation in order to reduce costs and enhance business profitability. Furthermore, the DMAIC improvement cycle was one of the most commonly used tool for constructing the Six Sigma model.

Kwak and Anbari (2006) mentioned Six Sigma is more complicated than Total Quality Management and traditional continuous quality improvements, because Six Sigma can be deemed to combine five following methods, include 1. Total Quality Management or continuous quality improvements, 2. Focus on customer strongly, 3. Plural analysis tool, 4. Financial performance, 5. Project management. In addition, the key successful factors for enterprises to employee the Six Sigma include high-level managers' participation, organization commitment, cultural change and effective project management. Therefore, this study integrate Balance Scorecard into DMAIC of Six Sigma, in order to strengthen the participation of enterprise's high-level managers and link the mission, values, vision, organization commitment, and cultural revolution, as shown in Figure 1.

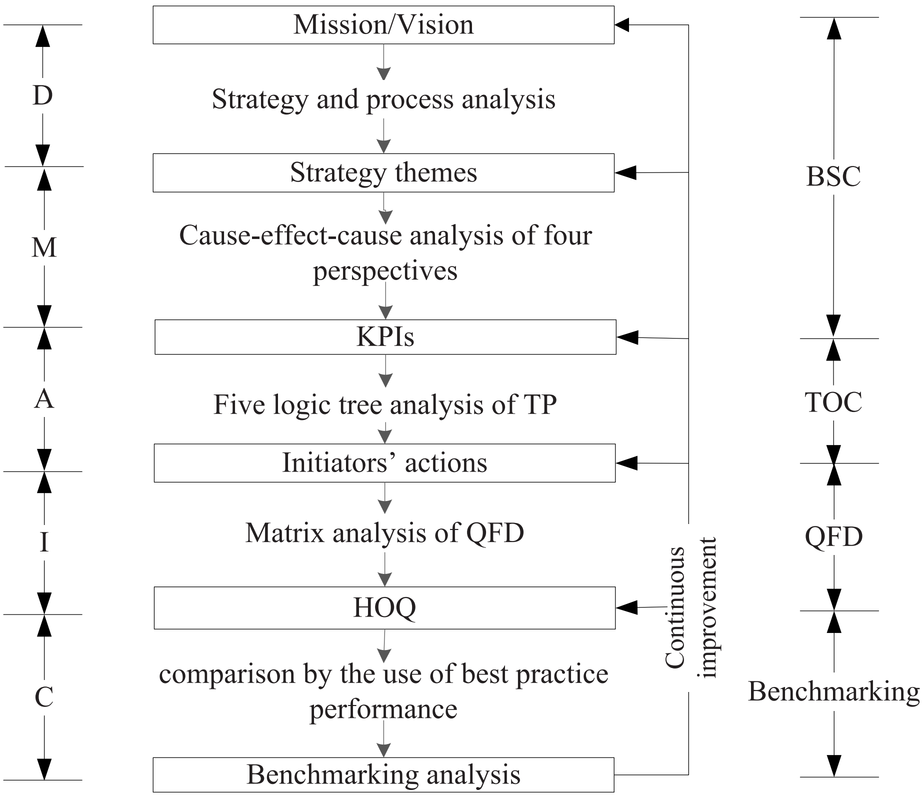


Figure 1: A Six Sigma based quality improvement framework

Balanced Scorecard (BSC) is proposed by Kaplan and Norton to integrate a company's mission, values, vision, and strategy into the four perspectives of BSC, which subsequently evolved into the company's performance targets and indicators (Kaplan and Norton, 1992). As airports activities include import, export, and transit of cargos as well as departure, arrival, and transit of passengers, the airports should focus on process management beyond organizational boundaries. Thus, there is a need to measure performance for the effective management of airports. Lai *et al*. (2002) stated that if one cannot measure a system, one cannot control the system; if one cannot control a system, one cannot manage the system; and if one cannot manage a system, one cannot improve the system. In fact, the lack of relevant performance measures has been recognized as one of the major problems in process management for airports. Accordingly, this study applied the Balance Scorecard as the first stage for developing performance measurements in order to establish the continuous quality improvement framework.

For the applications of Theory of Constraints (TOC), the bottlenecks will exist inevitably and influencing the performance all the time, if it is unable to find out its bottleneck operations and improves them for the organizations. At the same time, these bottlenecks is the constraints of the enterprises performance, it is a great factor to determine enterprises' operations quality too. Stein (1997) considers the existence of an enterprise is bound to some constraints, unless it can earn unlimited profits. These constraints can be divided into five types, namely, management constraints, behavioral constraints, capacity constraints, market constraints and logical constraints. Goldrtt in 1986 to propose TOC, he pointed out that the complexity of an organization not only by the impact of the number of constraints, but also decided to be these constraints relate to one another.

Moreover, in order to address the policy constraints and implement the process of on-going improvement, Goldratt (1992) develop a systematic approach—TP. This study deployed Thinking Process (TP) of TOC in the A phase of DMAIC, as for the complicated interfaces in air transport chain. Furthermore, there are two reasons for integrating TP into Six Sigma, first, enterprise should overcome some bottlenecks of process variations to reach four-sigma or five-sigma of process capability from three sigma of process capability, besides, it must have a total solution for bottleneck solving to achieve the six sigma of process capability. Secondly, the logic tree analysis of TP offers enterprises a total solution to solve the three questions via the five steps.

The five steps of logic tree analysis include Current Reality Tree (CRT), Evaporating Cloud (EC), Future Reality Tree (FRT), Pre-Requisite Tree (PRT), and Transition Tree (TT). Meanwhile, the three questions for the total solution to solve bottlenecks are "What to design?", "Design to what?", and "How to design?" Akao introduced Quality Function Deployment (QFD) in 1972 in Japan, as part of his work at the Mitsubishi Heavy Industries Kobe Shipyard to ensure that customer requirements are factored into invert aspect of the process. In recent years, the QFD has been widely used in a variety of problem areas and solution tools; meanwhile, Ozgener (2003) stated that the use of different purpose in the application of QFD matrix will be different when the changes in the problem areas.

Moreover, in the *I* phase, we integrate QFD with DMAIC in order to develop the actions priorities for quality improvement. There are three reasons for the integration, first, it is important to listen to voice of customer (VOC) for implementing the improvement project of Six Sigma, and deploy the VOC to all departments in the enterprise. Secondly, QFD is known as an improvement tool to enhance efficiency, quality, and competitiveness for enterprises. Third, is to ensure a team work for implementing the Six Sigma project. Thus, QFD provides a team work of problem solving tool for satisfy VOC matrix, evaluation matrix—competitiveness evaluation among the competitors for both VOC and technical requirements.

Benchmarking is a methodology of process comparison by the use of best practice performance as a benchmark to find out the self-improvements process for enterprises. There are some criteria for implementing the Benchmarking, e.g. select the process for the benchmarking, estimate the cost for implementing the Benchmarking, the selection and training for the Benchmarking team, develop the key indicators for the benchmarking, test the methodology of data analysis methods, analysis the flow by the key indicators, and collect the best practices for benchmarking.

4. Results and Discussions

Air Transport not only includes the complicated international interface, but operation failures will cause serious international influence and significant casualties. Therefore, Six Sigma to achieve the quality results perfectly is the most suitable to match the improvement tool of the air transportation industry. Thus, this study combined with the BSC, TOC, QFD, and Benchmarking into Six Sigma model for the process continuous improvement.

D (Define) represents the definition of customer needs, and find out the key process and key quality attributes. *M* (Measure) on behalf of the development of key indicators for the key process and quality attributes to calculate the values of each indicator. *A* (Analysis) expresses a large number of gathering information in order to facilitate analysis of a cause-effect relationship, and explore the root causes of variations. *I* (Improve) indicates the corresponding improvements to come up, and evaluate all the alternatives of improvement in order to propose the priorities to improve. *C* (Control) represents to monitor the results of improvements and feedback to improve the variations in order to ensure the sustainability of quality improvement.

Pande (2001) mentioned the company's mission/vision and the actions of Six Sigma project should be integrated closely for the full use of the function of Six Sigma. Along the same line, the enterprises' mission/vision can be transferred into the practical actions instead of unrealistic slogan by the high-level managers. Thus, this study integrated the Balanced Scorecard with Six Sigma in the *DM* phase of DMAIC in order to ensure the beginning of the Six Sigma project with the enterprise's mission closely linked.

In addition, there are two reasons of integrating Balanced Scorecard into Six Sigma, first, as enterprises need to invest substantial time and expenditure for implementing the Six Sigma project, while the Balanced Scorecard to emphasize the financial and non-financial perspectives of the balanced point of view can strengthen the weakness of financial perspective exactly for Six Sigma. Secondly, through the excellence of the balanced strategy analysis as well as the cause-effect-cause analysis and strategy map analysis via the four perspectives—financial perspective, customer perspective, internal process perspective, and learning and growth perspective—can be used to construct the key performance indicators (KPIs) for the Define stage of Six Sigma, the illustration for cargo terminal as shown in Table 2.

Table 2: Illustration of air cargo terminal for phase *D* of DMAIC

	Strategies	Strategic objectives	KPIs
Financial	1. Profitability 2. Cargo volume growth	Steady, continuous operation and profit growth.	1. Cargo volume growth 2. Space effectiveness 3. Market share 4. Net income 5. ROI
Customer	1. Customer loyalty 2. Customer satisfaction	Through the enhancement of customer satisfaction, improve customers' loyalty and cargo volume growth.	1. Service attitude 2. Aavailability of parking 3. Professionalism 4. Top ten customer volume 5. Customer retention
Internal process	1. Post sale process 2. Operation process 3. Innovative process	Improve customer satisfaction through the enhancement of innovation, operating efficiency, security and after sales service flow.	1. Non-general cargo service 2. Degree of automation 3. MHS 4. Density of storage position 5. Process efficiency 6. Cargo safety 7. Accuracy of invoices 8. Compensation for cargo missing/damaged
Learning & growth	1. Employee productivity 2. Employee turnover rate 3. Organizational climate 4. IT infrastructure 5. Employee competence	Through the learning and growth of employee, and company to reduce the employee turnover rate, increase the productivity, and then improve the internal process.	1. Employee productivity 2. Employee turnover rate 3. Organizational climate 4. IT infrastructure 5. Employee competence

General speaking, there are two kinds of organization's constraints which is physical constraints and policy constraints. There are 3 main physical constraints of an organization. Number one, vendors' constraints, most of the companies complain about the vendors, but this normally won't be a real constraint. For example, a company might face shortage of material supply, the root cause might be the ineffective of its purchase or inventory strategy, not caused by vendor – except for when vendor face a continuous lack of supply of material (normally it happens to the whole industry). Resource constraints are the second constraint, and market constraints are the third constraint (Goldratt, 1992). We developed the initiator's actions by the TP for the *A* phase from the results of *DM* phase for the illustration, as shown in Figure 2.

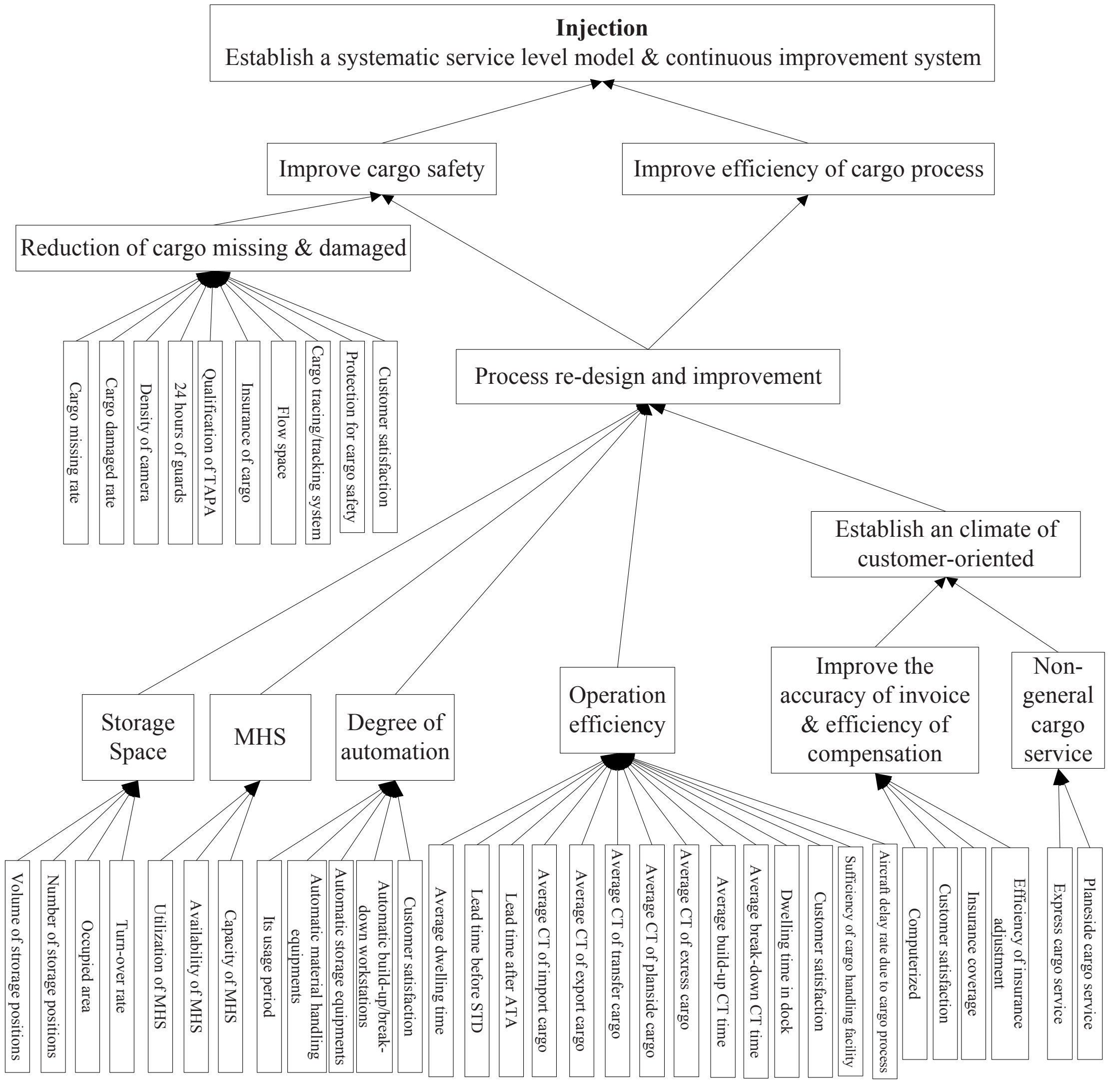


Figure 2: Initiators' actions for phase *A* of DMAIC

The trend of customer-oriented, complicated and changeable environment, QFD has been an essential methodology for the business to enhance on-going improvement. Therefore, we apply QFD as an effective instrument to implement the initiator's actions and evaluate the performance of competitors. Additional, QFD is a well-structured tool; HOQ is the core of QFD. Generally speaking, HOQ consists of nine matrices, they are customer requirements matrix, technical features matrix, relationship matrix, importance matrix, inter-correlation matrix, competitive evaluation matrix, technical importance matrix, technical evaluation matrix, and target values matrix. Customer requirements matrix, design requirements matrix, and relationship matrix are also known as main matrix of HOQ. Figure 3 presented part of the HOQ via the *I* phase from the phase *DM* and phase *A* for the illustration. At the same time, this study integrated Benchmarking with the *C* phase of DMAIC by Best Practice Performance Matrix of HOQ, the purpose is not only monitoring and predicting the results of improvements, but also correcting the gap for the continuous improvement.

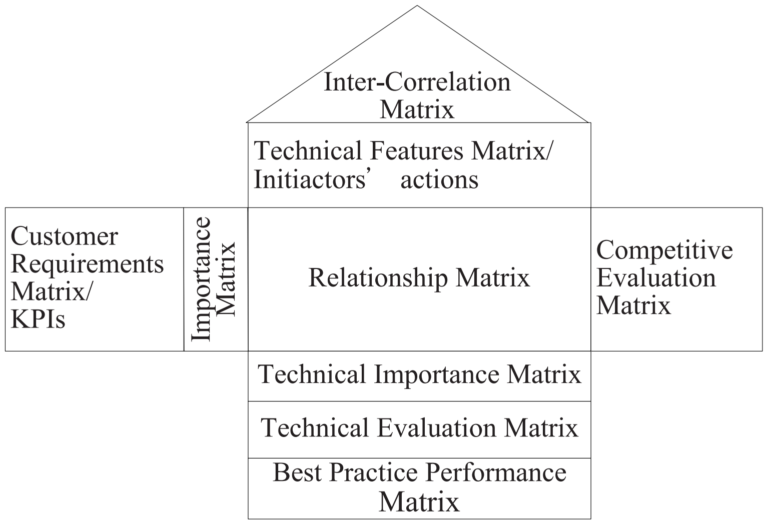


Figure 3: HOQ for phase *IC* of DMAIC

5. Conclusions

The airport is one of the most main facades for nations, and the aviation transport is also shown as the national development. However, the performance results of TTIA growth slid has had negative influence to the national competitiveness in recent years. According to statistical data of IATA and CAA, the accident rate was about 1.74 per one million flight hours, approximately three times more than global average 0.56 from 1993 to 2006 in Taiwan. Hence, this research aims at the TTIA to build a continuous improvement framework via Six Sigma for process quality.

Even though the science and technology of aviation industry is highly developed today, air accidents still occasionally happened, reasons including human factor, climatic factor and mechanical factor. Meanwhile, the proportion of accidents caused by human factor is the highest. This research developed an integrated framework base on DMAIC model of Six Sigma combining the Balance Scorecard, Theory of Constraint, Quality Function Deployment, and benchmarking management to achieve the ZD for airports.

Hong Kong Airport has won many times by Skytrax electing the most outstanding airport in the world since 2000. Hong Kong Airport also selected by Air Cargo News as the most outstanding cargo airport as well as the "Most Convenient Cargo" and "Most Friendly Airport for Cargo" airport by the Federation of Asia Pacific Air cargo Associations (FAPAA). Thus, this framework integrated Benchmarking into the C phase of DMAIC model. We hope to set up a benchmarking for our aviation industry, and strengthen practice of application and the contribution. Finally, this framework not only can improve the operation results of air terminals but also can provide the authorities to make decisions for process quality improvement of air transportation.

6. Acknowledgments

The author would like to thank the National Science Council of Taiwan for financially supporting this research under Project No. NSC97-2410-H-431-005.

References

Blackstone, Jr J. H. (2001), Theory of constraints—a status report, International Journal of Production Research, 39(6): 1053—1080.

Chen,C. H., Chang, H. H. and Tsou, C. S. (2007), A Hybrid Method for Measuring Service Quality: An Empirical Investment on Taiwan's KTV, International Conference on Business and Information, Tokyo, 1—11.

Chen, J. and Chen, J. C. (2001), QFD-based technical textbook evaluation—procedure and a case study," Journal of Industrial Technology, 18(1), http://www.nait.org/jit/Articles/chen110701.pdf, last accessed in December 2008.

Civil Aeronautics Administration (CAA), Taiwan, Civil Air Transportation Statistics**,** Retrieved December 26, 2008, from http://www.caa.gov.tw/big5/content/index.asp?sno=186

Civil Aeronautics Administration (CAA), Taiwan, Aviation Safety Bulletin**,** Retrieved December 26, 2008, from http://www.caa.gov.tw/en/content/index.asp?sno=67

Crosby, P. B. (1979), Quality Is Free, Mcgraw-Hil1: NY.
Evans, J. R. and Dean, Jr. J. W. (2003), Total quality: Management, organization, and strategy, 3rd ed., South-Western: OH.

Goldratt, E. M. and Fox, R. E. (1986), The Race, North River Press, Great Barrington: MA.

Goldratt, E. M. (1987), Essays on the theory of constraints, North River Press, Great Barrington: MA.

Goldratt, E. M. and Cox, J. (1992), The Goal: A Process of Ongoing Improvement, second revised edition, North River Press, Great Barrington: MA.

Heathington, K. W. and Jones, D. H. (1975), Identification of levels of service and capacity of airport landside elements, Transportation Research Board, Special Report 159: 72—91.

International Air Transport Association (IATA), Aviation Safety, Retrieved October 12, 2008, from http://www.iata.org/whatwedo/safety_security/safety/safety_data.htm

Kanplan, R. S. and Norton, D. P. (1992), The balanced scorecard: measures that drive performance," Harvard business review 70(1): 71—79.

Klein, D. J. and Debruine, M. (1995), A thinking process for establishing management policies," Review of Business, 16(3): 31—37.

Kwak, Y. H. and Anbari, F. T. (2006), Benefits, Obstacles, and Future of Six Sigma Approach," Technovation, 26: 708—715.

Lai, K. H., Ngai, E. W. T. and Cheng, T. C. E. (2002), measures for evaluating supply chain performance in transport logistics," Transportation research part E, 38: 439—456.

Lee, S. F. and Ko, A. S. O. (2000), Building balanced scorecard with SWOT analysis, and implementing, Sun Tzu's The Art of Business Management Strategies on QFD methodology, Managerial Auditing Journal, 15(1/2): 68—77.

Lee, S. F., Lo, K. K., Leung, R. F. and Ko, A. S. O. (2000), Strategy formulation framework for vocational education: integrating SWOT analysis, balanced scorecard, QFD methodology and MBNQA education criteria, Managerial Auditing Journal, 15(8): 407—414.

Lobo, I. and Zairi, M. (1999), competitive benchmarking in the air cargo industry: Part I, Benchmarking an International Journal, 6(2): 164—190.

Lobo, I. and Zairi, M. (1999), competitive benchmarking in the air cargo industry: Part II, Benchmarking an International Journal, 6(3): 220—238.

Özgener, S. (2003), Quality function deployment: a teamwork approach, TQM & Business excellence, 14(9): 969—979.

Pande, P. S., Neuman, R. P., and Cavanagh, R.R. (2001), The Six Sigma Way: how GE, Motorola, and other top companies honing their performance, McGraw-Hill: NY.

Rahman, S. (1998), Theory of constraints: a review of the philosophy and its applications, International Journal of Operations and Production Management, 18(4): 336—355

Rouse, P., Putterill, M. and Ryan, D. (2002), Integrated performance measurement design: insights from an application in aircraft maintenance, Management accounting research, 13(2): 229—248.

Seneviratne, P. N. and Martel, N. (1991), criteria for evaluating quality of service in air terminals, Transportation Research Record, 1461: 24—30.

Sokovic, M., Pavletic, D. and Fakin, S. (2005), Application of Six Sigma Methodology for Process Design, Journal of Material Processing Technology, 162-163: 777—783.

Spencer, M. S. (2000), Theory of Constraints in a Service Application: the Swine Graphics Case, International Journal of Production Research, 38(5): 1101—1108.

Stein, R. E. (1997), The theory of constraints: Applications in quality and manufacturing, 2nd ed., Marcel Dekker: NY.

Veen-Dirks, P. V. and Wijn, M. (2002), Strategic control: meshing critical success factors with the balanced scorecard, Long rang planning, 35: 407—427.

Wasserman, G. S. (1993), On how to prioritize design requirements during the QFD planning process," IIE Transactions, 25(3): 59—65.

Wilson, H. W. (2003), A balanced scorecard framework for assessing LTL motor carrier quality performance, Transportation quarterly, 57(3): 105—132.

Analysis on port-city growths in China

Shuk Man Sherman Cheung[1] and *Tsz Leung Yip*[1]

[1] Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hong Kong.

Abstract

This paper aims to model the economic production of port-cities and their evolution by considering the port production versus city factors. The paper uses the economic production equation to determine the growth of China's port cities over the period 1995-2005. There have been numerous studies examining general port growth and development on the basis of conceptual discussion. It is generally accepted that the Anyport Model and the Port Generation Model are considered the valuable models for a port developing into a port city. However, quantitative studies on port growth and factors that influence the growth rate are rather few. This study seeks to identify the factors that are critical to determine port city growth and attempts to develop an empirical model in the macroeconomic perspective. Seven major China port-cities and their interface with the reliance-ports are studied in the paper. This paper contributes to literature by identifying the port activities as economic production and the city factors as production inputs. This paper provides a quantitative analysis of port growth and policy insights on the interdependence between port and port-city.

Keywords: Port city, Anyport model, Cobb-Douglas production function

1. Introduction

Globalisation is a trend in the business world. Businesses are all looking globally for effective logistics solution. In the shipping and port industry, logistics integration is reinforced by globalisation. Forming networks with neighbour ports or having strategic alliance with remote destination ports are usual practices of scaled ports nowadays. The ports can enhance efficiency, strengthen competitiveness and achieve synergy effects and economies of scale. The role of ports is broadened and elaborated from a simple sea-shore interface to a comprehensive logistics centre. It is developed from an isolated facility to a community asset. The port and city are then inter-related and inter-dependence. A port facilitates the growth of its city and regional economy, whereas the growth of city pushes the development and evolution of its port. The city and port issues may be complementary as well as contradictory. On one hand, successful ports around the world are often city-ports. On the other hand, the port-city interface is one of the major determinants of the successfulness of modern ports, and therefore the formation of major port-cities.

Both port and city tend to focus on global outlook, logistics integration and transportation network. Under globalisation, there exists a close relationship between port and city developments, city-ports (and so as port-cities) serve the global markets with integrated logistics and supply chain flows. In modern shipping and port industry, in order to facilitate efficiency and trades, globalisation encourages standardisation in shipments and cargo handling. Containerisation is resulted. Containerisation advances the port facilities and develops cities. Major city-ports are having advanced container terminals. The port-cities are also having integrated logistical facilities to support the port operations and the city activities.

China is a continuously growing country in terms of economy and world position. The world is now looking into China for opportunities. With the China Open-Door Policy in 1978 and China joining into the World Trade Organisation (WTO) in 2000, trades play a crucial role to the growth and development of ports and coastal cities, as well as economies in China. In addition to the long coastal line, undoubtedly, the development of sea-bound international trades in China has a bright future.

From the geographical perspective, a port-city can be classified into two spatial scales. At a local scale, a port-city is the "area in transition" where combining port and urban jurisdiction and functions (Hoyle, 1989). At a wider scale, assuming land-sea connexion, a port-city is the nodal system as a whole within a regional area, consisting multiple cities and ports (Ducruet and Jeong, 2005). From the economic perspective, port-city is a city with port functions, which the city functions and port functions are keeping a balanced combination of centrality and intermediacy (Ducruet, 2005). A port-city is a product of the integration of port and city. Not only can it reflect the common characteristics of a general city, but also have its specific contents and rules of movement (Ducruet and Lee, 2006).

This paper aims to explore the economic function of port-cities and their evolution by considering the port production versus various factors of port-cities.

2. Literature Review

There are two hypotheses on the evolution of port and city: port-led growth and city-led growth. From the port-led hypothesis, the port development is described in the Anyport Model with 6 stages (Bird, 1963). Starting from the initial port site with small lateral quays adjacent to the town centre, the elaboration of wharfs is the product of evolving maritime technologies and improvements in cargo handling. This is also marked by changing spatial relationships between the port and the urban core, as docks are built further away from the central business district. In the final stages, increased specialization of cargo handling, growing sizes of ships, and ever increasing demands for space for cargo-handling and storage resulted in port activity being concentrated at sites far removed from the oldest facilities.

The Anyport Model implicates the changing relation between ports and their host cities. It describes the growing repulsion by the rest of the urban milieu. This aspect has been worked upon over the last two decades by a number of geographers investigating the redevelopment of harbour land. Hoyle (1989) proposed the Anyport-City Model to emphasize the changing linkages between the port and the city, instead of stressing the port infrastructure development. One of these urban linkages is the redevelopment of old port sites for other urban uses.

3. The Port-City Growth Model

Rather than remaining static, a port-city may be constantly growing over time. The port growth can be quantified by the port's activities. The port production should be the most remarkable factor of port activities. We consider a port-city as an organisation, the port production as the output of the organisation and the city factors as factors of production. In other words, the port factors (throughput, turnover or total trade value) are treated as the port productivity which is the combined result of city factors of production (Eq. 1). For the overall productivity of a port, it can be represented by Cobb-Douglas production function (Eq. 2).

$$Production = \mathrm{f}\,(Land, Capital, Labour) \tag{1}$$

$$Y = A\, L^{\alpha}\, K^{\beta}\, D^{\gamma} \tag{2}$$

where: Y = total port production (the monetary value of production in one year), L = Labour input, K = capital input, D = Land input, A = Constant, α, β and γ = constant output elasticities of labour, capital and land, respectively.

Output elasticity measures the responsiveness of output to a change in levels of either labour or capital used in production. For example if $\alpha = 0.15$, a 1% increase in labour would lead to approximately a 0.15% increase in output. Applying to the port-city, for example, '*Land*' can be quantified by the '*Infrastructure*', '*Capital*' by '*Economic*' factors and '*Labour*' by '*Demographic*' factors. The city factors affect the port factors and facilitate the port development, and vice versa. The port production is therefore the combination result of all the city factors. Presenting the factors by function formula, Eq. (3), port production is the dependent variable, whereas economic, demographic, fixed assets and FDI are independents in the given years. For easy comparison and statistic calculation, the factors are turned into logarithm, as Eq. (4).

$$\textit{Port Production} = f_1(\textit{Economic}, \textit{Demographic}, \textit{Infrastructure}) \quad (3)$$

$$\ln \textit{Port Production} = f_2(\ln \textit{Economic}, \ln \textit{Demographic}, \ln \textit{Infrastructure}) \quad (4)$$

Where
Port Production = *Throughput*, or *Total trade value*, or *Turnover*,
Economic = g_1 (*GDP*, *GDP per capita*),
Demographic = g_2 (*Population*, *Total employment*, *Employment in tertiary sector*),
Infrastructure = g_3 (*Fixed assets investment*, *Contracted foreign direct investment*), and
ln = Log of the figure of the year,

Combining port production and factors of port production, it has Eq. (5).

$$\ln Y = \ln A + \alpha \ln L + \beta \ln K + \gamma \ln D \quad (5)$$

where $\ln Y$ is the factor quantifying the port growth rate.

4. Data and Analysis

Table 1 summarises the data collected from China Statistics Yearbooks (1995 to 2007 data). The seven ports are the leading ports in China. The data for the port and the city where the port is located in are then collated.

Table 1: Data on Port and City Factors from 1995 – 2007

	Production			Economic		Demographic			Infrastructure	
	Port Throughput 10k tons	Port Freight 100m ton-km	Trade value Total 100m USD	GDP 100m USD	GDP per Capita	Population 10k persons	Total Employment 10k persons	Employment in tertiary 10k persons	Investment in Fixed Asset 100m Yuan	Contracted FDI 100m USD
Dalian	6,417 – 17,085	–	133 – 256	733 – 2,152	13,676 – 42,579	537 – 572	85 – 280	43 – 102	115 – 1,469	24 – 46
Tianjin	5,787 – 24,069	73 - 12108	65 – 716	918 – 5,018	10,281 – 45,829	942 – 1,115	403 – 614	159 – 276	393 – 2,389	35 – 73
Qingdao	5,103 – 18,678	–	88 – 330	710 – 2,696	10,331 – 38,892	690 – 749	103 – 393	42 – 254	79 – 1,486	7 – 80
Shanghai	16,388 – 44,317	117 – 13,695	190 – 2,830	2,463 – 12,001	18,943 – 57,695	1,415 – 1,858	670 – 906	284 – 503	1,602 – 4,459	15 – 149
Ningbo	6,853 – 26,881	–	–	796 – 2,449	15,069 – 51,460	530 – 560	57 – 414	30 – 96	146 – 1,503	5 – 41
Xiamen	1,314 – 4,771	–	76 – 286	308 – 1,007	25,248 – 50,130	123 – 160	94 – 96	15 – 32	79 – 648	7 – 19
Shenzhen	5,697 – 15,351	–	639 – 1,828	950 – 4,951	46,388 – 115,060	103 – 201	93 – 272	41 – 100	186 – 1,287	12 – 48

Source: China Data Online

The findings are generated by SPSS correlation and regression analysis programme. They consist of analysis of the port growth rate and corresponding city factors from 1995 to 2007. The correlation test is used to test the correlation between port and city factors. Port factors (i.e. annual throughput, port turnover and total trade value) are tested with city factors one by one. From Table 2, the correlations among port factors and city factors can be statistically supported. Based on the significance of the city factors to port factors, the model specification of Cobb-Douglas production function is then tested.

Table 2: Correlation Coefficient (R) of the Growth in Port and City Factors

Growth		Port Productivity		
		Throughput	Freight	Trade
Economic	GDP	0.912 **	0.563 **	0.862 **
	GDP per Capita	0.310 *	0.628 *	0.819 *
Demographic	Population	0.738 **	0.348	0.171
	Total employment	0.503 **	0.228	0.377 *
	Employment in tertiary sector	0.610 **	0.294	0.430 **
Infrastructure	Fixed Asset Investment	0.819 **	0.506 *	0.775 **
	Contracted FDI	0.760 **	0.386	0.648 **

** Correlation is statistically significant if $p < 0.01$ (2-tailed)
* Correlation is marginally significant if $p < 0.05$ (2-tailed)

Table 3 summarizes the regression outputs of the port production and city factors, where each of 3 models uses one single port factor as port production. The *R*-squares are 0.971, 0.690 and 0.983 indicate the high explaining power of the proposed models. Therefore the city factors can be used to predict the port production. Regarding the level of significance, the factors are highly significant when *p* is smaller than 0.01 (**) and are moderately significant when *p* is smaller than 0.05 (*). If the significant level is larger than 0.05, the independent factor will not be considered significant to the dependent factor. As the three models show similar regression results, the finding is robust and independent of different ways to define the port production. However, there are some discrepancies of variable signs among the three models.

Table 3: Significant (p) of the Growth in Port Throughput and City Factors

Growth		Port Throughput		
		Throughput	Freight	Trade
Economic	GDP	1.282**	16.652	1.731**
	GDP per Capita	–0.301*	–1.617	0.419**
Demographic	Population	0.365**	–7.260	–0.635**
	Total employment	–0.427**	15.786*	0.531**
	Employment in tertiary sector	0.039	–12.312	–0.304
Infrastructure	Fixed Asset Investment	–0.179*	–9.463*	–0.541**
	Contracted FDI	–0.075	–2.747*	0.112*
	Constant	3.939*	–11.054	–6.110**
	R^2	0.971	0.690	0.983

** Correlation is statistically significant if $p < 0.01$ (2-tailed)
* Correlation is marginally significant if $p < 0.05$ (2-tailed)

From Table 3, economic and demographic factors are highly significant to port throughput in general. In the 'throughput' model, the employment in tertiary sector and the contracted FDI are not significant to the port throughput. Due to the limited number of observations of port freight, the 'freight' model is not statistically satisfactory, in which only the total employment, fixed asset investment and the contracted FDI are significant in the model. The 'trade' model shows the most satisfactory results, and only the employment in tertiary sector is not significant to the total trade value.

The significance of economic and demographic factors to port productivity is understandable. In a port-city, the port and city are inter-reliance in which the economic factors will affect the capital input to port and the demographic factors would affect the labour input. Nevertheless, apart from the total trade value, the significance of fixed assets investment and contracted FDI are not obvious.

It is believed that the growth in fixed assets investment and contracted FDI are getting along with the port productivity of a port-city. The low significance of fixed assets investment and contracted FDI may be attributed to without considering the time lags in generating the regression analysis of economic production model. Investment on fixed assets such as superstructure in port and infrastructure of the city take a long time to complete. The investment decision may be made on the basis of performance of previous years. Similar to investment in fixed assets, the contracted FDI of a year may contribute the port production in the later years. The effect and significance of fixed assets investment and contracted FDI to port production cannot be immediately represented well by the annual data. Moreover, the contracted FDI is usually investing in businesses such as commercial bank and manufacturer, which are not directly related to port operation. The contracted FDI is much belonging to the capital input of a city. Therefore, the significance of contracted FDI to port productivity is also very low.

5. Implications

The first author further conducted site visits in the first half of year 2008, in addition to statistical analysis. The port-city interface of the national port-cities is revealed. The investigation in the port-city interface answers the 2 fundamental questions: (1) how port facilitates city functions; and (2) how city favours port functions.

5.1. How Port Facilitates City Functions

Port functions, from a primitive port to a sophisticated port, are direct cargo flow, simple trade, cargo handling, transhipment, warehousing, consolidation and information flow and other high value adding services. In the settings, the port factors are combined of annual throughput, port turnover and total trade value. At each enhancement and development of the port, the reliance-city is benefited.

When the port's productivity increases with more than simple cargo flow and handling, e.g. transhipment, early medium scale trading companies established around the port to facilitate shippers. These trading companies can improve the efficiency of cargo handling, and therefore rise the turnover of port. In the meanwhile, around the trading companies, which are sources of goods, markets and fairs agglomerate and formal commercial activities like banking and financing began. With commercial activities in the markets, traders and sellers require storage space for incoming or outgoing cargoes, therefore using the port as the warehouse.

From time to time, when trading is boosting around the port and the economy is rising in the city, port throughput can satisfy the growing demand of goods and different commodities. To further facilitate economic activities in the city, infrastructure such as highways and railways are constructed to transport the cargo from the port to the city. Apart from cargo, infrastructure also brings population into the city from outside.

Meanwhile, the consolidation and value added services impose enforcement onto traders to stick with the port. Having a fully functioned port and well established commercial activities and infrastructure, investment can be attracted from the region nearby and overseas. In a port-city, a productive port with whole package of services is beneficial and crucial to the development of city functions. The port-city interface helps the economic growth and demographic upgrading of the city.

5.2. How City Favours Port Functions

City functions, from a little village to a big city, are for better life of people who living in, education, commercial activities, manufacturing and political reasons. In the settings of the previous chapter, city factors of a port-city are economic factors (GDP and GDP per capita), demographic factors (Total employment and employment in the tertiary sector), investment in fixed assets and contracted foreign direct investment. At each enhancement of the city, the reliance-port is benefited. The combination of city factors is also the factors of port productivity.

When markets and fairs start to agglomerate around the port, the cargo at ports can be distributed quickly. The berthing time at port can be shortened and more trade turns in. This can increase the port turnover. When the commercial activities boost in the city, more goods from the incoming vessel is demanded. Having more port calls can boost the port throughput and numbers of direct shipments and transhipments. In the meanwhile, the rise in commercial activities enhance the economic conditions of the city, more trades are shipped through the port and therefore boost the total trade value in the port.

Moreover, to facilitate trade, infrastructure such as railways, highways and bridges are built. With these infrastructures, the connectivity between the port and the outer areas of the city can be achieved. The linkage between the port and the entire city and outside brings population to the city. The population can employed in the port-related value added and supporting logistics services, which is tertiary sector. These services make the port more competitive with more integrated port functions.

Once the port productivity is continuous to rise, the performance of the port becomes advanced and has attracted more capital from the city for port enhancement and expansion. The port can therefore become sophisticated and world ranked. In a port-city, the matured city development is essential and critical to port functions. The port-city interface helps enhancement and specialisation of the port to meet the world pace.

6. Conclusions

An economic production model is developed to study the growth of port-city. The economic production of ports is specified by Cobb-Douglas production function. Regression analysis is applied on port production and city factors by assuming a port-city as a production organisation. It is identified that the city economy is the most significant contributor of port production growth, but there are no consistent effects of the infrastructure and demographic factors. The study offers statistical evidences to the city-led evolution of port-cities and provides a solid foundation to derive economic policies of port development.

Previous conceptual discussion and qualitative analysis have been found very fruitful to understand the mechanisms of port-city evolution. This paper should have contributed to the issues of port-city development model, when the quantitative analysis are desired to determine the port activities against city inputs. However, further research is needed to compare the port growth models across the world and to explore the explaining power of various model specifications. This study should be extended to consider some other city factors and some other modes of port-city evolution. This is important because future port-city planning should include the port management.

7. Acknowledgements

The first author would like to thank, Dr. Tang Rong-Ming, from the Faculty of Business of the Zhejiang University, for giving her valuable knowledge in the China port-cities during her exchange study in the Zhejiang University in 2008. The study is supported by the Departmental Research Grant Number G-U371 Port City Competition.

References

Bird, J. H. (1963). *The Major Seaports of the United Kingdom*. London: Hutchison of London.

Ducruet, C. and Jeong, O. (2005). *European Port-city Interface and its Asian Application*. Research Report 2005-27, Korea Research Institute for Human Settlements, South Korea.

Ducruet, C. (2005). A metageography of port-city relationships. In: Wang, J., Olivier, D., Notteboom, T. and Slack, B. (eds), *Ports, Cities, and Global Supply Chains*, Ashgate: Aldershot, pp. 157-172.

Ducruet, C. and Lee, S.W. (2006). Frontline soldiers of globalisation: Port-city evolution and regional competition. *GeoJournal*, 67(2), 107-122.

Hoyle, B.S. (1989). The port-city interface: Trends, problems and examples. *Geoforum,* 20(4), 429-435.

Satisficing strategy in determining port location in Viet Nam

Ngoc-Hien Do[1], Quynh-Lam Ngoc Le[1], Ki-Chan Nam[2], and Kyu-Seok Kwak[2]

[1] Graduate School, Department of Logistics, Korea Maritime University, Busan, Korea
[2] Department of Logistics, Korea Maritime University, Busan, Korea

Abstract

The role of ports has been changed over decades because of development of technology, economy, and so on. Therefore, requirements or criteria to determine the port location have been varied. Actually, some of them conflict with others, so it is hard to choose the perfect port location. It satisfies some criteria but sacrifices other ones. The compromised alternative has been usually suggested. This strategy is called satisficing one, which was introduced by Simon (1957). In this paper, authors propose to use it in determination of port location. Displacement of ports from Ho Chi Minh City to Vung Tau province, Vietnam, is considered as a case study. The results of this paper could be referred as a reference when making decisions relative to port location and/or relocation.

Keywords: Satisficing strategy, Ports location, Vietnamese seaports

1. Introduction

The role of seaports increases simultaneously with the development of Vietnam's economy, which boosts the trade flows. Seaports systems have had attention and investment from government since the 1990s. At present, Vietnam has approximately 160 seaports, 305 wharves with total of 36164 km long and loading capacity of 177.58 million tons in 2007.

In general, they mainly serve economic centers, where Ho Chi Minh City is known as the biggest and most dynamic one in Vietnam, which handles about 75% of the total container volume. Most of seaports are city ports, which are situated in urban and semi-urban areas. They are classified as small ports able to serve ships up to 30.000 DWTs, and most of them slack in up-to-date.

Nowadays, the increase in the size of ship and far greater traffic volumes require demands on seaports to upgrade, because exiting ones are unable to adequately accommodate larger ships and additional traffic. New seaports have to meet demands of the country's industrialization and modernization process, meet the entire flow of sea-born import-export cargoes in the future, as well as take advantages of nature and shipping potentials to develop ports for better serving economic areas, industrial zones.

Many locations in Vietnam are suitable for building deep-sea ports, such as the 14 meter depth of Cai Mep area in the Southern, 17 meter depth of Da Nang, and 25 meter depth of Van Phong areas in Center, and 14 meter depth in the Northern. Once built at these areas, ports can accommodate vessels up to 15,000 TEUs (Van Phong area), and also expanse Vietnamese ports' capacity.

A combination of resources including supportive government policies, ample investment, and well thought out operations and information technology along with location and a natural deep harbor to help create a sustainable advantage for the port determines a high quality seaport system. Port of Singapore is achieved a sustainable competitive advantage relative to its location, which has a good natural with a superb sheltered harbor and a significant costs including infrastructure, well educated, and hard working labor forces (Gordon et al., 2005).

A specific research methodology designed to evaluate and compare competing seaport locations suggested by Baird (2006), which focused on container transshipment hub locations in northern Europe. The paper proposed to develop new transshipment terminal capacity that meets the future expected traffic flows. Transport distance and associated shipping costs and the vast natural deep-water harbor were used to evaluate alternatives. Following Malchow and Kanafani (2003), the most significant characteristic of a port is its location. Although landside expansions are a cornerstone of a strategic port planning, inland infrastructure/superstructure development plans are usually designed to satisfy the needs of shipping services, which meet the requirements of sea transport (UNCTAD, 2004).

Port location issue should be seriously considered. It is one of key factors affecting successfulness of seaports in the future. There are many criteria used to determine a ports location, so it is hard to find a perfect location that satisfies all criteria at the highest levels. This paper shows the way to determine a good seaports location that meets requirements, and a case study is considered as an illustration.

To evaluate alternatives many criteria are usually applied. However, it is not assured that the selected alternative is perfect in terms of satisfying all criteria at the highest levels. It actually both satisfies and sacrifices criteria in some aspects. The notion of "satisfice" was first introduced by Herbert Simon (1957), which is defined as "to be good enough for the real world although not perfect". It is applied in many decision making issues. For example, Tyson (2008) suggested a rationale for satisficing behavior based on two themes of Simon's (1957) work: cognitive constraints and environmental complexity (2007). It is also used in economic as determinants of the quality decision making in economic organizations with a significant influence (Minkes and Foxall, 2003). In this paper, the satisficing strategy is used to determine the seaport location in Vietnam.

2. Seaport Location Planning in Vietnam

The mainly function of seaports systems in Vietnam is to facilitate trade of economic centers. Therefore, major seaports in Vietnam are city-ports. They locate near economic centers such as Hochiminh City, Haiphong City and Danang City, so short distance from seaports to economic areas is their advantage. Besides, they have basic existing infrastructure/superstructure, good labors and so on. However, they are small seaports so that they cannot receive calls from big ships. Almost all are general ports, out-of date, and operate at low efficiency.

Vietnam's impressive and consistent growth over the last several years (second in Asia to China) has made Vietnam more and more visible on the global map, especially for those multinational corporations looking for an outsourcing and factory-relocating destination. While exports from Vietnam to North America have increased over last five years, its export market is still largely dominated by Intra-Asian trade. The import/export is shown in Figure 1.

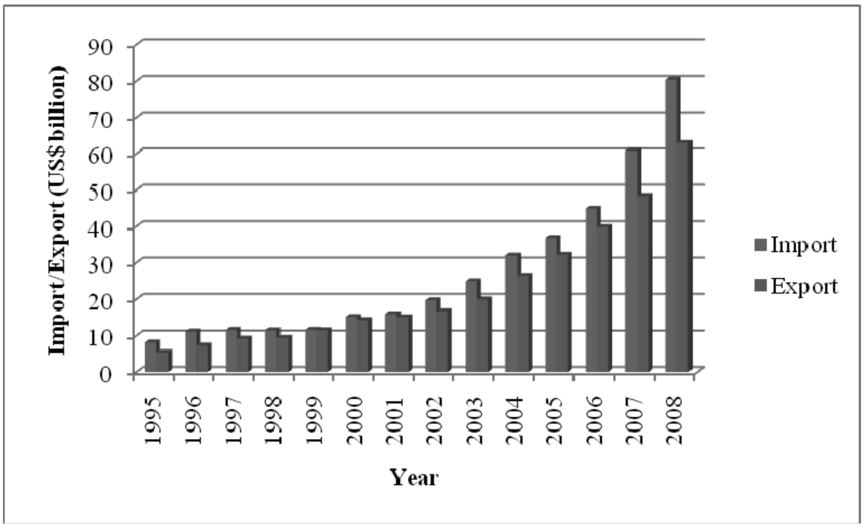


Figure 1: Import/Export of Vietnam 1995 - 2008 (US$ billions)
Source: Vietnam General Statistics Office

Beside price, labor force, Vietnam also attracts investors by a safety investment environment with transparent and clear political policies. Becoming 150th member of WTO is an opportunity for Vietnam in general and Vietnam port system in particularly. As the result of economic development, more import/export will go through ports. The cargo volume via seaports increases year by year as shown in Figure 2. Following the report about congestion in seaports of Vietnam, demands of cargo transport via seaports at Ho Chi Minh City area have overloaded in comparison with their capacity from 2007, so congestion problem in seaports should be considered.

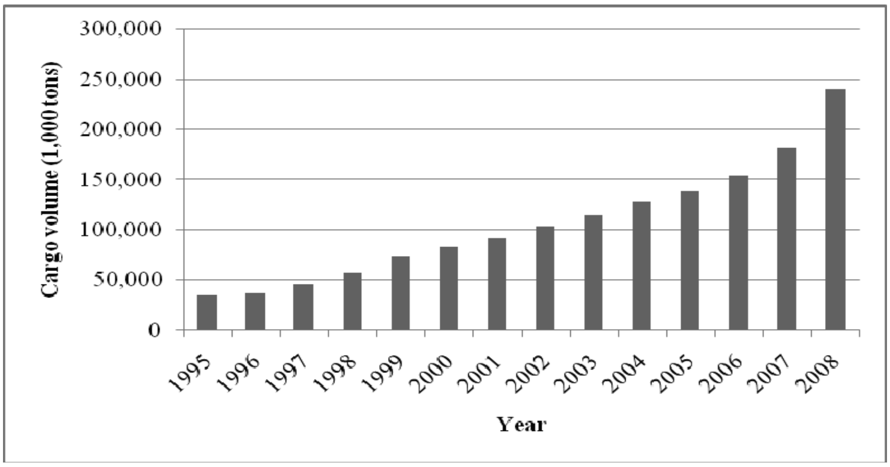


Figure 2: Cargo volume via Vietnamese seaports (in thousand Tons)
Source: Vietnam Seaport Association (VPA) and VINAMARINE

Therefore, a new stage of developing seaports system in Vietnam should be considered. It requires lower transports costs and extendible global reach of megacarriers, so at least seaports locate on deep-water that could serve big ships above 80.000 DWTs. Major respects of new seaports system include facilitating development of adequate port capacity, maritime access and hinterland connection to allow ports to fulfill their role, fostering the provision of competitive and efficient services in ports, and stimulating the wider community responsibility of ports. They need to make a closer co-operation instead of more competition each other.

Some requirements for developing new seaports system involve attentions. A clear, effective and transparent financial framework should be established; Relations with port service providers should be managed; Operational bottlenecks that hamper port efficiency need to be solved such as infrastructure capacity, technical-nautical services provided, cargo handing service and port labor, administrative bureaucracy, controls and inspections, and inefficiencies in hinterland transport; The sustainable development of port and port related capacity should be ensured about environmental legislation, positive contribution of ports, and stronger focus of resources on hinterland connections.

There are many requirements, challenges and opportunities for developing new seaports system in Vietnam. One of the key factors making new seaports success is their location that is good enough to meet some expected criteria. It is an essential condition to assure they could achieve expected goals. They could be developed from existing seaports, city ports, or they could be replaced by or moved to a suitable location. Besides, Vietnam has many areas that could be used for developing new seaports as shown in Figure 3 with 8 groups of port locations. Therefore, determining the best place for a new seaport system is an essential consideration.

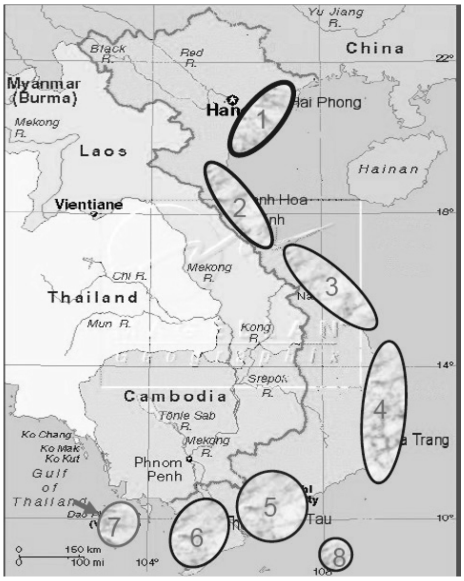


Figure 3: The possible locations for developing new seaports systems
Source: VPA

3. Criteria in the Determination of Seaport Location and the Satisficing Concept

3.1. Criteria in the Determination of Seaport Location

The seaport system location is determined by "where the ports are, where the market is, where the domestic freight is, where the international freight is, where the road system, rail system, or airport is, and where the labor is". It is selected based on analyze the trade-offs between freight and labor cost and speed of delivery requirement. It is good if it locates near a major highway, railroad and airport. It allows trucks carrying containers from ports to access the different levels. Actually, it should settle close to the population centers. Moreover, there is planning of land available at reasonable prices, which allows construction of large, of educated workforce.

The major functions of seaports system are to facilitate trade, so they need to offer and maintain higher levels of efficiency at reducing cost. Nowadays, a dramatic increase in size of containership that generates massive economy of scale effects and hence even lower transport costs requires attentions from seaports. Besides, they ensure environmentally sustainable. However, the efficiency, low cost and environmental sustainability could not reach simultaneously.

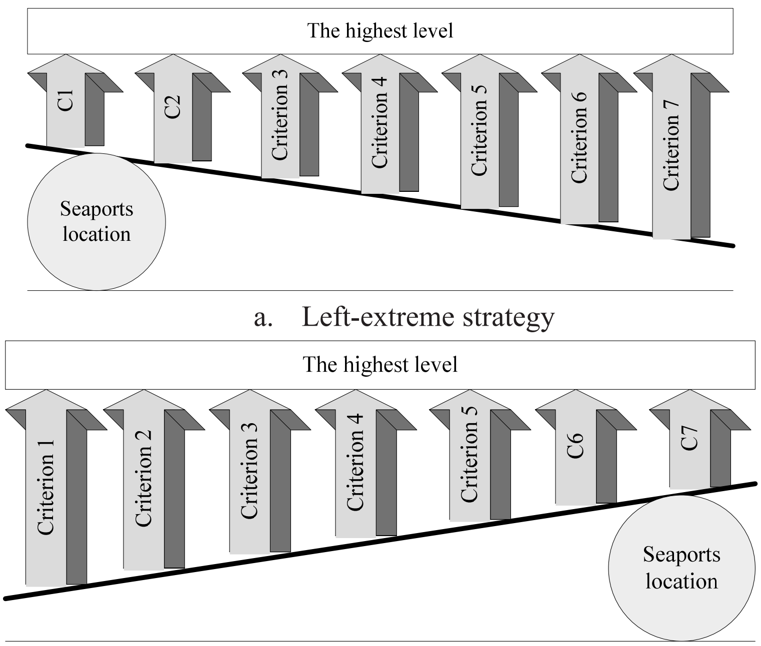
With all what mentioned above, some criteria that should be used to evaluate seaports location are proposed in this paper as shown in Table 1.

Table 1: Proposed criteria for port location

Criteria	Description
Natural water depth	That is suitable for building deep-water ports in order to accommodate big-sized vessels.
Existing transportation infrastructure/ superstructure network	Including highways, rail road, and airport. It is more convenient to access to ports so that easy to expand port's hinterland.
Labor forces	As transport is derived demand, then the more trade there is, the more jobs there will be in ports, as well as in other services. Labor force in the selected location should meet these requirements.
Land	Locating or relocating a port does not only meet the current expected capacity but also the demand in the future. Opportunity cost of land should be also put into consideration.
Distance to economic centers	Short distance to big economic centers is advantage of ports.
Environment effects	The impact to surrounding environment
Distance to international transport networks	Short distance to international sea routes is advantage for ports.

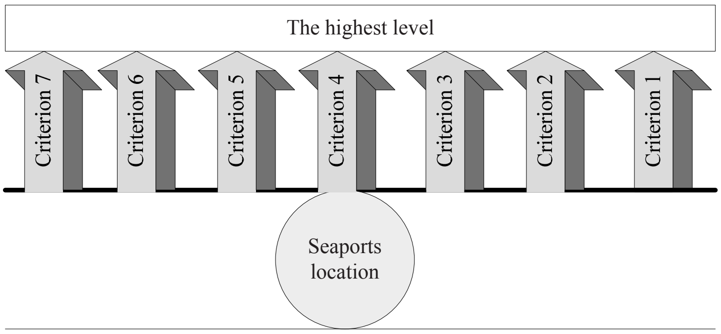
3.2. The Satisficing Strategy

Satisficing is a decision-making strategy which attempts to meet criteria for adequacy, rather than to identify an optimal solution. The idea of this strategy is explained as in Figure 4.



a. Left-extreme strategy

b. Right-extreme strategy



c. Satisficing alternative

Figure 4: Satisficing strategy

In order to illustration the satisficing strategy, it is assumed that three alternatives corresponding to three different locations are considered with regard to 7 criteria. The first and second location, see Figure 4a. and 4b. come with good results with regard to some criteria, but bad results for the other in comparison among three alternatives. The third alternative should be selected since it meets all criteria at an adequate level.

4. Application

The new seaport system in Vietnam has to meet the demands of growing shipping industry in the future. Moreover, it is expected to operate at the most efficient and cost effective, so planners should recognize existing terminals, deepen the harbor's channels and berths, and improve inland access. It is able to operate new high capacity, environmentally friendly cranes, which could alleviate traffic and port congestion. Therefore, it is difficult to determine the very strong position to compete in fast growing market, because seaports are a large boost for the local and Vietnam's economy.

The unequal location of seaports is a considered problem. The seaports in the north, which hold up to 25-30% of transport volume, still have abundant capacity. The ports in the central region are now lacking goods and only a part of capacity has been used. Meanwhile, the southern ports, amounting to 57% of total transport volume (90% in container transport) are overloaded (Ngo, 2007). Therefore, the new seaports system in the south, which serves for one of the most important and biggest economic center, Ho Chi Minh City, is a top priority.

There are many alternatives to develop a new seaports system in this case. The first one is considered to expand the existing seaports system. Most of them are located along rivers and inner the city, so this character used to be the ports' advantages such as utilized cities' infrastructure as well as located close to industrial zones. Besides, Ho Chi Minh City has a good labor force in both quantity and quality aspects. Moreover, if this alternative is selected, natural environment is changed at minimum level in comparison with other ones. However, it now shows weaknesses such as limitation of depth so that they cannot accommodate large sized vessels, which can only handle the small sized vessels of up to 1,500 TEUs. The limitation of hinterland is now seriously considered, where land used for developing new seaports is unavailable, deficient and expensive. And the terrible traffic tram problem at the big city hampers this selection.

The second place that could be considered to develop the new seaports system locates on the southern central area in Vietnam. Natural deep-water harbor is the strongest point, where the Van Phong bay has 25 meters depth, which suites for developing ports that could serve big ships up to 15,000 TEUs. These ports could serve the southern center provinces lasting from Binh Dinh to Binh Thuan as well as economic centers in the south. Besides, they also serve the northeastern Cambodia and northeastern Thailand via national highways 19 and 24 and play an important role in the economic zone and also for goods transited between southern China and ASEAN countries. Moreover, large land is available for developing ports and its price is still low. On the other hand, it locates far from Ho Chi Minh City around 400 kilometers, so it takes a long time for transport freight on the land. There is not existing infrastructure/superstructure for seaports, so it needs much investment in developing them. Labor force is not available or do not meet requirements. Because the new seaports system is constructed on the natural areas in this case, environment effects should be considered seriously.

As mentioned previously, there are the trade-offs among criteria in both alternatives. None of them satisfy the requirements of the new seaports system serving for economic centers in the south of Vietnam. Determining a suitable location for the new seaports system is a challenge of planners and researchers. Ba Ria – Vung Tau is a province known for its advantageous geographical location, where Cai Mep – Thi Vai area is a good candidate for relocating the seaports system from Ho Chi Minh City. It could be able to accommodate large vessels with draft of up to 14 meters. It is still available land for developing ports, and the land cost is not too expensive or cheap in comparisons with inner Ho Chi Minh City or Van Phong area in respectively. It requires not much investment in infrastructure/superstructure, because Vung Tau meets basic requirements in infrastructure/superstructure. Especially, the distance to Ho Chi Minh City is around 90 kilometers. With a good road system, it takes approximately 2 hours travel time by trucks. Because it locates not too far from Ho Chi Minh City, labor force might be interested in working for. If city ports are relocated in this place, congestion in ports or by ports is improved essentially. Besides, natural environment effects are not serious problems.

Therefore, 26 investment projects aimed to build ports along Thi Vai River system have been licensed. Of those, five projects have been completed and are in use. The other projects have been underway with the new container port, Cai Mep, and Ba Son shipyard. Ben Dinh – Sao Mai port complex is one of the key national projects for development in the period 2006 – 2010 under the national master plan for seaports. It is located downstream of the Thi Vai River close to the sea. The complex will provide oil, gas and container handling services and has a shipyard and a petroleum depot. It will have six container wharfs capable of handling 100,000 DWT vessels and around 50 million tons of goods annually.

5. Conclusions

This paper presents the satisficing strategy that is applied in decision making problem for seaports location. Some aspects and criteria relative to the seaports location are discussed. A case study of replacing city ports in Ho Chi Minh City to a suitable location was considered as an illustration for this strategy. The study shows that there is no perfect alternative, so a satisficing alternative should be selected, which satisfies criteria at the expected levels. Actually, some factors were suggested in this paper, whilst other ones could be used in suitable cases.

All seaports systems in Vietnam could be constructed in capable locations as shown in Figure 3, but because of some limitations such as budget, resources, and environment regulation the priority for each seaports system should be assigned. Each location has strong points beside weak points, so the seaports system should classified in main function, geography or capacity to focus developing. Besides, other strategies to support the ports industry should be used such as co-opetition (Song, 2003), privatization (Baird, 2002), regionalization (Notteboom, 2005), specialization.

References

Baird, A.J. (2006), Optimizing the container transshipment hub location in northern Europe, Journal of Transport Geography 14 (2006), Elsevier, pp. 195-214.

Baird, A.J. (2002), Privatization trends at the world’s top 100 container ports, Maritime Policy & Management, 29(3), pp. 271-284.

Ben Nghe port’s website: www.bennghepor t.com.

Danang port's website: www.danangportvn.com.

Department of Planning and Investment of KhanhHoa Province's website: www.khanhhoainvest.gov.vn.

Gordon, J.R.M, Lee, P.M and Lucas Jr, H.C (2005), A resource-based view of competitive advantage at the Port of Singapore, Journal of Strategic Information Systems 14 (2005), Elsevier, pp. 69-86.

HaiPhong port's website: www.haiphongport.com.vn.

Hochiminh City's website: www.hochiminhcity.gov.vn.

Malchow, M.B. and Kanafani, A. (2003), A disaggregate analysis of port selection, Transportation Research Part E 40 (2004), Elsevier, pp. 317-337.

Minkes, A.L. and Foxall, G.R. (2003), Herbert Simon and the concept of dispersed entrepreneurship, Journal of Economic Psychology 24 (2003), Elsevier, pp. 221-228.

Ngo, L.T. (2007), Seaports in Vietnam in deficiency or abundance?.

Notteboom, T.E. (2005), Port regionalization: towards a new phase in port development, Maritime Policy & Management, Volume 32, 2005, pp. 297-313.
Saigon port's website: www.csg.com.vn.

Simon, H.A. (1957), Models of Man: Social and Rational. New York: Wiley.

Song, D.W. (2003), Port co-opetition in concept and practice, Maritime Policy & Management, Volume 30, 2003, pp. 29-44.

Tyson, C.J. (2008), Cognitive constraints, contraction consistency, and the satisficing criterion, Journal of Economic Theory 138 (2008), Elsevier, pp. 51-70.

UNCTAD, (2004), Assessment of a seaport land interface: an analytical framework.

Vietnam General Statistics Office's website: www.gso.gov.vn.

Vietnam International Container Terminals' website: www.vict-vn.com.

Vietnam Maritime Administration's website: www.vinamarine.gov.vn.

Vietnam National Shipping Lines' website: www.vinalines.com.vn.

Vietnam Seaports Association's website: www.vpa.org.vn.

Analyses of the competitiveness of ports available for the Southwest and their attraction to the land-port in Guizhou

Qin Fang [1]*,Yonglei Jiang*[2] *and Zhongzhen Yang*[2]

[1] Department of Civil Engineering, Guizhou University, Guizhou, China.
[2] College of Transport Management, Dalian Maritime University, Dalia, China.

Abstract

In order to analyze the port choice behavior of the exported container in the Southwest in China if a land-port is constructed in Guizhou, this paper develops an attraction model of the seaport on land-port based on the gravity principle and establishes a system to evaluate the total competitiveness of the seaports and proposes a method to calculate the competitiveness with Analytic Hierarchy Process (AHP) approach. And at last, this paper analyzes the probabilities by which the exported containers transferred in Guizhou land-port choosing the seaports.

Keywords: Land-port, seaport, total competitiveness, and analytic hierarchy process

1. Introduction

The seaports available for Chinese Southwest are mainly located in Pearl River Delta, Yangtze River delta and the Beibu Gulf, however, the over long road transport becomes the bottleneck for the container import and export in the Southwest. In order to encourage the import and export container transport in the Southwest, we have proposed to construct the land-port in Guizhou to do the container collection, custom clearance, inspection and to delivery the collected containers to the seaports with artery railway liner (Guizhou Transportation Planning Survey and Design Institute, 2008). The land-port can integrate the flexibility of the road transport and the economies of scale of the railway transport to improve the accessibility of the Southwest in terms of international trade transport.

Figure 1 shows the method of building a network for international trade transport with the land-port in the Southwest. It is a typical hub-spoke transport network, in which containers are transported by artery railway liner from land-port to seaports, by artery shipping lines from seaports to oversea ports, while delivered by road feeder from the origins to the land-port. This kind of network can not only reduce the logistics cost but also improve the transport efficiency. In the system, the seaport choice behavior of containers in the land-port has great impacts on operating artery railway liner as well as seaport marketing strategy. Therefore, it is necessary to study the competitiveness of the seaports and their attractions on the land port in Guizho.

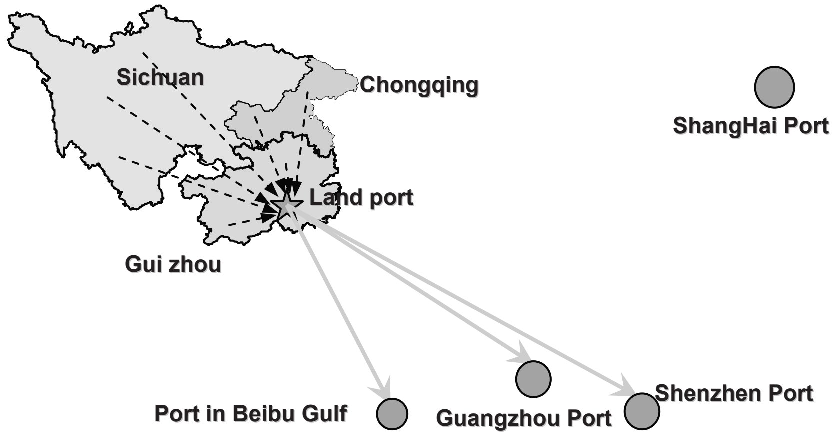


Figure 1: Network of International Transport based on Land-Port

2. Attraction of the Seaport on the Land-Port

The attraction of the seaports on a region can be described with gravity law, and the specific forms can be described with Eq. 1 (Yang et al., 2005).

$$F_{gj} = \alpha I_j / c_{gj}^2 \tag{1}$$

Here, F_{gj}=the attraction of seaport j on containers at land-port g ; I_j ="mass" of the seaport, the "mass" is term in the law of universal gravitation, here it refers to the total competitiveness of the seaport; c_{gj}=transport impedance between land-port and seaport, it means the generalized transport costs and can be calculated with Eq. 2 .

$$c_{gj} = \frac{c_{R,gj}}{c_{R,gj} + c_{T,gj}} c_{R,gj} + \frac{c_{T,gj}}{c_{R,gj} + c_{T,gj}} c_{T,gj} = \frac{c_{R,gj}^2 + c_{T,gj}^2}{c_{R,gj} + c_{T,gj}} \tag{2}$$

Here, $C_{R,ij}$ =railway transport cost, $C_{T,ij}$=road transport cost. Thus, according to the random utility theory, attraction probabilities of a seaport on the land-port can be calculated with Logit model (Ben-Akiva and Lerman, 1985) shown in Eq. 3.

$$P_{gj} = \frac{F_{gj}}{\sum_{k=1,\dots,n} F_{gk}} = \frac{I_j / c_{gj}^2}{\sum_{k=1,\dots,n} I_k / c_{gk}^2} \tag{3}$$

Eq. 3 shows the behavior with which containers in the land-port choosing the seaport randomly. Since $\sum_{j=1,\dots,n} P_{gj} = 1$, the model can be used to divide the market shares of the seaports on the land-port. In order to calculate P_{gj}, we firstly need to know the "mass (I_j)" of the seaport, namely the total competitiveness, as well as the generalized transport costs (c_{gj}) between the land-port and seaports. We will discuss the calculation methods afterward.

3. Model for Evaluating Container Port Competitiveness

3.1. Evaluating Index System

Many existing researches have studied on the evaluation of the port competitiveness, and the basic methodologies are similar. The first work of the evaluation is to build an evaluating indices system. Based on the work of Yand et al. (2005), an evaluating index system is established as shown in Figure 2. It is obviously that in order to evaluate the available ports with the index system, we firstly need to determine the weights of the elements in the second level, namely to quantitatively determine the importance of the five elements that influence the port competitiveness. Next, we also need to determine the weights of the elements in the third level and then we should quantify some qualitative elements in the system. Here we use AHP approach do the quantification, which is proposed by Saaty in 1970 (Gu and Zhou, 2005).

3.2. Weights of the Elements in Second Level

In order to know the relative importance between each pair of elements in second level, we did a questionnaire survey on related experts with mails and e-mails. We examined the consistency of the data on the reclaimed questionnaires and deleted the un-satisfied ones (CR<0.1, means the data with contradictory answers). With the left data, we calculate the standardized weight vector of each validated questionnaire, and finally we take their algebra means and standard the means to obtain the eigenvalue vector $(0.17,0.21,0.24,0.15,0.23)^T$ of the importance matrix of the elements in second layer. These values represent the import degrees of port scale, port infrastructure, port operation environment, port service level and port shipping condition on the port competitiveness.

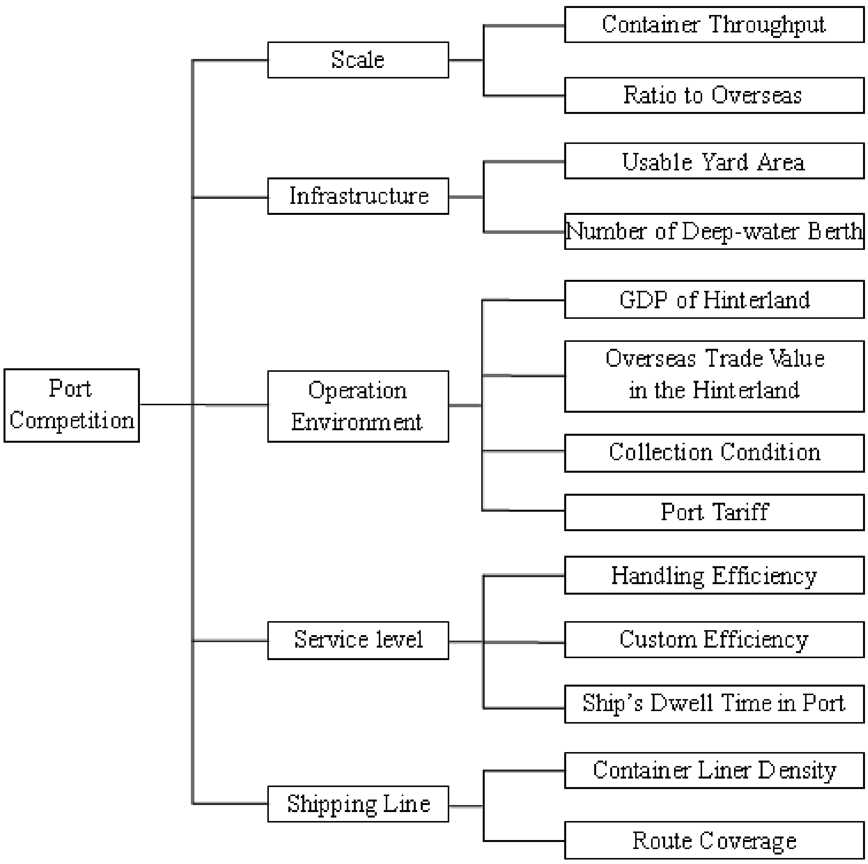


Figure 2: Indices for Evaluating Port Competitiveness

3.3. Quantifying the Indices and Determining their Membership Grade

The indices in level three include two qualitative ones, namely customs efficiency, collection condition. Here we quantify them with the scores given by the experts, and the algebra mean of all scores is taken as the quantified value. The left indices can be given the values with data from the ports (data resources: the Statistical Yearbook, the statistical information network, Chinese Port

Yearbook, transport development report). The actual values of the left indices are shown in Table 1 and the values of their membership function are listed in Table 2.

Table 1: Actual Value of some Indices

Index	Shanghai Port	Guangzhou Port	Shenzhen Port	Ports in Beibu Gulf
Container Throughput (M. TEU)	2171.9	660	1847.02	2.09
Ratio to Overseas (%)	45.3	26	72.1	70.3
Usable Yard Area (10000 M^2)	472.1	180	245.8	253
Number Of Deep-water Berth	131	59	61	29
GDP in Hinterland (Billion Yuan)	40781.7	32610	32610	18115
Overseas Trade Value in Hinterland (Billion Dollar)	5216.5	4963.5	4963.5	1253
Handling Efficiency (TEU/Hour)	300	150	250	80
Container Liner Density (Voyage/Month)	2106	392	1500	150
Route Coverage (Piece)	200	37	168	30
Ship's Dwell Time in Port (Day)	0.39	1	0.6	1.65
Port Tariff (Yuan/TEU)	515	470	600	660
Collection Condition	7.13	6.8	7	4
Custom Efficiency	7.1	6.8	7.1	5

Table 2: Values of the Membership Function of the Indices

Index	Shanghai Port	Guangzhou Port	Shenzhen Port	Ports in Beibu Gulf
Container Throughput (M. TEU)	1.000	0.30388	0.85042	0.00096
Ratio to Overseas (%)	0.628	0.36061	1.00000	0.97504
Usable Yard Area (10000 M^2)	1.000	0.38128	0.52065	0.53590
Number Of Deep-water Berth	1.000	0.45038	0.46565	0.22137
GDP in Hinterland (Billion Yuan)	1.000	0.79964	0.79964	0.44420
Overseas Trade Value in Hinterland (Billion Dollar)	1.000	0.95142	0.95142	0.24018
Handling Efficiency (TEU/Hour)	1.000	0.50000	1.00000	0.26667
Container Liner Density (Voyage/Month)	1.000	0.18614	0.71225	0.07122
Route Coverage (Piece)	1.000	0.18500	0.84000	0.15000
Ship's Dwell Time in Port (Day)	1.000	2.56410	1.53846	4.23077
Port Tariff (Yuan/TEU)	1.096	1.00000	1.27659	1.40426
Collection Condition	1.000	0.953727	0.98177	0.56101
Custom Efficiency	1.000	0.95775	1.00000	0.70422

3.4. Determining the Weights in Level Three with Fuzzy Entropy Method

Entropy is a concept in thermodynamics, which is used to express the energy failure degree in a material system. It is later introduced into the information theory, and the uncertainty of signal in the communication is called information entropy that is widely applied in many disciplines (Wang et al., 2006). Here, the method to calculate the entropies of the evaluation indices of the port is introduced as follows.

Assuming that there are n ports to be evaluated and m indices in level three, and the membership grades are given, then a $m \times n$ fuzzy evaluation matrix can be obtained as shown in Eq. 4.

$$Y = (y_{ij})_{m*n} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix} \tag{4}$$

Here, y_{ij} = membership grade of port *j* in terms of index *i*. if there are *m* indices and *n* being evaluated objects, the entropy of index *i* can be defined as:

$$H_i = -\frac{1}{\ln n}\sum_{j=1}^{n} f_{ij} * \ln f_{ij} \text{ , } i = 1,2,...,m \tag{5}$$

$$f_{ij} = y_{ij} / \sum_{j=1,...,n} y_{ij} \text{ , when } f_{ij}=0, \ \ln f_{ij}=0 \tag{6}$$

The value of the entropy reflects the utility of the index to the port competitiveness, the smaller the entropy the more important the index. Base on above equations, we can further obtain Eq. 7, which is the weight of the entropy of index *i*, then with Eq. 7 we can calculate the weights of the entropy of all indices to get Table 3.

$$\omega_i = \frac{1 - H_i}{m - \sum_{i=1,...,m} H_i}, \quad 0 \le \omega_i \le 1 \quad , i = 1,2,...,m \tag{7}$$

Table: 3 Weights of the Entropy of the Indices in Level Three

Index	Weight	Index	Weight
Container Throughput (M. TEU)	0.227082	Container Liner Density (Voyage/Month)	0.196238
Ratio to Overseas (%)	0.040779	Route Coverage (Piece)	0.158474
Usable Yard Area (10000 M^2)	0.040194	Ship's Dwell Time in Port (Day)	0.080683
Number Of Deep-water Berth	0.080513	Port Tariff (Yuan/TEU)	0.005106
GDP in Hinterland (Billion Yuan)	0.022132	Collection Condition	0.014039
Overseas Trade Value in Hinterland (Billion Dollar)	0.060392	Custom Efficiency	0.005670
Handling Efficiency (TEU/Hour)	0.068698		

3.5. Calculation of Port Competitiveness

Assuming that the weights obtained with AHP in level two is W, entropy weights of sub-indices from the method of fuzzy entropy are W_A, W_B, W_C, W_D and the corresponding fuzzy evaluation matrixes are Y_A, Y_B, Y_C, Y_D, then the competitiveness *I* of the ports can be estimated with Eq. 8 (Wang et al., 2004). In our study the competitiveness of the four ports are shown in Table 4. It can be seen that Shanghai port has the highest competitiveness, while the ports in Beibu Gulf have lower competitiveness.

$$I = W^T * Y = W^T * \begin{bmatrix} W_A^T * Y_A \\ W_B^T * Y_B \\ W_C^T * Y_C \\ W_D^T * Y_D \\ \vdots \end{bmatrix} = (I_1, I_2, I_3, I_4, \cdots) \tag{8}$$

Table 4: Total Competitiveness of the Ports

Index	Shanghai Port	Guangzhou Port	Shenzhen Port	Port of Beibu Gulf
Score	0.180668	0.099683	0.165478	0.089946

4. Attraction of Seaports on the Land-Port

For simplification, we take 20-feet heavy container as the unit to calculate transport costs and the costs and time for a container from Guiyang to Shanghai, Guangzhou, Shenzhen and Beihai Gulf are listed in Table 5 (note: the being built Xia (Xaimen)-Rong (Chengdu) expressway and Gui (Guiyang)-Guang (Guangzhou) high speed railway are taken into account).

Table 5: Railway and Road Transport Cost and Time

Origin - Destination	Railway		Roadway	
	Cost (RMB/TEU)	Time (Day)	Cost (RMB/TEU)	Time (Day)
Guiyang-Shanghai	2,442	5	6,600	4
Guiyang-Guangzhou	1,250	3	3,600	2
Guiyang - Shenzhen	1,400	3	3,900	2
Guiyang- Beihai Gulf	1,563	4	4,015	2

In order to integrate time and cost into one factor, we change time into money term with $vot = M \times i / 365$ (Guan and Nishii, 2000; Wang, 2001), here M is the value of time of the freight in a container and i is the loading ratio. With the data in "Chongqing Statistical Yearbook 2007", the value of time of container departures from Guizhou is $547000 \times 10000 \times 7.5 \times 44\% \times 7.29\% \div 202451 \div 365$ =17.9RMB/day. Based on the value, the generalized costs of container transported by railway and highway from Guiyang to seaports are shown in Table 6 and the weighted inland transported cost c_{gj} are shown in Table 7.

Table 6: Generalized Costs of Container from Guiyang to the Seaports

	Railway (RMB/TEU)	Highway (RMB/TEU)
Guiyang - Shanghai	2531	6672
Guiyang - Guangzhou	1304	3636
Guiyang - Shenzhen	1454	3936
Guiyang – Beihai Gulf	1635	4050

Table 7: Weighted Inland Transport Cost from Guiyang to the Seaports

Origin - Destination	Total Cost (RMB/TEU)	Origin - Destination	Total Cost (RMB/TEU)
Guiyang- Shanghai	5533	Guiyang-Shenzhen	3267
Guiyang-Guangzhou	3020	Guiyang – Beibu Gulf	3355

With the above data and Eq. 3, we calculate the attractions of the seaports on the containers in the land-port to get Table 8. Obviously, land-port construction causes the exported containers in the Southwest to mainly select the ports of Shenzhen and Guangzhou. Ports in Beibu Gulf are ranked in the third, while Shanghai port with its strong competitiveness can only attract 14% containers in the land-port. It means that with Guizhou land-port, Xia-Rong expressway and Gui-Guang high speed railway, 86% exported containers in the Southwest will no longer go to Shanghai, but choose the ports in Pearl River Delta and Beibu Gulf, which are much nearer than Shanghai.

Table 8: Attraction of the Seaports on the exported Containers in the Land-Port

	Shanghai	Guangzhou	Shenzhen	Beibu Gulf
Attracting ratios (%)	14	27	39	20

5. Conclusion

The construction and operation of land-port can improve regional transport network for the freights traded between countries. Organizing foreign trade container transport based on land-port may shorten the distance from the inland where is far from the seaports to its overseas partners in terms of transport time and cost. The calculated results show that, due to the economies of scale of the long distance railway transport between the land-port and the seaports, the seaports with the highest competitiveness may not be able to get the biggest market share. It also indicates that the integrated transport system is important for the transport of foreign trade containers. The planning and construction of the logistics nodes such as land-ports, and railway transport liner are extremely essential to the improvement of the integrated transport network.

References

Ben-Akiva, M. and Lerman, S., (1985). Discrete Choice Analysis: theory and Application to Travel Demand, The MIT Press, .

Guan, H.Z. and Nishii, K., (2000). The research of assessment methods and application of time value of the goods, Science and technology of highway traffic, 17 (5): 107-110.

Gu, Y.Z., Zhou, X.Z., (2005). BCG matrix analysis for port logistics business based on AHP, journal of Shanghai Maritime University, 04 Issues in 2005.

Guizhou Transportation Planning Survey and Design Institute (2008). “Research Rport for construction of regional land port in Guizhou”, Guiyang, China.

Wang, C.Y., Xiao, P.M. and Xiao, L., (2006). The application of entropy weight fuzzy comprehensive evaluation in the port competitiveness evaluation, 27 (2): 131-134.

Wang, H.Y. (2001). The research of determination method for the time value of passenger and cargo transportation, Master Thesis, ChangAn University, XiAn, China.

Wang H.Y., Zhou, W., and Wang, Y.Q., (2004). The research of determination method for the time value of goods traffic, science and technology of highway traffic, 21 (7): 131-134.

Yang, Y.B., Yang, Z.Z., Yang, Z., and Zuo, Z., (2005). Evaluation of Competition Ability and Market Share for Container Port, Proceedings of the Eastern Asia Society for Transportation Studies, Vol.5, 2483-2493.

An airline human factors risk management model: RSF/RIF and score formula based approach

Prof. Dr Triant Flouris[1] and Dr Ayse Kucuk Yilmaz[2]

[1] School of Aviation Sciences, Daniel Webster College, Nashua, NH 03063-1300, USA
[2] School of Civil Aviation, Anadolu University, 26470 Eskisehir, Turkey

Abstract

This paper addresses the question "is human factors based risk unmanageable in the airline business?" We suggest a human factors (HF) risk management model for airlines containing a new full-set of risk shaping/influencing factors (RSF's/RIF's) taxonomy and suggest a scoring formula. This model is expected to support corporate risk management and human resource management efforts aimed at identifying and managing human factors related issues and to provide a systematic process-based approach to airline management. Managers can provide holistic snapshots of their company's human resources by this human factor based risk management concept. We apply our theoretical model to the airline industry and more specifically: a. discuss the importance of human factors in airline management and operations, b. develop a human factors based risk list for airlines, c. assess the cost of human based risks on an airline's financial and operational performance, and d. develop a human factors based risk analysis model for the airline business.

Keywords: Airline Management, Aviation Human Factors, Human Resource Management, Risk Management, Risk Analysis, Risk Scoring.

1. Introduction

Airlines, as other companies face a multitude of diverse risks in their dynamic business environment. Risks are an expected part of business life and entrepreneurship. Risks, business systems, and their activities are generally interrelated with human factors either directly or indirectly. The human element impacts all managerial and operational systems and activities in a company. Therefore, managing human factors is an inseparable part of business management.

Our Human Factor Risk Management (HFRM) model is structured by managing activities of human based risks for businesses and provides a comprehensive and coherent framework for the evolution and management of human factors. The HFRM model continues to anticipate potential human factor based risks and develops appropriate responses for various scenarios.

Our HFRM model can help provide a more precise understanding of risk factors associated with the human element and to furnish insights on their appropriate identification and prioritization via scoring and management.

This study reviews both leading enterprise risk management (ERM) guidelines and strategic human resource management (SHRM) literature in the development process of the HFRM model. These are reviewed in view of their context of human factor based risks. A number of weaknesses in current literature and works on both ERM and human resource management have been identified. These weaknesses come mostly from a single acting approach to human factors, the absence of internal strategies, and lack of systematic and holistic approach for managing human based risks.

SHRM and ERM literature have generally focused on the risk source dimension of human factors thus ignored or insufficiently focused on the dual role of human factors. In addition, human based risk work did not use taxonomy and related risk scoring approaches. The HFRM model supports SHRM and whole management efforts via considering the multi-dimensional nature of human factors. Our model is designed as a combination of holistic risk management (to financial, strategic, operational and threat risks) and strategic human resource management.

The human element is both a critical aspect of aviation safety and airline management in view of airline's financial and operational performance. Historically, safety in the aviation system was sought through enhanced systems, equipment design and certification procedures. A technological plateau was achieved in the 1970s yet incidents and accidents continued to occur. This directed the attention of the international community to the human component. The influence of human capabilities and in particular limitations in safety of operations has been evident for many years. However, little attention or interest has been directed to them (Aviation Watch, 2009).

Managing human factors based risk within this complex environment involves implementation proactive approach, well planned risk management process and preventive measures, such as continuously improving process, human factor analysis, training for improve risk awareness, create and maintain risk culture and consideration for the various elements (Adopted from National Steering Committee on Patient Safety, 2004):

- Personnel (evaluate to ensure optimal numbers for workload, proper credentials and staff physical / mental well-being).
- Equipment (evaluate to ensure that needed devices are present, functioning properly, monitored for safety and regularly serviced with a plan for phased and emergency replacement).
- Environment (evaluate for physical designs that may inhibit or increase risks to those receiving or providing care).
- Administration (create an organizational culture of safety, evaluate and plan for effective policies and procedures — including a policy for reporting actual and potential risks to those receiving or providing care).

The commercial airline is an extremely competitive, safety-sensitive, high technology service industry. People, employees and customers, not products and machines, must be the arena of an organization's core competence. The implications are vast and pervasive affecting no less than the organization's structure, strategy, culture, and numerous operational activities (Appelbaum and Fewster 2002).

Our study is organized into four main sections. After the introduction, we explain strategic human resource management in airline industry and key drivers of the human factors risk management model in section two. Our methodology is provided in the third section. A verbal model is used for developing the HFRM model, which contains a full-set of risk shaping/influencing factors (RSF/RIF) and a score formula. Finally, we conclude with a summary and significance of our model as well as with a brief discussion of future research possibilities.

2. Strategic Human Resource Management and Drivers of the Human Factor Risk Management Model

Strategic human resource management (SHRM) focuses on the relationship between an organization's strategy and the management of its human resources (Milmore et al., 2007). By combining the HRM function with business strategy, SHRM reflects a more flexible arrangement and utilization of human resources to achieve set organizational goals, and accordingly helps organizations gain a competitive advantage (Wei, 2006). The Strategic perspective of HR has grown out of researchers' desire to demonstrate the importance of human resource practices in organizational performance (Delery and Doty, 1996). The field of HR strategy differs from traditional HR management research in two important ways (Becker and Huselid, 2006):

i. SHRM focuses on organizational performance rather than individual performance.

ii. SHRM emphasizes the role of HR management systems as solutions to business problems rather than individual HR management practices.

With respect to HRM: increasing work-force diversity; the increasing size and complexity of companies; the absolute and relative rise in the labour costs of firms; the intensification and globalization of competition; the individualization of HRM-issues; structural developments in the labour market; and innovations in technology, have been the factors that have lead to this increased acknowledgement of HRM-related issues. With respect to the airline industry: it is one of the fastest growing industries in the world and has gained importance through the current era of globalization. Today operations are being dealt with on a worldwide basis, and airline travel plays an increasing role in bridging geographical distance. According to the International Air Transport Association (IATA), the number of passengers travelling on scheduled flights has increased on average by over six percent annually for the last three decades. In comparison with 2003, international airline passenger traffic increased by over fifteen percent in 2004, and in certain regions, such as Asia- Pacific, which includes China, passenger numbers increased by over twenty percent, showing that the industry is becoming ever more significant (Wilson, 2005).

The term "human factors" has grown increasingly popular as the commercial aviation industry has realized that human error, rather than mechanical failure, underlies most aviation accidents and incidents. If interpreted narrowly, human factor is often considered synonymous with crew resource management (CRM) or maintenance resource management (MRM). However, it is much broader in both its knowledge base and scope. Human factors involves gathering information about human abilities, limitations, and other characteristics and applying it to tools, machines, systems, tasks, jobs, and environments to produce safe, comfortable, and effective human use. In aviation, human factors is dedicated to better understanding how humans can most safely and efficiently be integrated with the technology. That understanding is then translated into design, training, policies, or procedures to help humans perform better (Graeber, 2009).

Aviation is an excellent example in which a high-risk industry implemented co-ordinated and comprehensive strategies to reduce preventable accidents. Also, the study of human factors engineering has led to an understanding that, although adverse events will occur in any human endeavor, they can be minimized through the design of equipment or tools, design of the tasks themselves, the environmental conditions of work, the training of staff, and the selection of workers. Airline regulators, plane manufacturers, and commercial airline carriers have combined human-factors engineering with the knowledge that failures in communication and co-ordination among team members have led to tragic aviation accidents. Their collaboration resulted in a wide variety of mandatory and voluntary processes that have dramatically improved passenger safety (National Steering Committee on Patient Safety, 2004):

- Redundancy in key operating systems
- Simulator training to improve teamwork and prepare for sudden emergencies
- Restrictions on the number of consecutive hours worked
- Mandatory reporting of designated aviation accidents / incidents
- Voluntary reporting of near misses
- Extensive use of information technology for the provision of flight information and weather conditions
- Comprehensive and objective investigation of accidents with reporting of the probable cause
- Procedural checklists with alarms for key equipment and/or human failures

Human Factor Risk Management recognizes that people do not always understand, communicate or perform consistently. Each individual brings to the workplace a unique background and technical ability, and has different needs and priorities (COSO, 2004, p. 5.). Both sources and managers of risk, human resources have dual roles in corporate management systems and risk management. Human resources are one of the keys to success in corporate management systems. For this reason, the aforementioned dual roles (risk source and risk manager) require a different approach to both human resource management and the corporate management systems. This is due to the fact that the human element has different, highly dynamic and hardly controllable characteristics from other corporate sources.

The drivers for developing a new Human Factors Risk Management (HFRM) model are as follows:

- integration of human factor risk management into the organization as a part of achieving their overall goal of a managed corporate culture
- increase the human factor contribution to company functions and activities
- meet requirements for managing human factors
- reduce costs arising from human performance limitations and add value through improved human performance
- meet demand of business owners and high level managers

Desired outcomes from our HFRM model, taxonomy and score formula are as follows:

1. human factors will be considered as leading risk factors in corporate management system and organization
2. human based factors will be managed to achieve corporate objectives (e.g., financial outcomes, operational performance)
3. threats and opportunity based awareness and responsibility will be increased amongst the managers and other personnel in the organization
4. human factor management will be considered an essential part of the business planning process

The HFRM model intends to maximize the benefits of existing functions and activities in all departments via a human-centric approach. Expected additional benefits include:

1. a strategy for human factors risk management across the organization
2. a generic framework, which enables flexible and tailored approaches
3. a systematic managerial tool for the best management of human factors based risks
4. a common language and culture for corporate risk management
5. a tool for increasing human based opportunities and decrease human based threats
6. a common approach for internal monitoring, control, and review
7. a strong managerial tool for continuous control of Human Factors based risks

3. Methodology

We use qualitative research methodology and verbal modeling in setting up our model. Additionally, we use quantitative risk analysis methodology in order to aid in the assessment of the human factors. A new full-set of the RSF/RIF taxonomy has been determined and weighted by AHP, and then results are used in the new score formula. Details of our model and related taxonomy are not given in this article.

Faced with rapid change organizations need to develop more focused and coherent approaches for managing people. The HFRM model allows for the management of human risk (like any risk) to be a continuous process of identifying, analyzing and mapping areas that have the potential to cause threats and can provide an opportunity in implementing system improvements. We developed a model that would allow human risk to be considered systematically and which would gain maximum advantage from tools and processes already existing. Detailed information about our HFRM Model Elements and Tools is not given in this article.

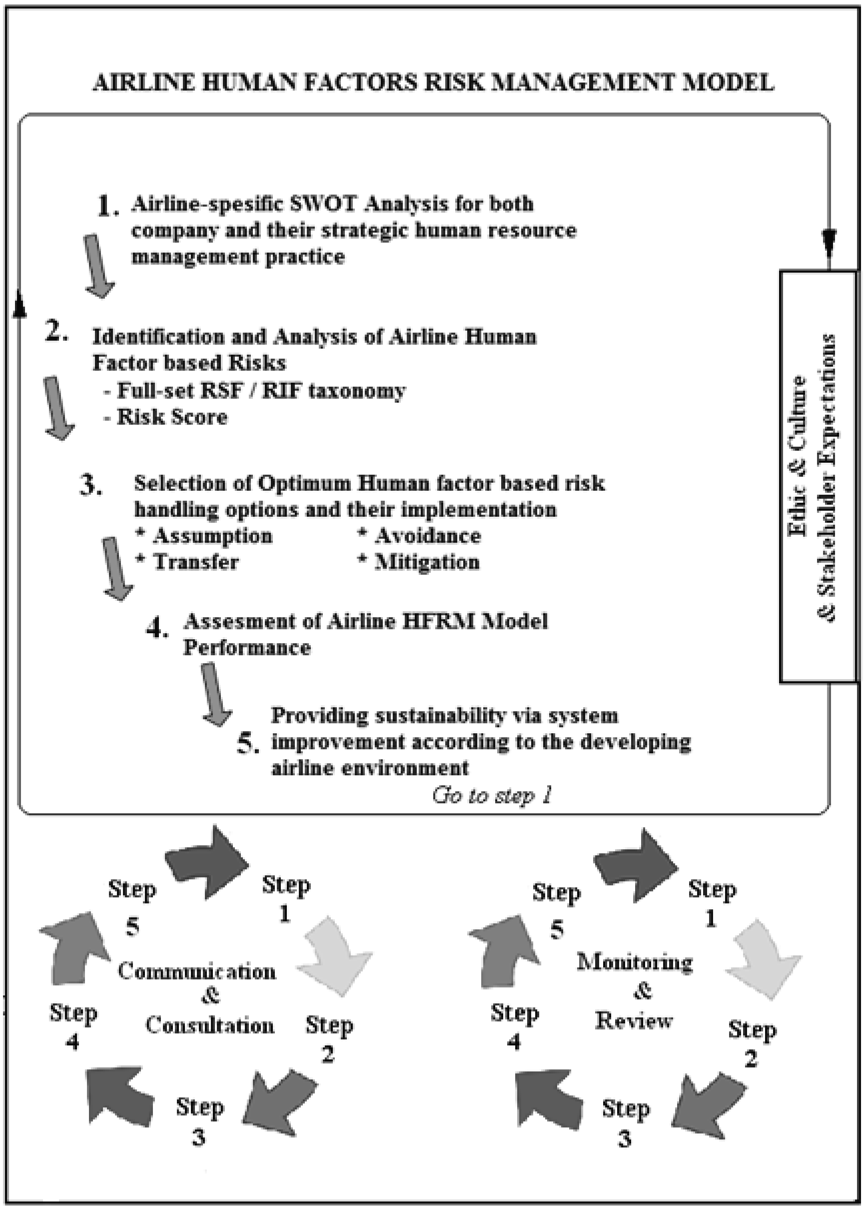


Figure 1: Human Factor Risk Management Model

Human Factor Risk Management (HFRM) model provides a generic framework for the establishment and implementation of the management process within organizations. A number of large companies have recognized the value in adopting some kind of a risk model. The proposed model is primarily a process for applying human factors tools and techniques in a coordinated and systematic way. The model has a number of elements that deal with and link management processes to corporate sustainability. The model also takes into account work ethic, culture, and stakeholder expectations as organizational culture is affected by these concepts. This model can be applied in any industrial and business management segment. Our HFRM process includes a continuous monitoring, review, communication and consultation system along with its five steps. These four steps are integral to the entire HFM process. Communication and consultation will be reflected in each step of the process. Figure 1 provides an overview of the model and its elements as a continuous loop.

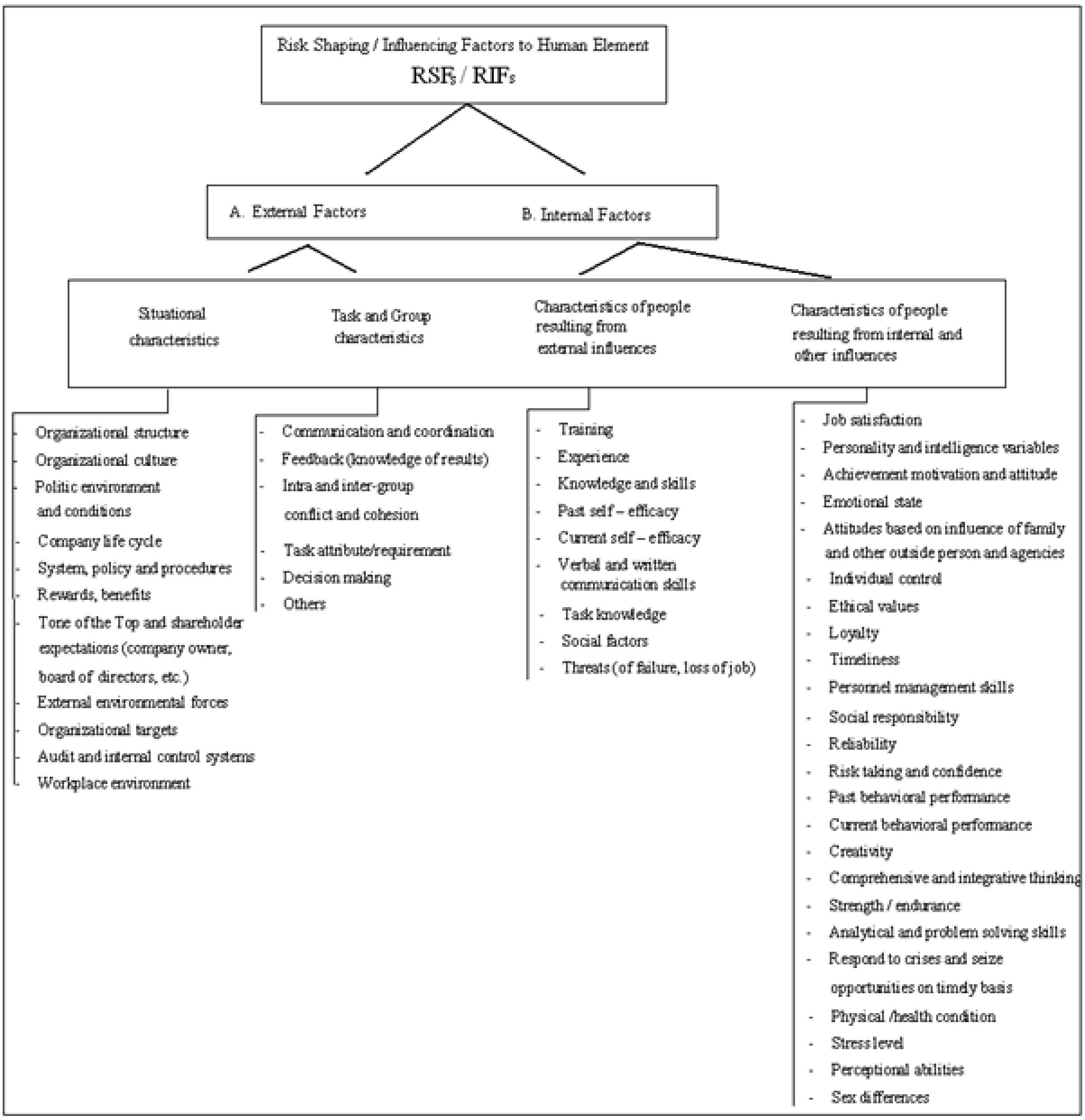


Figure 2: Full-set RSF/RIF taxonomy for HFRM model

Our full-set RSFs/RIFs taxonomy has been based on Performance Influencing /Shaping Factors (PIF /PSF) of Miller and Swain (1987) and Ya-Lih Lin and Sheue-Ling Hwang (1992). These are the most appropriate taxonomy when compared to others for our systematic approach of the HFRM model. All the factors researched and collected from the above-mentioned sources are collated into a new full-set RIF/RSF taxonomy. After that, we omitted RIF/RSF items that had little relevance to human risks. Alternatively, items that were determined to have a bearing on human factors were included. Items involving the stressor factors were also removed because it was difficult to calculate risk score from the approach of human risk factors. These factors affecting and influencing the human performance in business management can be classified into two main groups; external and internal risk shaping/ influencing factors (RSFs /RIFs). At last two main, four sub group categories and 45 items were determined. Figure 2 shows our full-set RSFs/RIFs taxonomy items and categories.

After that new full-set RIF/RSF taxonomy, the Analytical Hierarchy Process weights the impact score of the human factor. Weighted results are used in calculation of total score: Likelihood <multiplied by> impact. Corporate high-level managers and human resources managers or selective jury members are responsible in this process. They report to present company management according to the score formula results. Impact levels of RSFs/RIFs are ranked by Saaty's Analytic Hierarchy Process.

Impact score of RSFs / RIFs

Icon	Name	Normalized by Cluster	Limiting
No Icon	IMPACT SCORE	0.00000	0.000000
No Icon	Characteristics of people resulting from external ~	0.27960	0.279604
No Icon	Characteristics of people resulting from internal ~	0.34171	0.341707
No Icon	Situational characteristics	0.25263	0.252634
No Icon	Task and Group characteristics	0.12606	0.126055

Figure 3: Impact score results by AHP

Analytical Hierarchy Process results show that internal factors are highly influenced and shaped by human risk factors (see figure 3). All of the RIF's factors are important to achieving organizational objectives and human performance. Impact factor scores will use to calculation risk likelihood as following:

Impact (Characteristics of people resulting from external influences) **x** 0.279604;
Impact (Characteristics of people resulting from internal and other influences) **x** 0.341707;
Impact (Situational characteristics) **x** 0.252634
Impact (Task and Group characteristics) **x** 0.126055.

The our score formula is depicted in Figure 4. Prioritization via the human score formula is a critical outcome of our model. Managers can use this formula in their human resource performance analysis. The calculation of the risk score includes internal and external factors. RIFs/RSFs and score formulas can use personnel performance assessments in the company. For this reason this model provides for a systematic approach and risk score formula as a supporting tool to human resources and its performance management.

This formula and its result have great applicability and importance in assessing human resources in a company in terms of current performance situations. High scores achieved mean that employees provide advantages to a company. If an employee achieves a low score, this person should improve in terms of their job related activities such as (depending on the applicability to the company) sales performance, training, marketing skills, human relationships, etc. Personnel empowerment should be one of the required fields.

If the results of the score calculation are analyzed in detail, the strengths and weaknesses of personnel can be seen clearly and understood by the management of a company that employs this system. This can create the opportunity to reinforce the evaluated personnel's strengths. Managers can then work on optimizing personnel capabilities and they can derive maximum benefit from personnel capabilities and skills. The most fundamental contribution of the HFM model is the effective management of human based threats and opportunities towards achieving corporate objectives. Thus, human based threats can be minimized and human based opportunities can be maximized. Therefore, a company can be provided with reasonable assurance towards organizational and strategic objectives linked to corporate sustainability.

Human Factor Score Formula

Likelihood Scale

$L = \{ 0 < L < 1 \}$ — Likelihood calculation is included favorizing conditions, adverse conditions and uncertainty in this formula. These should be considered by selective jury members in the calculation of likelihood.

Impact Scale

$I = \{ 0 < I < 100 \}$

Internal Factors:

I_f = Characteristics of people resulting from internal and other influences

E_f = Characteristics of people resulting from external influences

External Factors :

S_f = Situational characteristics

T_f = Task and Group characteristics

R = Total Risk Score

$$I_f = \frac{\sum_{i=1}^{n} I_{f_i} \times L_{I_{f_i}}}{n} \times 0.34$$

$$E_f = \frac{\sum_{i=1}^{n} E_{f_i} \times L_{E_{f_i}}}{n} \times 0.28$$

$$S_f = \frac{\sum_{i=1}^{n} S_{f_i} \times L_{S_{f_i}}}{n} \times 0.25$$

$$T_f = \frac{\sum_{i=1}^{n} T_{f_i} \times L_{T_{f_i}}}{n} \times 0.13$$

$$R = I_f + E_f + S_f + T_f$$

Calculated rounded weighted rates of Impact by Analytic Hierarchy Process are :

to Characteristics of people resulting from external influences is = 0. 28

to Characteristics of people resulting from internal and other influences is = 0.34

to Situational characteristics is = 0.25

to Task and Group characteristics is = 0.13

Figure 4: Score Formula to risk analysis of Human Factors

Like many airlines, Qantas has advanced systems for detecting and managing human risk in flight operations. The Qantas Human Factors Strategy provides a framework for applying Human Factors risk management principles in any operational area. The model uses a safety management systems (SMS) approach to manage human performance limitations. The Human Factors Strategy is intended to maximize the benefits of existing practice whilst outlining a common approach for Human Factors management across the Qantas group. The proposed solution is primarily a framework for applying human factors tools and techniques in a coordinated and systemic way. The model has a number of elements concerned with risk management processes as part of normal operations, as well as "triggers" for when particular HF attention is required for example, following an event or incident, the acquisition of new equipment or the need to manage change. The model also takes into account that human behaviors do not occur within a vacuum, but are influenced by an organization's safety culture, which should also be subject to periodic measurement, assessment, and improvement. The model can be applied generically across all the operational businesses and has sufficient flexibility to cope with customized tools and processes within each element (Raggett, 2006:2).

4. Conclusions

Our HFRM model can be useful to get much clear see and recognize of human factor at all points by airlines and is beginning to improve the management of human risk to optimization of airline operation and its performance. HRM is a major contributor to the success of any organization. Human factor should be an important piece of the enterprise risk management at airlines and aviation operations. A human risk management model based framework that allows systematic evaluation of human factors research and interventions could contribute to minimizing human based threats and optimizing aviation operation's performance. Our HRM framework is designed to support airline operators to focus their efforts on human factors issues that may negatively affect the performance of the aviation operations.

In this study, we have offered a new Human Factor Risk Management (HFRM) model. We developed a fresh full-set of human factor based risk shaping/influencing factors (RSFs/RIFs) taxonomy for use in the Human Factor Risk Management process. We calculated the impact weighting of the obtained RSFs/RIFs items that we evaluated through an analytical hierarchy process (AHP) technique. Impact weighted scores were used in scoring formula to prioritization of human factor based risks. The proposed full-set RSFs/RIFs taxonomy and score formula is used in our Human Factor Management (HFM) model. An appropriate analysis/assessment framework must be developed in order to use the taxonomy in other human resource areas.

This paper presents a model for managing human factors in airline operations via a well-grounded, proactive, and systematic process based approach. The HFRM model is a very helpful tool for airline managers in assessing and managing their human resources; therefore, his model is offered as managerial tool. Likelihood and impact elements of our formula may improve employee performance. An effective HFRM model can provide a sustainable competitive advantage, critical to the success of airline.

Our research aims to contribute to the fields of corporate sustainability management, risk management, airline management, human factors management, and human resource management. Since the processes of the model are capable of dealing with all kinds of feedback and dependence when modeling a complex decision environment, we advocate that this HFRM model, full-set RSF/RIF taxonomy and formula are useful and workable. The HFRM model deals with complexity of human factors and provides insights to managing human factor based risks.

References

Appelbaum, S.H. and Fewster, B.M. (2002), Global aviation human resource management: contemporary recruitment and selection and diversity and equal opportunity practices, Equal Opportunities International, Volume 21, Number 7, 2002, pp. 66-80(15), accessible from http://www.ingentaconnect.com/content/mcb/030/2002/00000021/00000007/art00006

Aviation Watch, (2009), System Safety and Human Factors, accessible from http://www.aviationwatch.co.uk/#fn3

Becker, B.E. and Huselid, M.A. (2006). Strategic Human Resources Management: Where Do We Go From Here?, Journal of Management, Southern Management Association.Vol. 32 No. 6, December 2006, 898-925.

Committee of Sponsoring Organizations of the Tradeway Commission (2002), COSO's ERM Integrated - Framework (draft version), accessible from http://www.erm.coso.org.

Delery, J.E. and Doty, D. (1996). Harold Modes of Theorizing in Strategic Human Resource Management: Tests of Universalistic, Contingency, and Configurational Performance Predictions, The Academy of Management Journal, Vol. 39, No. 4 (Aug., 1996), pp. 802.

Graeber, C. (2009), The Role of Human Factors in Improving Aviation Safety, Aero Magazine, No. 8,Boeing Commercial Airplanes Group, accessible from http://www.boeing.com/commercial/aeromagazine/aero_08/human_textonly.html, January 30, 2009)

Lin, Y.L. and Hwang, S.L. (1992). The application of the log linear model to quantify human errors. Reliability Engineering & System Safety [RELIAB. ENG. SYST. SAF.]. Vol. 37, no. 2, pp. 157-165. 1992.

Milmore, M., Lewis, P., Saunders, M., Thornhill, A. and Morrow, T. (2007) Strategic Human Resource Management: Contemporary Issues, Harlow: FT/Prentice Hall, p.37.

Miller, D.P. and Swain A.D. (Solvandy, G. ed.) (1987). Handbook of Human Factors. Wiley InterScience. Prudue Univ.; Glendon, A. Ian ; Sharon Clarke and Eugene F. McKenna. (2006). Human Safety and Risk Management, CRC Pres, 2006, p.151.

National Steering Committee on Patient Safety (2004), Building a Safer System: A National Integrated Strategy for Improving Patient Safety in Canadian Health Care, pp. 5-6.

Raggett, L. (2006), A new human factors risk management program for Qantas, ess2006: evolving system safety - The 7th International Symposium of the Australian Aviation Psychology Association. 9 - 12 November 2006, Manly Pacific Hotel, Sydney, Australia, accessible from http://www.aavpa.org/seminars/ess2006/pdf/pdf%20papers/Raggett.pdf*,* p.2.

Wei, L. (2006). Strategic Human Resource Management: Determinants of Fit, Research and Practice in Human Resource Management*,* 14(2), 49-60.

Wilson, S. (2005), Human Resource Management in the Airline Industry - The Example of Star Alliance, Master thesis, Arts in International Business, Universitat Erlangen, Nürnberg, Friedrich Alexander Universität Erlangen Nürnberg (WISO) pp.2-3.. accessible from http://www.grin.com/e-book/41155/human-resource-management-in-the-airline-industry-the-example-of-star#,

Maritime transport patterns in container ships in the European Union

Fernando González-Laxe and Isabel Novo-Corti [1]

University Institute of Maritime Studies, University of A Coruña, Spain.
[1]For contact:
Department of Economic Analysis and Business Administration, School of Economics, Campus de Elviña
Tel : +34 981 167000, Ext: 2445
Fax: +34 981 167070
E-mail : inovo@udc.es

Abstract

European Transport Policy, underlined in the white paper two priority objectives to be attained by 2010: the regulation of competition between modes and the consolidation of a link-up of modes widely linked to intermodality. To ensure sustainable mobility in Europe, it has been defined a global strategy and a rules to protect the environment and promote economic activity at the same time.

The increasing level of seaborne transport is only comparable to road transport growing. The growing process of containerisation and building an infrastructure for regular maritime lines, Short Sea Shipping, and the project of motorways of the sea corridors, feet in this scene. On the other hand, the proper economic development, the better port infrastructures and services for management, storage and distribution of goods have supported the increasing size and number of containers handled, especially of SSS. The shown dynamics seems to draw a new stage, with new expectations for transport sector, with more concentrated markets affecting to all agents engaged. In this paper we intend to make an approach to the levels of concentration of maritime transport in European countries and also to the main important ports, focussing both loaded and unloaded goods. Our aim is to find possible changes in those concentration levels and affecting market distribution and structure in the maritime transport sector, mainly in container ships, linked with the new economic situation.

Key words: Maritime Transport, Concentration.

JEL Classification: D43, L11, L91, L92, R40

1. Introduction

The Maritime transport role in whole transport is increasing and it's also being promoted by institutions in order to achieve less environmental aggressive modes. Nevertheless the evolution of transport growth all over the world has been increasing in all modes more or less at the same levels, making a difficult task for governments to promote economic and social development at the same time that they try to preserve the natural environment.

Sustainability is a very important goal for Europe. The wide relation between Economic growth and transport network became in a great increasing of all transport modes. Everybody knows that protecting the environment and making it compatible with economic development, has to do with the searching for the most ecological transport mode without constraining economic growth.

In this way, European Union Transport Policy is trying to move freight transport from road to sea, and now we are beginning to see some results (as the guidelines of White Paper of Transport have drawn -

European Commission. (2001). The results are not in a lineal way for all maritime transport: as it was easy to expect the containers traffic is the one that is showing the most important levels of growth.

Our goal in this paper is to analyze mentioned transformations in the context of European Union, paying a special attention to container ships movements. For doing that, we have focussed on EUROSTAT maritime transport data and we have elaborated some concentration and specialization index, that lead us to achieve some conclusions about changing patterns in European maritime traffic of freight.

We first make a draw of maritime transport in European TEN-T, focussing on the role of economies of scale and the main objectives and the concentration process in agents involved in maritime traffic. Next, we analyze the evolution of maritime traffic in each European Unión member country and in its' main ports, we complete this point of view focussing in 20 container ports in Europe studying the distribution of charge between them, and the evolution in the last years. In the other hand, we make an overview of number of vessels and the weight of transported charge, specifically of container ships, in order to detect some changing patterns in the structure of maritime freight movements in Europe. Eventually, some conclusions are shown as a conclusion of our findings.

2. Marine Transport in the European Ten-T

Under this perspective, an important development of maritime transport has appeared, which not only has increased, it also has been growing and changing, at the same time. Indeed, there are different maritime transport modes with different ways of development. The cargo market is usually divided into several different markets: container, Ro-Ro, conventional general cargo (break-bulk), dry bulk, liquid bulk. The maritime transport industry is supported by distribution and storage networks and infrastructures.

An examination of the most relevant features of practices in the maritime transport sector reveals: a) standardization of norms; b) homogenization of services; and c) objectives are set in order to achieve economies of scale and agglomeration; i.e. organizations produced by greater concentration and centralization of economic agents in the sector, in the form of consortiums, alliances or strategic partnerships. It`s said that, actually, maritime transport, to be precise, short sea shipping is the only real hope in holding back the spectacular growth in road freight transport all over the EU.

Increasing containerising process is promoted for the globalisation and economics developments, increasing economic activity, trade liberalisation, reduced import tariffs, outsourcing, increasing containerisation of dry bulk and break-bulk cargoes Henderson et al. (2002). The process of containerisation is an important element in the current context of maritime transport. It has increased dramatically since the sixties. Technical, technological and logistical advances have transformed and brought the maritime transport sector up to date Kuby and Reid (1992), Talley (2000). Not only is new cargo transported, but in addition the total capacity of ships has increased. This assertion can be supported by the following data: a) As the European Commission Staff Working Paper 1139 (2007) indicates, container shipping has been the fastest growing sector of the maritime industry during the last two decades, the number of containers shipped worldwide has nearly doubled in the past six years from 60.5 million twenty-foot equivalent units (TEU) in 2000 to 112 million TEU in 2006. In Europe the largest part of container traffic is concentrated in North Western Europe (the "Hamburg-Le Havre range") whose ports registered in 2005 a 41.7 million TEU, representing a share of 56% of the total European container traffic. In particular, the three largest European ports Rotterdam, Hamburg and Antwerp handled 23.86 million TEU. b) the containership market has maintained an annual growth rate of 10% over the last decade, aiming to reach 10 million TEUs by the year 2008, when in 1996 3,196 TEUs were entered; 5,071 in 2002 and 7,691 in 2005, by the top 100 global operators in terms of container capacity; c) the amounts accumulated in ports have also increased, maintaining a growth rate of 6% over the last decade; d) transfers at ports are on the increase, meaning greater imbalances when assessing container transport, their classifications as full/ empty, or analysing import and export and e) average ship size has increased, indicating a trend towards naval giganticism, e.g. currently there are ships with 12,000 TEU transport capacity, when fourth generation ships (known as Post Panamax during the 1988-2000 period) transported between 4 and 6,000 TEUs, and the fifth generation ships (known as Post Panama Plus) could hold between 6,000 and 10,000 TEUs. (Paixao and Marlow (2003); González-Laxe (2005).

The container transport industry consolidates two aspects: firstly, determining different vehicles related to ship property and rental; secondly, concentration dynamics i.e. increasing traffic density on particular regular lines in order to achieve a more competitive position. The results of such dynamics are evident: a) the leading ten companies provide 60% of the total supply, while ten years ago this figure reached only 43.6%; b) the perspectives for consolidation in the containerisation sector are increasing as new construction requirements for vessels will be carried out by the five leading companies and c) this means that a maritime industry oligopoly may form, according to Allix (2006).

Such processes feed off each other through partnership agreements; consortiums or alliances in the quest to reduce risk, increase display capacity, and minimise transport costs by widening the range of activity. Similarly, the processes of merging and acquiring businesses herald a new era, both in terms of the composition of economic power and the alignment of strategies employed by regular line operators, Frémont and Soppé (2004), leading to a high concentration of maritime agents.

Adding all these points, if we are looking for economies of scale and the concentration process in agents involved in maritime transport, we can see a characteristic stage: Regular lines for transportation, regular ports where inward or outward the freight and vessels increasing more and more. It seems to entail a new structure for afford this new era in maritime transport: bigger and stronger companies, enforced ports structure and services, and a better intermodality system. The panorama seems to appoint to a new stage were concentration will be necessary for competitiveness for all agents.

Shifting the balance between modes involves looking beyond the rightful place of each particular mode and securing intermodality. The biggest missing link is the lack of a close connection between sea, inland waterways and rail. For centuries sea and river dominated goods transport in Europe major towns were built on rivers or on estuaries and the large trade fairs in the Middle Ages were always held at river or sea ports. Nowadays, despite a slight revival, water transport is the poor relation even though it is a mode which is not expensive and does less damage to the environment than road transport -White Book: The European Transport Policy (2001).

Taking into account Europe geography, its history and globalisation process the European Union is still dependent on the maritime transport. Nearly 90% of its external trade and more than 40% of its internal trade goes by sea; on the whole nearly 2 billion tons of freight are loaded and unloaded EU ports each year; maritime companies belonging to European Union nationals control nearly 40% of the world fleet; the majority of EU trade is carried on vessels controlled by EU interests; and finally the maritime transport sector - also including shipbuilding, ports, fishing and related industries and services - employs some 3 million people in the European Union.

Key aspects and trend in maritime traffic	
ROUTES AND REGIONS 1. Concentration of maritime routes in the great central East-West and North-South areas. 2. Concentration of flows on the Long Branch routes. 3. High growth rate of traffic with the East. 4. Imbalance in flows between regions.	***PORTS*** 1. Development of hub ports at connection points on branch routes and feeder routes. 2. Evolution from maritime-land interfaces towards logistical platforms and intermodal nodes. 3. Increase in ferry activity.
SHIPOWNERS 1. Process of fusions and acquisitions between global operators. 2. Streamlining traffic for economies of scale. 3. Using ships with increased capacity. 4. Local alliances between large and small ship-owners for feeder routes. ***5.*** Vertical alliances with logistical operators	***TERMINAL OPERATORS*** 1. Growth in the participation of global operators. 2. Vertical integration of ship-owners, terminal operators and logistical operators. 3. Dominant presence of global operators in Asia, Europe and North America. ***4.*** The markets in Latin America, Oceania and Africa are developing.

Figure 1: Based on a study by Pérez Fiaño (2007).

The strategies of maritime companies, terminal operators and port authorities are set out in Figure 2 outlining the different main objectives of maritime agents, results and impacts. The diversity of concepts and strategies of the agents and institutions is particularly noticeable, from maximising profits and position on the market (formulated by the Maritime Companies) to customer loyalty and logistical services to increasing service value (by the Terminal Operators) or maximising profits in cargo maintenance (Port Authorities' aim). It is equally important to emphasise different concepts in terms of the variety of instruments used: Maritime Companies tend towards prioritising their shares when studying fees and costs; terminal operators are characterised by their attention to prices and use of technology; port authorities place emphasis on maritime access, followed by territorial regulation and concessions Kent and Ashar (2001).

OBJECTIVES AND INSTRUMENTS OF MARITIME AGENTS			
	Maritime Companies	***Terminal Operators***	***Port Authorities***
OBJETIVES	Maximizing profits; improve position on the markets; controlling logistical chains	Maximizing profits; customer loyalty and logistical services; and increasing value-added.	Contribution to minimizing costs through logistical chains and maximizing cargo maintenance
INSTRUMENTS	Fees; controlling costs in terms of capacity, cargo volume, time, cooperation etc, marketing and services.	Prices; maintenance technology for improving quality, speed, safety, information etc.	Maritime access, territorial regulation and concessions; socio-economic negotiation; pricing policies.
IMPACTS	Large ships; streamlining sailing networks; alliances and consortiums; and specialized terminals.	Economies of Scale, industrial logistics	Information about maritime access, guarantee of social and economic stability, industrial strategy and concessions policy.

Figure 2: Our own elaboration.

3. Evolution of European Maritime Traffic: An Overview for Countries and Marin Ports

Concentration is inherent in the maritime transport sector in regular lines (mergers and acquisitions are usual). Its form, nature and effects have improved modes and the organization of the sector -and also changed production- Foued (2007). For an overview of concentration levels in maritime traffic in the EU we focus on three key aspects: The maritime traffic held in the whole of Europe and its distribution among countries, the maritime traffic held for each country contrasted with the traffic in main ports (that is to say, ports with more than 1 million Tm -tones- of freight) and the movement of cargo in top ports in the EU, in the context of vessel maritime traffic.

Methodology supporting our research is based in one hand, on traditional concentration index, in the other hand on specialization Bird index. The more usual index are the CR ad the Hirchman-Herfindahl concentration indexes. The first one shows the aggregated percentage for the top agents, in this way, the CR4 is reporting information about the accumulated percentage for the 4 top agents; although this index is criticized for omitting the number of agents involved, its intuitive information makes it a very useful tool. The second one is $I = \sum_{i=1}^{n} p_i^2$, where $p_i^2 = \frac{\text{Handled goods for "i" agent}}{\text{Handled goods for all agents}}$. This index belongs to the Hannah-Kay characterized for taking into account the whole concentration curve, as opposed to the CRi index. $HKI_a = \left[\sum_{i=1}^{n} p_i^{(1+a)}\right]^{1/a}; \quad a > -1 \quad and \quad a \neq 0$. The parameter in the exponent (*1+a*; in this case is *2*) modulates the weight level assigned to the agents with a major market quota. In this case when the market is distributed between big and small agents, the concentration measure of Hannah-Kay indices are usually as big as big the parameter. HHI is a particular case in this group.

The objective in the first stage of this paper is to see if some concentration of maritime traffic is taking place in some countries or in some special geographical areas, and to detect movements of loaded or unloaded freight from smaller ports to bigger ones in the same country, we also analyze the possible movements of containerizing freight from one top port to another among the 20top. The second stage fits on evolution of container ships by size of vessels in last seven years, in order to complement the concentration level observed in countries and ports.

Analyzing maritime traffic in UE countries, we can appreciate that amount of cargo handled in the EU-27 ports in 2006 was 3.834 million tonnes (See Table 1). We have analyzed the level of this traffic in each country in order to detect possible changes among ports; in other words, we tried to find out if there is some concentration of traffic of goods in any port caused by the diminishing of traffic in another port. First we have analyzed the total figures (inwards plus outwards), but we want to underline that it is correct to say that this is the movement of cargo of goods handled, but we must be careful with the meaning of "global transport of goods": It is obvious that these totals may include a "double counting" (it is possible that goods loaded in one port would be then unloaded in another port. If both are reported data to Eurostat, the movement of cargo is being double-counted). To avoid possible confusion with double-counting we also analyze inwards and outwards cargo. The level of goods handled has increased by 3.2% since 2005 (in 2005 it was 3,717 million tonnes). It grew in all Member States except Latvia (- 5%), Poland (-3%), Romania (-2%) and the United Kingdom (-0.2%), but the most relevant rise has been seen in Slovenia (23%). This has to do with the dry bulk goods handled in the port of Koper; Finland (11%) and Bulgaria (11%). have also increased the handled good level. The United Kingdom is still the leading EU-27, despite the slight decrease mentioned above, representing more than 15% of the EU-27 total. Italy is the second one, with a share of 14%, followed by the Netherlands (12%) and Spain (11%). Greece and Spain show the highest increase in the same period. However, in these cases the increases are mostly due to the improvement of the statistical coverage[1]. Since 1997 goods handled in EU-15 ports have increased in 463.568 million tonnes (24.7%). A similar trend is seen for EU-12. The progressive growth is not the same for all countries. By analyzing global trends for the top countries we find that the United Kingdom has grown far below the other top countries. Italy, the Netherlands, Spain, and Greece have grown at a

[1] Greece: The statistical coverage of data has considerably improved between 2001 and 2002, being these reference years. In particular, collection of data on ferry boats started from the last quarter of 2001. From 1997 to 2003, in the tables from the "Passengers" collection the number of passengers corresponds only to the number of non-cruise passengers ("ferry passengers"). Spain: Data include Ceuta and Melilla. The statistical coverage has significantly improved in 2001 (inclusion of new ports). Data only cover "main ports". Data for the period 2003-2005 are provisional and likely to be revised. (Eurostat Methodological notes, statistics in focus 62/2008)

rate of 18.6%, 17.6%, 50.9% and 49.3%, respectively. EU-15 increased its handled goods in 20.9%, and the Euro Area (12) went up to 26.8%.

We can appreciate increasing values for United Kingdom 21.8% Italy 19.0%, the Netherlands 14.1%, Spain 56,1% and Greece 49.9% inwards and -18.2%, 18.4%, 30.3%, 39.8% and 49,0% outwards, respectively, analysing loaded and unloaded cargo. The leader shows a less balanced growth, we also found that the low growth that has taken place in the last years is due to the loaded goods, because the unloaded ones support acceptable growth levels. Other top countries grow in an unequal way, as loaded or unloaded goods are considered. The most balanced growth is shown by Italy and Greece. For Italy and Spain inwards flows rise over outwards flows, contrary to this, in the Netherlands the level of inwards good is quite small. The quantity of outwards goods is relatively low in comparison with the inwards ones; this fact might explain these growth taxes. In addition, while the inwards goods are growing every year, the outwards range shows positive and negative growths along the period analyzed.

The accumulated percentage of freight (CRi) for main countries is shown in table 2. The results indicate that only five countries achieved more than 60% of freight, not only for total freight, but also for inwards and outwards. In all European countries evolution of cargo handled is quite similar, in the sense that it doesn't show significant differences among countries. There are no signs of changes in cargo from one country to another. The observation of inwards and outwards cargo movements shows approximately the same results. The Hirschman-Herfindahl concentration index -see table 3- indicates low levels of concentration in all cases (inwards, outwards and total goods handled). A slightly decreasing trend is shown for freight outwards and it remains in the same levels for total goods handled due to levels of traffic inwards (bigger than outwards), that remain in the same concentration level along the period of study. According to what we stated above, we can't conclude that some concentration process was taking place in any country of the EU-15 from 1997 to 2006. By analyzing the HHI for EU-27 from 2003 to 2006, it is possible to appreciate the same tendency, as shown in table 4. Concentration values are low, particularly for outwards freight. Nevertheless, the main movements of freight are inwards flows, and the value of HH concentration index in this case is also low, but bigger than the outwards. The bigger number of countries taken into account when analyzed within EU-27, shows an decreasing in HHI values as opposed to EU-15.

4. The Whole Maritime Traffic and Main Ports

We have proved that maritime traffic is increasing in the European Union, and we have also elaborated CRi and HHI concentration indexes to evaluate the concentration level by countries. Our next step is to analyze the performance of each country, so that we are going to search the level of whole maritime traffic and the main ports traffic. In order to do that, we must analyze possible interchanges of cargo among ports that belong to the same country. We have made a distinction between the main ports in the country (with handled goods over 1 million tonnes) and the other ports. Eurostat provides quarterly statistics for main ports, as well as annual statistics for all goods handled in each country (in other words, for all the ports in each member state). We have homogenized both series and put the information together to elaborate an index: Handled goods for main ports/Total Handled goods in all ports, in percentage terms. This indicator (shown in Table 5) shows relatively homogeneous results for all countries in the European Union[2]. Data only show a complete series for ten countries available in the EU: Belgium, Denmark, Germany, Ireland, Greece, Italy, the Netherlands, Portugal, Finland and Sweden. By analysing countries with a complete series, we can appreciate that most maritime traffic is supported by the main ports, almost in all cases it is over 70%. Focusing on the evolution, the dominating feature is the stability, except for Greece and Ireland. Greece has begun to reduce its maritime traffic in main ports since 2000 showing the lowest level in 2002, while in 2004 the main ports reached a level under 70%. This shows less concentration in

[2] We need to underline the exceptional data of year 2001 for Belgium and France. This difference may be originated in different database sourcing (although both come from Eurostat) or in data reported from countries to Eurostat. In any case it is not a very significant divergence with our analysis.

maritime traffic than at the beginning of the serial. Ireland presents a decrease at the beginning of the period but since 2001 it remains steady. Finland goes down in 2001. Other countries show the same path, with levels of concentration between the 72% of Denmark and the 99.4% of the Netherlands all through the period. The average values stay around 88%, decreasing slightly at the end of the period.The highest concentration levels are in the Netherlands (99.4% in 1997 and 99.8% in 2006), Greece (99.4% in 1997 and 67.7% in 2006), Finland (93.3% in 1997 and 88.5% in 2006), Belgium (98.1% in 1997 and 98.7% in 2006) and Germany (96,6% in 1997 and 97,3% en 2006), although there are some differences regarding the cases, in particular because while the Netherlands, Germany and Belgium show an increase in the concentration of handled goods in their main ports, Greece and Finland show exactly the opposite trend. For 2005 and 2006 we can analyze complete series for the 20 countries, the 15 European Union, the 13 Euro area and the 12 Euro area. Data for all European countries in 2005 and 2006 show stability in percentage of maritime traffic held by the main ports of all countries.

We can summarize that there is a relevant concentration of freight in main ports, as it was expected. Nevertheless, if we were searching for an increasing or decreasing tendency in concentration levels, we could not find any radical change. As data show, the same percentage of freight supported by main ports applies to EU-25, EU-15, EU-13 and EU-12, even with diminishing on one point. The countries that have increased concentration of maritime traffic in main ports are Ireland, France, Italy, Portugal and United Kingdom while Estonia, Greece and Cyprus have that concentration diminished. Taking account the appreciations shown above, we can’t come to the conclusion that there is an increasing concentration. No obvious conclusion can be come up. More empirical evidence is needed to sustain a solid estimation, but we can argue that may be some increasing concentration is beginning to appear in some countries.

To afford the high level of competition, firms are thinking about saving costs through scale economies. This helps shipping lines to invest in containerships with more capacity. The increase of vessels size concentrated the main growth in vessels between 5000 and 7500 TEU and in excess of 7500 TEU (the larger size range). The capacity has increased twelve-fold in the last ten years, with an annual growth of about 30% and a trend towards big size vessels[3]. The presence of scale economies is linked to this process Cullinane and Khanna (1999) and it may involve a change in market structure and even in maritime traffic flows direction. In our research about changes in maritime freight volume searching for a concentration levels evolution, now we focus on the analysis in vessel traffic in the European Union. To avoid double-counting problems we only take into account inwards vessel traffic.

First of all, we focus on levels of vessel traffic in EU countries in terms of the total number of vessels and the number of container ships. Secondly, we analyze data for both regarding gross registered tonnage (GRT). Finally, we elaborate some comparative indexes to determine the relation GTR/Nº in order to achieve some results about vessels size evolution. The number of vessels handled in EU 15 in the period 2000-2007 is characterized by a progressive growth (1.763.454 vessels in 2000 to 1.88.257 in 2007), showing the highest level in 2004 (2.062.587 and in 2006 with 2.0208.907). In reference to the kind of vessels, most are non-specialised general cargo carriers, liquid bulk ships (tankers) (showing a slightly decreasing tendency), container ships ((in progressive growth), dry bulk carriers (decreasing slightly), miscellaneous vessels (dredgers, research vessels, others), specialized carriers, vessels for offshore activities, fishing vessels, dry cargo barges, tugs and others.

The most important increase in number of vessels is recorded in Spain (314% between 2000 and 2007), and Portugal (121%). Most countries show a rising tendency, except France (-30%), Greece (-32%), Italy (-26%) and Denmark (-2%). In terms of cargo, a progressive growth is shown for EU-15

---

[3] In 2000 10% of the total fleet was represented by vessels with a capacity in excess of 5000 TEU, by 2010 the share of this vessel size is expected to represent 40% of the total fleet. On the Far East – Europe route the average vessel size in 2000 was 4500 – 5500 TEU; in 2010 it is expected to be 8000 – 9000 TEU, with a further increase by 70% by 2015. The largest operational container vessels have a capacity in excess of 12000 TEU. A similar trend is visible in the Ro- Ro sub-markets of car carrying, ferry market and unaccompanied freight.

and most European countries. In countries with a decreasing tendency in number of vessels, only France and Italy show the same decreasing tendency for gross registered tonnage. Italy shows a big fall in the last year (2007), but FROM 2000 to 2006 it has been progressively rising. France shows the opposite tendency: it had been slowly decreasing in the period 2000-2006, but in 2077 it began to grow. The most important growth is shown by Portugal and Spain (39% and 33%, respectively). Nevertheless, to have an overall view of the evolution of vessel traffic, we have analyzed some indexes showing the evolution for tones/number of vessels. In both cases (total vessels and containers), we can appreciate –see table 6- an increasing tendency within the European Union. Once having analyzed the previous data, we have clearly seen the increase in vessel size and in cargo circulation. However the same concentration level remains steady.

Finally, to complete this analysis, we have studied the level and evolution of container traffic in the top 20 ports of EU. We have noticed an increasing traffic of containers, but it's not linked to changes in concentration levels in the 20 top ports; that is to say, the distribution of freights among ports remains the same. In fact, we realized that the percentage of participation in the whole traffic is almost constant in all ports. There are two important characteristics. The first one is a possible slight signal of concentration for the three top container ports: Rotterdam (NL), Hamburg (DE) and Antwerp (BE); and the second one is the special behaviour of Algeciras (ES), because it does not show a clear tendency. These oscillations in level of cargo are not explained by any other Spanish port in the Top-20 schedule, because neither Barcelona, nor Bilbao, nor Valencia, nor Las Palmas have special oscillations. What is happening here is exactly quite the opposite. Through the observation of data for Spanish ports, it is possible to prove exactly what we are stating: Barcelona, Bilbao, Las Palmas and Valencia, portray a steady maritime traffic with a slight increasing tendency towards the end of the period of study. None of them seems to absorb the Algeciras variations, as these oscillations are explained for the Eurostat methodological notes, where it is reported that data for Algeciras are underestimated in 2004. If we accept this explanation, then Algeciras behaviour is the same as for the other ports. To further explain, we have studied the concentration index for top-20 ports, which are shown in tables 7 and 8 in the attachment. We have used the CR index and Herfindahl-Hirschman index. Both of indexes got similar results: There is a relatively high level of concentration in 3 or 4 top ports, but it seems to stay in the same values along the estimated period. Regarding the concentration index, a steady level of concentration is shown; it even seems more like a decreasing trend than an increasing one. The evolution of each port separately is observed in the next table using the CRi concentration index. Combining the information offered by both indexes in the most favourable case for defenders of increasing concentration it would be possible to conclude that there may be some concentration levels concerning CR4 or other values of CR index. Port markets have been traditionally perceived as oligopolistic markets (especially due to their own geographical situation). It is significant to highlight that for all those years, the top 4 ports are the same: Rotterdam (NL), Hamburg (DE), Antwerp (BE), Bremen & Bremerhaven (DE) , except in 2000 when the top 4 were formed by Rotterdam (NL), Hamburg (DE), Felixstowe (UK), Bremen & Bremerhaven (DE). Nevertheless, the levels of cargo volume are really similar.

We have appreciated that concentration levels in the 20 top ports in the EU, together with the large expansion in handled freight, are probably creating congestion problems for these ports, and may be an important reason for not increasing maritime traffic. If our guess turns true, then the next ports on the top list (for example ports between 20 and 40 positions) would be in a situation of absorbing maritime traffics from the smaller ports, because their size could be more adequate to afford the raising goods handled. What has changed is the congestion level. Congestion has had a huge impact on the whole supply chain. Still, we must not dismiss the idea of concentration. This may happen to appear in ports which have not been considered as top ports. The rising concentrate activity may be in full bloom in less important ports and this way proves that this concentration process exists.

Maritime containerised trade is mainly delivered by regular liner services. Lam et al. (2007) argue that the general concentration of the maritime shipping liner industry has developed in parallel to the increase in container ship size. We have analyzed the evolution of number of container ships focussing in it size and we came to the conclusion that in Europe some patterns are changing, but these changes are not in a sustainable trend.

In tables 9 and 10 we can appreciate the increasing trend between 2000 and 2007. It indicates that the most important increase in number of vessels has been for the small containers ships (more than 100 and less than 500) or for the middle sized (more than 6000 and less than 10000), at the same time is show a increasing trend for the biggest sized containers, but it's quite difficult to make a strong deduction because del absolute number is increasing beginning from zero.

If we analyze the same data, but attending to the annual growth data, it's possible to evaluate an unequal trend for different sizes of vessels, nevertheless it helps to make an overview where its shown that the relevant growth for vessels sized from 100 to less than 500 became from a strong rise at the beginning of the period but it began to decrease since 2005.

5. Conclusions

Port organisation has been one of the most debated issues in the field of maritime management and policies. There are no standard models in terms of port property and institutional structure; there are multiple models and varieties. Over time and with a view to ensuring insertion in an increasingly internationalised economy, ports have adopted different administrative systems and institutional management strategies. There is also greater global understanding of the interests at stake, meaning that public and private agents provide services which they excel at, and their public body role is defined by three missions: to catalyse, legalise and facilitate. To summarise, port management models tend to be a combination of three vectors: the degree of public/private participation; the management strategy (centralised or decentralised); the availability of facilities, assets and port services.

Therefore, ports are not separate bodies or within isolated markets; and it is necessary that they are understood as being within the contexts of integrated logistics and supply chains. And equally, although there is no taxonomy that allows port administrators to set down parameters or reference points for interaction with other ports, it is no less certain that the complexity and variety of port business, on more than one level, require a comparative study of organisational, operational, physical and spatial, and finally legal and regulatory differences.

Ports are becoming increasingly "pro-active" in the supply chain; in business transactions, and in management practices. The different European and South American concepts provide evidence of two very different situations (Pérez Fiaño, 2007). Whereas the former has a solid policy and institutional structure, the latter lacks harmonising features. While the former aims to liberalise port services, even though the current position is one of reluctance to the State losing regulating control in port management; the former is easier to push in this direction. If in Europe the emphasis is on investment in the large centres of integration and in Trans-European transport networks, such maximising of potential is limited in South America. And finally, if in Europe the direction for policy and strategies is channelled into establishing sea highways; on the South American continent boosting regional cabotage is still in the inauguration and institutional launching phase.

The Hirschman-Herfindahl concentration index is decreasing from 1997 to 2006. Hence, there are no solid reasons to suspect about increasing concentration. This means that despite maritime traffic in Europe being quite concentrated in some countries, showing the existence of concentration in the maritime transport sector, the levels of concentration do not point out changing signs.

There is an important concentration of freight in main ports for most member states of the European Union, these concentration levels show a very slight rise, and the increasing concentration ratio involves most countries in the European Union. Nevertheless, we can't conclude that there is a growing concentration, contrary to our expectations. Therefore, maybe some increasing concentration is beginning to appear, but more empirical evidence is required to sustain a firm presumption. Focusing on the top 20 ports in the European Union, there is a relatively high level of concentration in

3 or 4 top ports, but it seems to stay in the same values along the period of study. The number of ports equivalent (using HHI concentration index) is around 11 (the total number of ports analyzed are 20). The congestion problems probably affecting the top 20 European ports could translate the research field to the next 20 or 30 top ports, because it's likely to find some increasing concentration level in these big ports, not included in the top 20 lists. The increasing containerization processes appear in the European Union, not only in terms of number of container ships (and also other traffics), but also in terms of volume or transported freight. At the same time there is an increase in vessels size, as ratios GRT/Nº vessels and GRT/Nº of containers are increasing.

References

Allix, Y. (2006), De l´atomisation à l´oligopolisation: stratégies de concentration dans les industries maritimes et portuaires conteneurisées, Ecole de Management de Normandie. Working Paper 43/2006.

Cullinane, K. and Khanna, M. (1999), Economies of Scale in Large Containerships. Journal of Transport Economics and Policy. Vol. 33, Nº 2, pp. 185-208.

European Commission Staff Working Paper 1139 (2007).

European Commission Staff Working Document. Accompanying document to Communication on a European Ports Policy. Full Impact Assessment COM(2007) 616 final. SEC(2007) 1340.

European Commission. (2001), The white paper. European transport policy for 2010: time to decide. Italy.

Foued, A. (2007), La concentration dans le transport maritime des lignes regulieres. Economie Appliquee, December 2007, v. 60, iss. 4, pp. 163-96.

Frémont, A. and Soppé, M. (2004), Les stratégies des armateurs de lignes régulières le desserte maritime et la domination asiatique et le renouveau européen, Seminaire INRETS, Paris.

González-Laxe, F. (2005), Puertos y transportes marítimos: ejes de una nueva articulación global, Revista de Economía Mundial, 12, 123-148.

Henderson, J., Dickens, P., Hess, M., Coe, N. and Yeung, H.W.C. (2002), Global production networks and the analysis of economic development. Review of international Political Economy, Vol. 9, pp. 436-464.

Kent, P.E. and Ashar, A. (2001), Port Competition Regulation. A Tool for Monitoring for Anticompetitive Behavior. International Journal of Maritime Economics. Vol 3, Nº 1, pp. 27-51.

Kuby, M. and Reid, N. (1992), Technological change and the Concentration of the U.S. General Cargo Port System: 1970-88. Economic Geography. Vol. 68. Nº 3, pp. 272-289.

Lam, J.S.L., Yap, W.Y.Y. and Cullinane, K. (2007), Structure, Conduct and Performance on the major liner shipping routes, Maritime Policy and Management. Vol 34, No 4 pp 359-381.

Paixao, A. C. and Marlow, P.B. (2003), Fourth Generation Ports. A question of agility, International Journal of Physical Distribution and Logistics Management, 33(4), 355-376.

Pérez Fiaño, J.E. (2007), Infraestructura portuaria y logística del transporte interoceánico regional. Conf. Foro Iniciativa Cuenca del Pacífico. Cali, 29/I/2007.

Talley, W.K. (2000), Ocean Container Shipping: Impacts of a Technological Improvement.

Appendices

Table 1: Goods Handled in European Union ports (Million Tonnes). (Elaborate with Eurostat data).

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
25 European Union	:	:	:	:	:	:	3.393.255	3.504.690	3.644.361	3.760.121	:
15 European Union	2.887.233	2.951.780	2.930.539	2.974.043	3.037.645	3.091.023	3.188.830	3.304.564	3.433.783	3.545.911	:
13 Euro area	:	:	:	:	2.233.623	2.293.094	2.378.549	2.475.833	2.583.680	2.689.494	:
12 Euro area	2.054.801	2.122.694	2.111.363	2.145.169	2.224.477	2.283.789	2.367.760	2.463.770	2.571.054	2.674.011	:
Belgium	161.621	171.026	165.557	179.381	174.181	173.824	181.110	187.889	206.539	218.941	:
Bulgaria	:	:	:	:	20.192	20.390	21.358	23.125	24.841	27.513	:
Denmark	124.010	104.966	97.213	96.533	93.972	94.283	103.954	100.373	99.688	107.674	:
Germany	213.318	217.388	221.623	242.535	246.050	246.353	254.834	271.869	284.865	302.789	315.051
Estonia	:	:	:	:	40.383	44.682	47.048	44.808	46.546	49.998	:
Ireland	36.333	39.958	42.928	45.273	45.795	44.919	46.165	47.720	52.146	53.326	:
Greece	101.311	110.546	112.549	127.750	122.171	147.692	162.534	157.892	151.250	159.425	:
Spain	270.634	280.254	295.715	234.913	315.120	326.001	343.716	373.065	400.019	414.378	:
France	305.079	319.000	315.153	325.789	318.188	319.032	330.135	334.035	341.470	350.334	:
Italy	434.295	444.956	425.914	446.641	444.804	457.958	477.028	484.984	508.946	520.183	:
Cyprus	:	:	:	:	:	7.220	7.258	6.837	7.290	7.924	:
Latvia	:	:	:	:	56.827	51.978	54.652	54.829	59.698	56.861	61.083
Lithuania	:	:	:	:	20.953	24.405	30.242	25.842	26.146	27.235	29.253
Malta	:	:	:	:	:	:	3.417	3.474	3.503	3.578	3.228
Netherlands	402.162	405.384	395.664	405.802	405.853	413.312	410.330	440.722	460.940	477.238	:
Poland	:	:	:	:	46.210	48.111	51.020	52.272	54.769	53.131	:
Portugal	54.734	57.619	58.794	56.404	56.164	55.599	57.470	59.071	65.301	66.861	68.229
Romania	:	:	:	:	27.619	32.698	35.925	40.594	47.864	46.700	:
Slovenia	:	:	:	:	9.146	9.305	10.788	12.063	12.625	15.483	:
Finland	75.314	76.562	77.467	80.681	96.150	99.099	104.439	106.524	99.577	110.536	:
Sweden	149.892	155.618	156.349	159.291	152.830	154.626	161.454	167.350	178.122	180.487	185.039
United Kingdom	558.530	568.502	565.614	573.050	566.366	558.325	555.662	573.070	584.919	583.739	:

Source: Authors own findings, based on EUROSTAT data

Table 2: Maritime Traffic in European Union ports (Million Tonnes) (inwards): Elaborate with Eurostat data.

25 European Union	:	:	:	:	:	:	2.144.422	2.228.228	2.302.823	2.395.680	:
15 European Union	1.869.649	1.942.046	1.908.023	1.952.306	2.023.988	2.043.605	2.100.345	2.182.351	2.254.074	2.333.307	:
13 Euro area	:	:	:	:	1.567.219	1.594.938	1.638.796	1.703.599	1.759.401	1.824.537	:
12 Euro area	1.423.952	1.491.640	1.463.280	1.496.263	1.560.536	1.588.254	1.631.088	1.694.807	1.750.359	1.813.944	:
Belgium	100.028	110.885	102.476	110.947	106.970	104.642	103.940	108.946	116.603	125.486	:
Bulgaria	:	:	:	:	11.927	11.600	12.696	13.614	14.293	16.250	:
Denmark	71.644	59.153	54.429	52.876	51.678	50.418	56.908	54.568	53.545	59.682	:
Germany	140.468	144.912	143.149	152.210	156.476	153.935	159.202	168.490	172.257	182.246	192.027
Estonia	:	:	:	:	3.762	3.718	4.664	4.042	4.557	6.765	:
Ireland	25.574	28.690	30.726	31.679	32.634	32.182	33.234	34.703	37.654	38.563	:
Greece	59.529	66.234	67.350	75.292	73.013	88.468	94.343	92.740	88.173	94.323	:
Spain	184.943	198.032	212.995	171.588	229.394	241.023	249.633	269.906	290.528	298.636	:
France	222.328	234.745	227.433	238.610	231.819	232.788	237.278	240.524	243.747	250.550	:
Italy	298.110	305.675	291.430	315.155	318.414	322.824	334.819	338.374	348.235	358.108	:
Cyprus	:	:	:	:	:	5.619	5.665	5.573	6.019	6.508	:
Latvia	:	:	:	:	2.498	3.319	3.794	3.549	4.679	6.787	7.672
Lithuania	:	:	:	:	3.475	3.797	4.097	4.372	4.710	8.466	10.907
Malta	:	:	:	:	:	:	2.976	3.027	3.321	3.347	3.077
Netherlands	313.210	320.015	304.167	315.850	317.330	318.088	318.489	339.510	350.952	358.912	:
Poland	:	:	:	:	14.684	14.943	15.171	16.522	16.420	19.906	:
Portugal	40.685	43.431	45.409	43.820	43.779	42.758	42.794	43.135	47.473	46.886	47.055
Romania	:	:	:	:	14.546	15.412	18.658	22.371	25.168	24.556	:
Slovenia	:	:	:	:	6.684	6.683	7.709	8.792	9.043	10.594	:
Finland	39.076	39.020	38.145	41.110	50.706	51.546	57.356	58.479	54.737	60.235	:
Sweden	79.006	85.154	83.350	86.824	82.885	84.105	88.583	90.552	95.780	94.569	97.106
United Kingdom	295046	306100	306963	316344	328889	320828	323766	342425	354391	365112	:

Source: Authors own findings, based on EUROSTAT data

Table 3: Maritime Traffic in European Union ports (Million Tonnes) (outwards): Elaborate with Eurostat data.

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
25 European Union	:	:	:	:	:	:	1.248.833	1.276.462	1.341.538	1.364.441	:
15 European Union	1.017.584	1.009.733	1.022.516	1.021.736	1.013.658	1.047.418	1.088.485	1.122.212	1.179.710	1.212.604	:
13 Euro area	:	:	:	:	666.404	698.156	739.753	772.234	824.278	864.957	:
12 Euro area	630.848	631.054	648.083	648.906	663.942	695.535	736.673	768.963	820.696	860.067	:
Belgium	61.592	60.140	63.080	68.434	67.211	69.182	77.170	78.943	89.936	93.455	:
Bulgaria	:	:	:	:	8.266	8.790	8.661	9.511	10.548	11.263	:
Denmark	52.366	45.813	42.784	43.656	42.294	43.865	47.046	45.806	46.143	47.991	:
Germany	72.850	72.476	78.474	90.325	89.574	92.418	95.632	103.379	112.608	120.543	123.023
Estonia	:	:	:	:	36.621	40.964	42.383	40.767	41.988	43.234	:
Ireland	10.758	11.268	12.202	13.594	13.161	12.737	12.931	13.017	14.492	14.763	:
Greece	41.783	44.313	45.199	52.457	49.158	59.225	68.191	65.152	63.078	65.102	:
Spain	85.691	82.222	82.720	63.324	85.726	84.978	94.083	103.159	109.492	115.742	:
France	82.751	84.255	87.720	87.179	86.368	86.244	92.858	93.511	97.723	99.784	:
Italy	136.184	139.281	134.484	131.486	126.390	135.134	142.209	146.610	160.711	162.076	:
Cyprus	:	:	:	:	:	1.601	1.593	1.264	1.271	1.416	:
Latvia	:	:	:	:	54.329	48.659	50.858	51.281	55.019	50.074	53.412
Lithuania	:	:	:	:	17.478	20.608	26.146	21.470	21.435	18.769	18.346
Malta	:	:	:	:	:	:	441	447	182	230	151
Netherlands	88.952	85.369	91.498	89.952	88.523	95.224	91.841	101.212	109.988	118.326	:
Poland	:	:	:	:	31.526	33.168	35.848	35.751	38.350	33.225	:
Portugal	14.049	14.188	13.385	12.584	12.386	12.841	14.676	15.936	17.828	19.975	21.174
Romania	:	:	:	:	13.073	17.286	17.267	18.223	22.696	22.144	:
Slovenia	:	:	:	:	2.462	2.621	3.080	3.271	3.583	4.889	:
Finland	36.237	37.543	39.322	39.571	45.444	47.553	47.083	48.045	44.840	50.301	:
Sweden	70.885	70.465	72.998	72.467	69.945	70.521	72.870	76.798	82.342	85.918	87.933
United Kingdom	263484	262402	258651	256706	237477	237497	231896	230645	230529	218627	:

Source: Authors own findings, based on EUROSTAT data

Table 4: Hirschma-Herfindalh Concentration Index for some European Countries. Own elaboration based on Eurostat data.

Hirshman- Herfindalh Index	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Goods Handled	0,115	0,115	0,114	0,114	0,112	0,111	0,109	0,11	0,11	0,108
Inwards	0,05	0,052	0,05	0,052	0,053	0,052	0,05	0,051	0	0,051
Outwards	0,126	0,127	0,123	0,122	0,116	0,114	0,11	0,108	0,107	0,103

Source: Authors own findings, based on EUROSTAT data

Table 5: Handled goods in main ports/Handled goods in all ports (%). Own elaboration based on Eurostat data.

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
25 European Union	NA	NA	NA	NA	NA	NA	NA	NA	80,62%	80,74%
15 European Union	NA	NA	NA	82,05%	80,34%	81,54%	81,49%	81,78%	82,20%	82,26%
13 Euro area	NA	NA	NA	NA	NA	87,59%	87,65%	87,69%	87,52%	87,37%
12 Euro area	NA	NA	NA	89,11%	85,93%	87,55%	87,60%	87,76%	87,60%	87,49%
Belgium	98,12%	98,18%	98,05%	98,06%	100,98%	98,39%	98,56%	98,65%	98,72%	98,66%
Bulgaria	NA	NA	NA	NA	0,00%	99,56%	100,00%	100,02%	100,00%	99,99%
Denmark	71,96%	74,12%	73,81%	76,66%	78,75%	77,99%	76,73%	80,27%	82,39%	81,82%
Germany	96,65%	96,90%	96,74%	96,02%	94,65%	95,91%	97,11%	96,90%	97,23%	97,26%
Estonia	NA	NA	NA	NA	NA	99,98%	96,24%	100,00%	96,79%	95,10%
Ireland	78,49%	77,18%	75,23%	76,19%	75,32%	89,59%	89,43%	89,18%	88,11%	89,47%
Greece	99,41%	99,17%	99,37%	75,30%	78,73%	65,95%	68,47%	68,28%	69,22%	67,75%
Spain	NA	NA	NA	94,47%	70,43%	93,53%	93,37%	93,23%	93,38%	93,35%
France	NA	95,95%	95,27%	98,84%	101,20%	95,11%	95,55%	95,04%	95,40%	95,81%
Italy	81,41%	82,88%	82,68%	84,47%	84,82%	85,17%	85,06%	85,60%	83,41%	83,70%
Cyprus	NA	NA	NA	NA	NA	NA	NA	NA	100,69%	97,70%
Latvia	NA	NA	NA	NA	NA	98,07%	98,47%	97,96%	97,90%	97,99%
Lithuania	NA	NA	NA	NA	NA	100,07%	NA	100,00%	100,00%	100,00%
Malta	NA	NA	NA	NA	NA	NA	99,97%	100,00%	100,00%	99,97%
Netherlands	99,39%	99,29%	99,99%	99,37%	99,35%	99,59%	99,62%	99,81%	99,71%	99,80%
Poland	NA	NA	NA	NA	NA	NA	NA	99,23%	99,19%	99,11%
Portugal	87,03%	86,19%	86,99%	89,91%	90,29%	90,50%	89,11%	89,95%	86,87%	89,17%
Romania	NA	NA	NA	NA	NA	92,48%	94,36%	100,00%	100,00%	98,51%
Slovenia	NA	NA	NA	NA	NA	99,37%	99,38%	99,36%	99,33%	99,41%
Finland	93,27%	93,64%	91,81%	93,45%	78,41%	88,11%	88,32%	88,92%	89,15%	88,53%
Sweden	81,26%	81,23%	83,04%	82,92%	86,43%	82,00%	82,75%	83,21%	85,11%	84,77%
United Kingdom	NA	NA	NA	88,08%	89,12%	88,37%	88,50%	88,60%	88,17%	89,21%

Source: Authors own findings, based on EUROSTAT data

Table 6: Vessel Traficc in European Union. Some index.

VESSEL TRAFICC - SOME INDEX	2.000	2.001	2.002	2.003	2.004	2.005	2.006	2.007
NUMBER OF VESSELS								
European Union (15 countries)	1.763.454	1.776.354	1.907.889	1.993.960	2.062.587	1.981.640	2.028.907	1.884.257
Euro area (BE, DE, IE, GR, ES, FR, IT, LU, NL, AT, PT, FI)	1.121.141	1.164.189	1.292.909	1.386.925	1.456.345	1.382.437	1.438.297	1.278.926
NUMBER OF CONTAINERS								
European Union (15 countries)	57.688	68.440	74.353	79.093	84.037	83.248	87.226	88.738
Euro area (BE, DE, IE, GR, ES, FR, IT, LU, NL, AT, PT, FI)	48.918	59.455	64.977	69.787	74.312	73.224	76.630	77.140
GROSS REGISTRERED TONNAGE (GRT) TOTAL VESSELS								
European Union (15 countries)	10.411.675	10.942.330	11.946.228	12.314.558	12.656.071	12.731.882	13.100.681	13.130.157
Euro area (BE, DE, IE, GR, ES, FR, IT, LU, NL, AT, PT, FI)	6.351.775	6.887.394	7.830.702	8.133.310	8.377.461	8.501.784	8.887.797	8.707.924
GROSS REGISTRERED TONNAGE (GRT) CONTANINERS								
European Union (15 countries)	942.720	1.152.647	1.309.390	1.324.018	1.486.700	1.488.305	1.653.561	1.658.302
Euro area (BE, DE, IE, GR, ES, FR, IT, LU, NL, AT, PT, FI)	771.071	966.453	1.109.317	1.118.283	1.261.171	1.270.380	1.414.834	1.408.096
GRT/Nº VESSELS								
European Union (15 countries)	5,90	6,16	6,26	6,18	6,14	6,42	6,46	6,97
Euro area (BE, DE, IE, GR, ES, FR, IT, LU, NL, AT, PT, FI)	5,67	5,92	6,06	5,86	5,75	6,15	6,18	6,81
GRT/Nº CONTAINERS								
European Union (15 countries)	16,34	16,84	17,61	16,74	17,69	17,88	18,96	18,69
Euro area (BE, DE, IE, GR, ES, FR, IT, LU, NL, AT, PT, FI)	15,76	16,26	17,07	16,02	16,97	17,35	18,46	18,25

Source: Authors own findings, based on EUROSTAT data

Table 7: CRi Concentration Index for Top-20 Ports in EU (Ranking 2006)

TOP-20 CONTAINER PORTS IN 2006 CRi concentration Index							
	2000	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**
1	19,4%	17,0%	17,1%	16,8%	17,7%	17,4%	16,8%
2	32,7%	30,2%	31,2%	31,2%	32,8%	32,6%	32,4%
3	41,5%	38,6%	39,5%	40,6%	43,7%	44,3%	44,1%
4	49,7%	47,0%	47,5%	48,1%	51,3%	51,4%	52,0%
5	57,9%	54,9%	55,0%	52,9%	58,1%	57,4%	57,8%
6	65,8%	61,7%	62,1%	58,7%	63,9%	63,3%	63,1%
7	70,2%	66,6%	66,9%	66,0%	68,6%	68,5%	68,0%
8	74,3%	70,9%	71,5%	70,8%	73,2%	73,1%	72,6%
9	78,4%	75,2%	76,0%	74,9%	77,7%	77,1%	76,7%
10	82,0%	79,5%	80,0%	79,7%	81,0%	81,0%	80,4%
11	85,4%	83,4%	83,7%	82,9%	84,1%	83,7%	83,0%
12	88,8%	86,8%	87,0%	86,7%	87,2%	86,3%	85,5%
13	91,1%	90,1%	89,9%	89,0%	89,6%	88,6%	87,8%
14	93,1%	92,2%	92,1%	89,0%	91,7%	90,5%	89,9%
15	95,2%	94,3%	94,1%	92,7%	93,7%	92,3%	91,9%
16	97,2%	96,2%	96,0%	94,7%	95,5%	94,0%	93,8%
17	98,7%	98,0%	97,9%	96,6%	97,1%	95,6%	95,4%
18	100,0%	99,2%	99,1%	97,7%	98,2%	97,3%	97,0%
19	100,0%	100,0%	100,0%	98,5%	99,2%	98,7%	98,6%
20	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

Source: Authors own findings, based on EUROSTAT data

Table 7 b

TOP-20 CONTAINER PORTS IN 2006 0N THE BASIS OF VOLUME OF CONTAINERS HANDLED IN (1000 TEUs)								
		2000	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**
1	**Rotterdam (NL)**	6.253	6.061	6.505	7.118	8.242	9.195	9.575
2	**Hamburg (DE)**	4.275	4.665	5.376	6.126	7.004	8.084	8.878
3	**Antwerpen (BE) (2)**	2.641	3.001	3.153	4.012	5.055	6.221	6.718
4	**Bremen & Bremerhaven (DE)**	2.643	2.972	3.032	3.191	3.529	3.741	4.504
5	**Algeciras (ES) (3) (4)**	NA	1.737	1.732	2.024	970	3.184	3.262
6	**Felixstowe (UK)**	2.825	2.839	2.682	2.482	2.717	2.760	3.030
7	**Gioia Tauro (IT)**	2.575	2.393	2.883	3.094	3.170	3.123	2.835
8	**Valencia (ES) (4)**	1.313	1.512	1.826	2.012	2.156	2.415	2.615
9	**Barcelona (ES) (4)**	1.389	1.404	1.122	1.765	2.084	2.071	2.315
10	**Le Havre (FR)**	1.334	1.550	1.754	2.015	2.158	2.144	2.119
11	**Southampton (UK)**	1.092	1.213	1.275	1.375	1.435	1.384	1.502
12	**Piraeus (EL)**	1.096	1.164	1.395	1.606	1.551	1.401	1.413
13	**Las Palmas (ES) (4)**	648	664	726	966	1.111	1.222	1.303
14	**Constanta (RO)**	NA	NA	NA	NA	391	867	1.170
15	**Genova (IT)**	1.179	1.536	1.499	1.591	1.437	1.038	1.146
16	**La Spezia (IT)**	661	758	780	836	879	916	1.086
17	**Marseille (FR)**	725	745	811	835	920	911	950
18	**Bilbao (ES) (4)**	425	447	454	468	498	863	899
19	**Zeebrugge (BE)**	488	279	329	328	458	682	895
20	**Göteborg (SE)**	652	624	725	634	722	772	812
Total top 20 ports (5)		36.588	39.168	43.706	47.380	53.077	52.994	57.027

(1) TEU = Twenty-foot Equivalent Unit (unit of volume equivalent to a 20 foot ISO container).

(2) Partial data up to 2nd quarter 2004.

(3) Data for 2004 are underestimated.

(4) Data for the period 2003-2005 are provisional and likely to be revised.

(5) Information about the ports being part of the top 20 ports during the reference year concerned. The composition of the top 20 changes over time

Table 2: Statistics in focus 62/2008. Eurostat

Table 8: Herfindal - Hirschmann Concentration Index for Top-20 Ports in EU (Ranking 2006)

Year	Herfindahl-Hirschmann Index	Number of equivalent ports
2000	0,09	10,66
2001	0,08	11,80
2002	0,09	11,59
2003	0,09	11,60
2004	0,09	10,94
2005	0,09	10,99
2006	0,09	11,08

Source: Authors own findings, based on EUROSTAT data

Table 9

Maritime transport - Vessel traffic - Main ports - Number of container ships by size of vessels - Direction: inwards only									
	2000	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**% Growth 2000-2007**
Total	57.688	68.440	74.353	79.093	84.037	83.248	87.226	91.983	59%
Less than 100	0	3	2	2	2	1	3	14	---
From 100 to less than 500	276	734	2.263	5.987	6.036	4.135	3.493	3.395	1130%
From 500 to less than 1000	462	106	69	61	88	289	73	55	-88%
From 1000 to less than 2000	1.211	788	4.481	5.752	6.520	7.815	8.031	7.474	517%
From 2000 to less than 3000	6.213	6.247	6.382	6.657	7.757	6.877	5.385	4.423	-29%
From 3000 to less than 4000	7.530	11.368	9.080	9.126	8.961	8.283	6.556	7.182	-5%
From 4000 to less than 5000	3.659	3.992	3.272	2.852	2.571	3.399	4.356	3.931	7%
From 5000 to less than 6000	4.152	3.936	3.320	2.778	3.349	3.578	5.334	5.569	34%
From 6000 to less than 7000	3.949	5.960	6.415	7.314	7.561	7.254	7.488	8.250	109%
From 7000 to less than 8000	1.534	1.702	1.510	2.145	3.251	3.756	5.123	7.843	411%
From 8000 to less than 9000	550	1.579	2.458	2.261	2.039	2.373	3.087	3.630	560%
From 9000 to less than 10000	2.222	2.235	3.454	3.036	3.272	3.485	3.868	4.987	124%
From 10000 to less than 20000	7.788	8.286	8.351	8.859	8.280	8.310	9.139	9.820	26%
From 20000 to less than 30000	4.882	5.046	4.371	4.286	5.842	5.955	6.132	6.655	36%
From 30000 to less than 40000	4.678	4.980	5.409	4.631	6.112	5.519	5.666	5.475	17%
From 40000 to less than 50000	2.768	2.577	2.252	2.762	3.531	2.874	2.868	2.042	-26%
From 50000 to less than 80000	3.659	3.931	5.312	6.009	7.563	7.769	7.955	7.545	106%
From 80000 to less than 100000	618	835	1.000	1.121	1.287	1.550	2.320	2.976	382%
From 100000 to less than 150000	0	0	0	0	0	22	331	595	---
From 150000 to less than 200000	0	0	0	1	0	0	12	117	---
From 200000 to less than 250000	0	1	0	0	0	0	0	0	---
From 250000 to less than 300000	2	0	0	1	1	2	0	1	-50%
More than 300000	0	1	0	1	3	1	1	0	---
Unknown	1.535	4.133	4.952	3.451	11	1	5	4	-100%

Source: Autors own findings based on Eurostat data.

Table 10

Maritime transport - Vessel traffic - Main ports - Number of container ships by size of vessels - Direction: inwards only - % Anual growth							
	2001	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**
Total	18,6%	8,6%	6,4%	6,3%	-0,9%	4,8%	5,5%
Less than 100	---	-33,3%	0,0%	0,0%	-50,0%	200,0%	366,7%
From 100 to less than 500	165,9%	208,3%	164,6%	0,8%	-31,5%	-15,5%	-2,8%
From 500 to less than 1000	-77,1%	-34,9%	-11,6%	44,3%	228,4%	-74,7%	-24,7%
From 1000 to less than 2000	-34,9%	468,7%	28,4%	13,4%	19,9%	2,8%	-6,9%
From 2000 to less than 3000	0,5%	2,2%	4,3%	16,5%	-11,3%	-21,7%	-17,9%
From 3000 to less than 4000	51,0%	-20,1%	0,5%	-1,8%	-7,6%	-20,8%	9,5%
From 4000 to less than 5000	9,1%	-18,0%	-12,8%	-9,9%	32,2%	28,2%	-9,8%
From 5000 to less than 6000	-5,2%	-15,7%	-16,3%	20,6%	6,8%	49,1%	4,4%
From 6000 to less than 7000	50,9%	7,6%	14,0%	3,4%	-4,1%	3,2%	10,2%
From 7000 to less than 8000	11,0%	-11,3%	42,1%	51,6%	15,5%	36,4%	53,1%
From 8000 to less than 9000	187,1%	55,7%	-8,0%	-9,8%	16,4%	30,1%	17,6%
From 9000 to less than 10000	0,6%	54,5%	-12,1%	7,8%	6,5%	11,0%	28,9%
From 10000 to less than 20000	6,4%	0,8%	6,1%	-6,5%	0,4%	10,0%	7,5%
From 20000 to less than 30000	3,4%	-13,4%	-1,9%	36,3%	1,9%	3,0%	8,5%
From 30000 to less than 40000	6,5%	8,6%	-14,4%	32,0%	-9,7%	2,7%	-3,4%
From 40000 to less than 50000	-6,9%	-12,6%	22,6%	27,8%	-18,6%	-0,2%	-28,8%
From 50000 to less than 80000	7,4%	35,1%	13,1%	25,9%	2,7%	2,4%	-5,2%
From 80000 to less than 100000	35,1%	19,8%	12,1%	14,8%	20,4%	49,7%	28,3%
From 100000 to less than 150000	---	---	---	---	---	1404,5%	79,8%
From 150000 to less than 200000	---	---	---	---	---	---	875,0%
From 200000 to less than 250000	---	-100,0%	---	---	---	---	---
From 250000 to less than 300000	-100,0%	---	---	0,0%	100,0%	-100,0%	---
More than 300000	---	-100,0%	---	200,0%	-66,7%	0,0%	-100,0%
Unknown	169,3%	19,8%	-30,3%	-99,7%	-90,9%	400,0%	-20,0%

Source: Autors own findings based on Eurostat data.

Analysis of random disruptive events in shipping and port operations

Saut Gurning and Stephen Cahoon

Department of Maritime Logistics and Management, Australian Maritime College, University of Tasmania, Australia.

Abstract

Dynamic demands of global maritime services and existing maritime infrastructure constraints imposed by a rapidly changing business environment make it increasingly necessary for service providers in the maritime industry to be aware of the interruption consequences from various factors. The main questions individual maritime companies face is whether, why, how, and with whom they should interact if one particular disruptive event takes place along their supply-chain management activities. This paper discusses various maritime disruption events based on a multi-disciplinary literature review. It also provides a qualitative method for analysing stages of disruptive events in the maritime stage of the supply-chain including the identification of direct and indirect driving factors. Costs and time analyses are presented based on supply-chain management (SCM) and logistics processes in assessing the impact of maritime disruptions. The application of this analysis in recognising maritime disruptions may provide a valuable tool for mitigating the dynamics of the shipping and port markets, as it indicates potentially effective redesign strategies when a specific source of uncertainty is encountered in a supply-chain. In this paper we study the events of maritime disruptions driven by factors of disruption in port and shipping operations.

Keywords: maritime disruptions, supply-chain management.

1. Introduction

Supply-chain operators are facing a variety of uncertainties on their particular leg of the supply-chain due to various internal and external factors interrupting the shipment of goods from the source to the planned destinations. For example Handfield and Maccormark (2008), Gaonkar and Viswandham (2007) and Rodrigues et al. (2008) argue that global outsourcing strategies implemented by dominant international supply-chain operators create uncontrollable deviations from the original plan even though disruptions may occur on each specific stages of the chain flow. This includes the maritime related processes - as a transportation cluster service in the downstream operation of supply-chain management - that experience a range of uncertainties. These variables are the risks associated with the transport of cargo in shipping- related industries such as the shortage of ship fleets for one particular route or cargo at a certain period of time, natural hazard conditions at sea such as severe wave height or a supply/demand imbalance of containers including container imbalance (Jula et al., 2006; Lin and Tseng, 2007).

Further possible factors of uncertainty in port and shipping-related services creating disruption in the supply-chain include congestion (as in Cullinane and Song, 1998; Pettitt, 2007), strikes of stevedores or port workers (for example Nathan, 2005; Peter, 2000), and security related incidents at the port (Barnes and Oloruntoba, 2005; Carafano, 2007; Van De Voort et al., 2003). These uncertainties subsequently may generate chaotic conditions along the chain which commonly results in four categories of outcomes. These are delays, deviations, disruptions, and disasters which may produce unavoidable divergences from the original plans of supply-chain operations. These criteria may also be regarded as the stages through which disturbances evolve, in their stage from a recurrent risk of individual events to destructive events. Stage One or the first stage through which a risk passes is referred to as the delay stage; here the focus is on the recurrent changes displayed by the performance of a supply-chain and the cancelation of previous planned (Wright, 2008; Zsidisin and Smith, 2005). Stage 2 is the deviation stage when one or more performance parameters of a supply-chain diverge from their estimated or mean value, without significant transformation to original supply-chain structure (Gaonkar and Viswanadham, 2007). Stage 3 is the disruption stage which occurs when existing services are unavailable due to direct and indirect factors interrupting the services' provisions (Gaonkar and Viswanadham, 2007; Handfield et al., 2008; Yu and Qi, 2004). The last stage is the disaster stage which results in a service platform being damaged and as a consequence, supply-chain entities are unable to provide services (Chang, 2000; Gaonkar and Viswanadham, 2007).

If uncertainties in a supply-chain are not responded to appropriately and in a timely fashion, they may severely affect the availability of specific cargoes and impact on the supply-chain operators' functions in terms of financial performance, market position and the competition level of its entities (as in Homan, 2007). For example, maritime operations may be categorised as a critical node if specific disruption events occur in a port area, shipping operations and forwarding services. Its function to interface and integrate transportation and distribution operations is essential. Furthermore, due to the global services linking supply-chain networks, the propagation effects of maritime disruption as initial stage passed to other stages either downstream or upstream level of supply-chain are presumably enormous (Craighead et al., 2007) .

2. The Drivers of Supply-Chain Disruption events in the Maritime Leg

Supply-chain disruption in maritime operations has not been discussed intensively as a disruption management topic in an overall context of the supply-chain. In this study, 18 variables are considered as dominant factor that can impact maritime stage in the supply-chain derived from supply-chain literatures and also as a result of exploration study of wheat supply-chain between Australia and Indonesia in 2006-2008. Table 1 lists the 18 variables reported in relation to maritime disruptions between 1986 and 2008 including the causal factors identified and reported by each study. From the table it is found that the factors of severe weather condition and port strikes were cited more frequently than any of the other categories in the literature. Severe weather condition and port strikes, therefore, may be classified as minor disruptions (Bearing-Point and Hewlett-Packard, 2005) as they occur frequently. The next most cited factors classified as major disruptions were port congestion, security issues, and earthquake. These particular events are more related under operational and external-based disruptions. From these studies, minor-based disruptions tend to precede more occurrence disruptions as it created more unavailability of maritime services compared to major-based disruptons.

In a port disruption report by Frittelli (2005) it was found that disruption events (by Katrina hurricane in 2005) had significantly reduced 50 percent of port' service level, resumed its partial service after six months (particularly on cargo handling operation) due to the damage of handling cranes, considerable damaged of port buildings, railroads, key truck-link, bridges, and shipping channels in the Port of New Orleans, Baton Rouge, South Louisiana, Gulfport, Mobile, Alabama, and Pascagoula. Further, Coy et al. (2005), Frittelli (2005), and Shultz (2006) explored the effect of operational disruptive events in relation to ship handling, warehousing, cargo loading and inland transport services in ports had deviated the performance of grain supply-chain from US to various international markets. Similar to this, Australian Transport Safety Bureau (ATSB) (2007) reported the impact of disruptive events (by major storms) caused considerable higher interruptive occurrence of Newcastle Port services in 2007 compared to the period of 2004-2006. The disruptive conditions during that year generated a significant increase of ship queue numbers, interrupted rail connection (because of flooding) and the closure of the port as a result of ship grounding in the port channel.

Table 1: Researchers and Factors Identified / Discussed

		Akaha, 1986	Rose et al., 1997	Chang, 2000	Ramage, 2001	Watanabe, 2002	Connolly, 2004	Alexander and Irwin, 2005	BearingPoint, 2005	Chu and Hansen, 2005	Coy et al., 2005	Davis et a.,l 2005	Early-Warning, 2005	Frittelli, 2005	Kleindorfer and Saad, 2005	Arnold et al., 2006	CSA, 2006	Shultz, 2006	Auckland, 2007	Business Line, 2007	CBC, 2007	Conrad et al., 2007	Drum, 2007	Heger, 2007	Mahbub, 2007	Pettit, 2007	Reuters, 2007	Yank, 2007	Brown, 2008	Gurning, 2008	Horlock, 2008	Robert, 2008	Saidi, 2008	Seba, 2008	Stratfor, 2008	Todd, 2008	Tsukimori and Jukwey, 2008	Wright, 2008	Nightingale, 2008	**Total**
No.	**Disruptive Events**	**Researchers and Factors Identified / Discussed**																																						
1	**Security issues**						•		•							•						•																		4
2	**Political events**	•										•															•													3
3	**Rail related operation**												•								•																			2
4	**Port strikes**					•			•	•									•		•	•									•							•		**8**
5	**Customs & administration**								•											•		•																		3
6	**Severe weather condition**							•	•		•			•	•			•					•		•			•	•	•				•						**12**
7	**Earthquakes**		•	•											•														•											4
8	**Electrical outages**																					•													•					2
9	**Equipment down / shortage**																															•	•							2
10	**Empty containers**																																			•				1
11	**Ship accidents in port areas**				•				•															•																3
12	**Shipping-Port disputes**																•																							1
13	**Port congestion**								•																	•							•				•			4
14	**Ship shortages**								•									•												•										3
15	**Fuel and bunkering Costs**																	•																						1
16	**Inland accesibilty problems**																													•							•			2
17	**Telecomunication system**																					•																		1
18	**Shortage of service demand**																																						•	1

For shipping related services, various disruptive factors may create a short-term unavailability of specific route or networks, shortage of transport capacity leading to higher shipping costs, and longer distribution time. For example, a study of Cyclone impact on LNG and crude-oil terminals in Iran (Qalhat terminal) and Oman (Mina Al Fahal) by Yank (2007) informed that the closure of these terminals for 14-18 days for LNG loading operation interrupted oil and LNG shipping routes from those resources. Further, an exploration study of 42 cases of maritime disruptive events in Indonesia during 2006-2007 found that various disruptive factors (mainly because of severe weather conditions, earthquakes, higher bunkering costs) created the unavailability of certain shipping services from 3 days to two weeks in ferry transport, domestic oil transportation, coal transfer and food transport within the islands (Gurning, 2008a; Gurning, 2008b; Mahbub, 2007).

Those possible factors may be categorised into two group of disruption risks (as listed in the table 2) namely direct and indirect factors due to its propagation effects (Blackhurst et al., 2004; Zsidisin et al., 2004) and the multi-level of disruption factors (Jones and Towill, 1998; Peck, 2006). Factors or events that may generate initial delays, deviations and eventually disruptions within the boundaries of maritime operations can be defined as direct maritime disruption risk. Events related to security and safety risks, service related risks, and infrastructure risks are categories of disruptive drivers under this category. In contrast, disruptive events indirectly stimulated from external aspects of maritime boundaries, subsequently are defined as indirect maritime disruptions such as market risks, organisation and relationship risks, and environmental risks.

Table 2: The Categories of Maritime Disruption Risks

TYPE OF MARITIME DISRUPTION RISKS	
DIRECT	**INDIRECT**
Security and safety	**Market risks**
- Ship accidents	- Shortage of Demand
- Ship pollutions	- Shortage of ships
- Political events	- Financial Crisis
- Terrorist attack	- Trade imbalance
Service related risks	**Organisation and relationship**
- Operational and equipment	- Employment / Port Workers
- Competition	- Legal and policy
- Fuel and bunkering	- Resources
- Electrical shortage	- Customs process
- Congestion	- Ships inspections
- Infrastructure related risks	**Environmental risks**
- Communication facility	- Severe weather
- Lack of development	- Earthquakes
- Inland transport connections	- Flood

3. Maritime disruption case of wheat-chain between Australia and Indonesia

The wheat-chain between Australia and Indonesia in the period of 2006-2008 experienced various uncertainty factors mainly due to natural reasons which tended to be generated by maritime disruptive events. These disruptions created significant higher prices of wheat in the raw and flour markets, especially to Indonesia's consumers. In 2006-2008, due to the effect of drought, the harvest quantity of Australian wheat was approximately 11 million tons (at the end of 2006) compared to 24 million tons in 2005. The result in January 2006 was the price of hard wheat (APH1 and APH 13) rising to US$ 170 per ton FOB to Indonesia. Further, in October 2006 (as shown in figure 1), the price rose more rapidly to US$ 227 per ton FOB. If the trading term based on CIF in 2007 is compared to the

January 2007 price, the wheat level achieved US$ 326 per ton CNF in contrast to US$ 212 per ton CNF in January 2006 (Drewry, 2007; Fearnsearch, 2007; Gunawan, 2007, pers.comm). The natural-based wheat disruption of the drought therefore caused increased prices of wheat in the range of 50 to 60 per cent in 2006-2007.

More significantly, in the dry-bulk shipping operation, the uncertainties of freight rates (categorised as disruptive market factor of wheat supply-chain) contributed considerably to the increase of wheat price in the period of February 2007 – February 2008. The increase of wheat price due to maritime-related operations was nearly 230 per cent (if CNF wheat price is considered). The main reason for this impact is due to the imbalance of dry-bulk shipping market which started in the middle of 2006, subsequently creating a significant increase in the charter rate for the dry bulk fleet especially for Panamax and Handymax (for example Blas, 2007; BULOG, 2007; Clarkson, 2007; WEA, 2008). Between January and September 2007, the cost continued rising from US$54 to US$ 95 per ton representing an increase of approximately 76 per cent (Blas, 2007; Clarkson, 2007). However, millers and wheat traders in Indonesia were not able to respond to this increasing trend directly by increasing the selling price of the flour because of the relatively low ability of Indonesia consumers (Peter, 2007; Siagian, 2007). For example, during 2006-2007, Bogasari Ltd, the biggest wheat-miller in Indonesia, increased the selling price of their flour product to the market by only 12 per cent which was far below the increase of 76 per cent due to rise in shipping freight in January-September 2007.

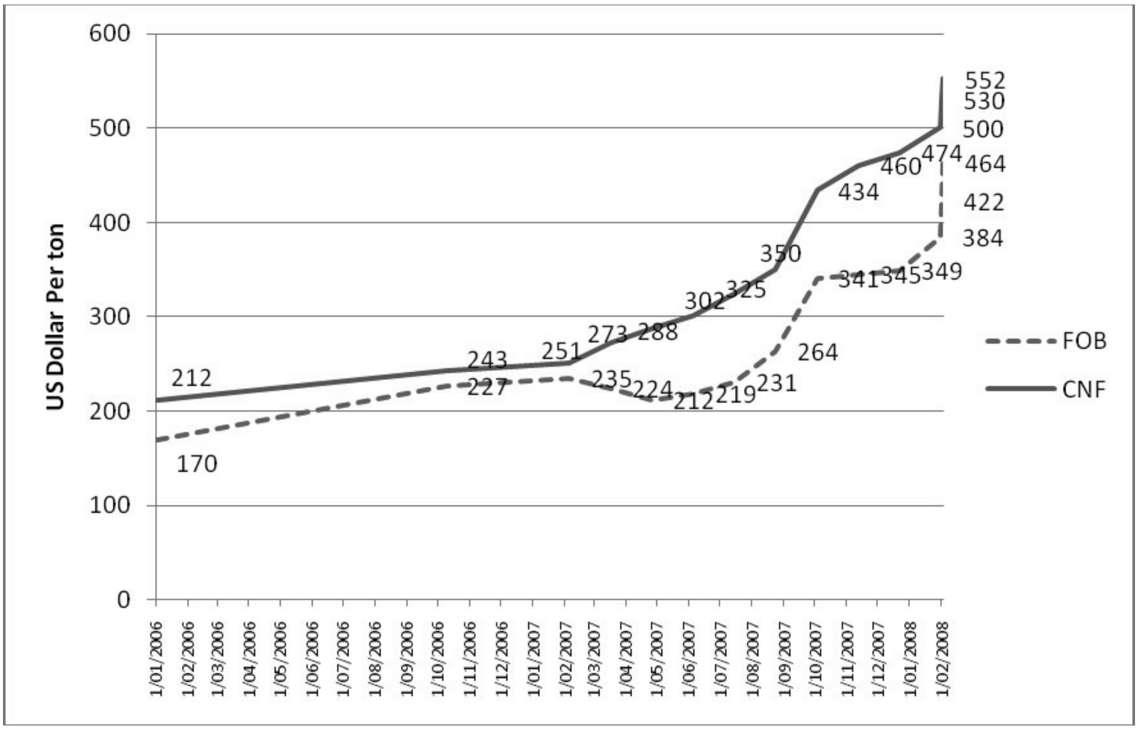


Figure 1: The 2006-2008 Trend of Indonesia Wheat Price from Australia
Source: APTINDO (2008) ; Edijatmiko (2007); FAO (2008); Gunawan (2007); WEA (2008)

In responding to disruptive factors of maritime operations, the wheat industry, especially between Australia and Indonesia, has been using containerised wheat transport in contrast to dry bulk pattern which has, consequently created a substantial problem to wheat trade. Gurning and Grewal (2007) found that using containers for consignment of wheat from Australia to Indonesia was much cheaper compared to dry bulk shipments. Temporarily, bulk shipping and dry bulk terminal operation were presumably no longer able to rely solely on the benefits of exploiting bulk commodities and trades to sustain growth and competitive success compared to containerised shipping due to its loss of economic scale (DFAT, 2007; Ray, 2007).

In relation to the availability of containers, particularly in Australia, there is an imbalance in container trade between Asia and Australia, which results in a continuous repositioning of empties to Asia from Australia. Annually, there are about 200,000 empty containers left in Australia with an estimated five million tons of capacity (WEA, 2007). Hence the transformation of bulk cargo transport to containerised transport has been a reasonable alternative to the higher freight rate of bulk shipping during 2006-2007. In addition, the transportation cost using containers can possibly be reduced further if the volume of container traffic reaches levels at which economies of scale can be obtained (7000-8000 TEU), with the assumption that there will be no supply disruptions due to drought.

The other important impact of maritime related uncertainties is the result of longer transportation time for the wheat distribution process especially in reaching rural consumers in inter-islands' locations in Indonesia. In September 2007, the longest period of time for consumers to wait for the supply of flour and wheat related products in Indonesia was about 30 days (Edijatmiko, 2007; Siagian, 2007). In Australia, the average days of delays in the same period was 14 days due to limitations of two important activities namely rail transportation and loading operation at wheat terminal operations (APTINDO, 2008; WEA, 2007; WEA, 2008). In relation to that, in the period of 2005-2006, Australian wheat farmers complained about the performance of the national rail-system which as being slow and limited in capacity (CSIRO, 2007; DFAT, 2007). Moreover, it was reported that there was a limitation of carrying volume for wheat transportation including an insufficient number of trains allocated to each of the ports of Portland, Geelong, Melbourne and Port Kembla. This was due to a reduction in number of wagons allocated by Pacific National to those ports (as in CSIRO, 2007). The result was a further decrease in the transport capacity of rail facilities for the wheat chain in Australia. CSIRO in 2007 investigated some major bottlenecks in Australian port infrastructure in relation to the wheat chain (Commonwealth Scientific and Industrial Research Organisation (CSIRO) 2007). As reported by CSIRO which surveyed the food logistics progress of Australia in the period of 2004-2006, they found four main problems: (i) the draught problems of certain ports especially those terminals handling wheat based commodities; (ii) the insufficient capacity of the rail track to ports with the need for heavy duty rail; (iii) lack of road access to ports and, (iv) urban encroachment around grain terminals which affects port handling operations. Consequently, congestion has created considerable delays at all Australian dry bulk terminals, averaging about six days in September 2007 (G-Ports, 2007).

Another important factor not considered by CSIRO is the small parcel size for containerized wheat regularly demanded by Indonesian millers. The Indonesian wheat and flour market is controlled by Bogasari Ltd. which has about 75 per cent of market share with the other 25 per cent distributed among six medium and small millers (APTINDO, 2003; APTINDO, 2008). Consequently, these small millers need a particular shipping arrangement for small parcels of imported wheat in the range of 4,000-5,000 tons per month. Therefore, it is neither appropriate nor efficient to use either Handymax or Panamax ships which have a minimum carrying capacity of more than 40,000 tons per shipments (Peter, 2007; Siagian, 2007; Wheat Exports Australia (WEA) 2007). In response to this, the common strategy implemented to date by wheat shippers and consignees is to share the space on a dry bulk ship. Another alternative is to use bags or FIBC (flexible intermediate bulk container) for bulk movement (APTINDO, 2007).

In Indonesia, the disruptive factors in the maritime leg that further impact on domestic wheat distribution include the congestion problems of inland access to and from the main ports of Indonesia, and problem of inter-island networks of ferry and short sea-shipping services (as shown in Figure 2). In relation to wheat trade, Indonesia has more complex problems with port infrastructure than Australia due to limitations of port draught, narrow channel problems, and high port waiting levels caused by the lack of terminal availability. In terms of inland transport systems, this is the most disadvantaged area for Indonesia because of the lack of rail facility and connections to support port operations in general and the wheat handling process in particular. Consequently, inland transferring operations at Indonesia's main ports should rely on trucking services which currently is insufficiently provided in terms of quantities, performance, and costs (for example as discussed in Ditjenhubla, 2006). Moreover, port congestion in Indonesia as discussed by Gurning (2008b, p. 112) was revealed as including a 'lack of terminal storage, limited draught of berth, and lengthy customs procedures' affecting one cargo handling services in dominant ports managed by Indonesian port corporations I-IV. These factors subsequently become major reasons for the increase of terminal handling charges by shipping companies, container handling charges by terminal operators and longer ship time at ports (Ditjenhubla, 2006; Gurning, 2008b; Patunruet al., 2007). In terms of domestic distribution in the period of 2006-2008, it was found that the supply-chain based on wheat products has also been interrupted by natural factors such as severe waves, flood and wind in the Indonesian waters (BMG, 2007; BMG, 2008).

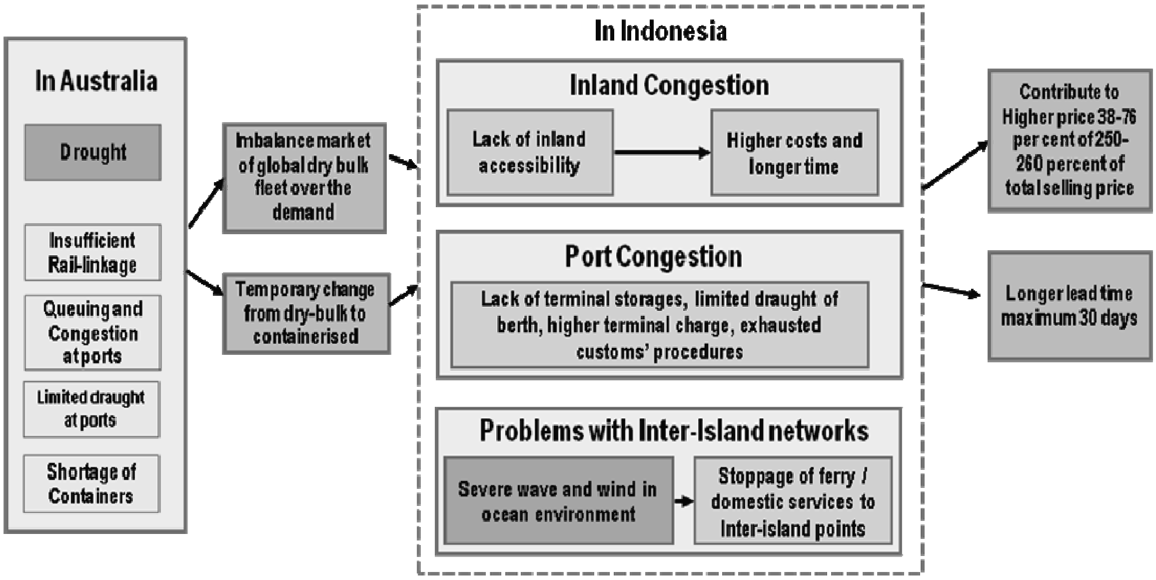


Figure 2: Maritime Disruption Mechanism of Australia-Indonesia Wheat-Chain in 2006-2007

4. Conceptual Strategies of Maritime Disruption Management

The majority of supply-disruption papers (as mentioned in table 3) focus on the combination of contingency rerouting and inventory/sourcing mitigation strategies in response to maritime disruptions (Cavinato, 2004; Craighead et al., 2007; Mccormack, 2008; Tang, 2006; Tomlin, 2006; Zsidisin and Smith, 2005). Derived from these literatures, it may be identified that, the dominant reactions of maritime users in the supply-chain management by adjusting new route of maritime leg and providing strategic stock (no alternative source available) including providing back-up systems are critical initial-steps in disruption risk management and contingency plans for responding to worst case scenarios of maritime disruptions. Gupta (2003), Kleindorfer and Saad (2005) and Sheffi (2001) are the supply-chain disruption papers that consider the mixture of inventory/sourcing mitigation and business continuity concept in managing maritime disruptions. These strategies are recommended for a firm (as maritime user) that faces constant demand and sources from two identical-cost and infinite-capacity suppliers. Further, Blackhurst et al. (2005) and Craighead et al. (2007) emphasised the importance of maritime operators to minimise the ultimate loss from maritime disruptions by considering the trade-off between robustness of the supply chain to maritime disruptions and the overall efficiency under normal operations of maritime services. This may be achieved through risk-sharing decision through insurance plans, redesign of supply-chain flow through a certain maritime leg, and damage control plans (as discussed in Elkins et al., 2008; Howick and Eden, 2001). The goal of managing maritime disruption is to alleviate the consequences of disruptions and risks or, simply put, to increase the *robustness* of a supply chain through maritime leg. However, there are very few qualitative concepts of managing maritime disruptions that concern about the perspective of time-based manners (pre, in, and post disruption stages).

Table 3: The Trend of Conceptual Maritime Disruption Management from Literatures

Year	Disruption Management Issues	Events in Maritime Disruptions	Researchers	Objectives	Methods	Disruption management strategies			
						ISM	CR	BCM	RP
2001	Security on supply chain risk assessment	Terrorist attack at port	Sheffi	Propose postponement delays and imported inputs	Risk Pooling Analysis	√	-	√	-
2003	Business Continuity planning	Multimodal application of maritime services	Gupta	Developed business continuity scenario of a firm facing disruptions	Continuity cycle	-	-	√	
2004	Formal risk assessment	Inbound supply after maritime services	Zsidisin, Carter, and Cavinato	Propose formal risk assessment procedures and techniques	Agency theory	√	√	-	-
2005	Disruption visibility	Port stoppages	Blackhurst *et al*..	Propose actions related to disruption discovery, recovery and redesign	Industrial community responses	√	√	-	√
2005	Analytical framework of disruption response	Port closure	Kleindorfer and Saad	Measure the implications of Risk Management program data for disruption management system	Standard logistic regression model	√	-	√	-
2006	Resilient supply chain	Stoppage of maritime operation because of severe weather	Tang	Robust strategy of mitigation process disruption	Lesson learned from cases of Nokia, Dell, Li and Fung.	√	√	-	-
	The quantification of mitigation and contingency strategies	Blockages of one particular shipping routes	Tomlin	Proposed flexible scenarios of optimal disruption-management strategies	Optimal ordering Policy under Markov chain	√	√	-	-
2007	Disruption severity	Port Strike	Craighead *et.al*	Develop warning and recovery planning	Complexity analysis	√	√	-	√
2008	Risk Monitoring system	Supply chain port entry	Handfield *et al*..	Mapping the high critical nodes	Incident report system	-	√	-	-
	Supply Chain risk assessment event	Port Strike event	McCormac	Develop procedures to estimate the probability supplier attributes	Risk Probability Index and multi-use matrix	√	√	-	-

Note: ISM: Inventory and sourcing mitigation; CR: contingency rerouting; BCM: business continuity planning; RP: recovery planning

5. Conclusions

In the global supply chain structure, maritime leg forms a major critical infrastructure on supply-chain operations as it may be impacted by direct and indirect disruptive events. These range from minor weather disruptions to hurricanes, ship shortage to work stoppages and from security breaches to potential terrorist actions. The impacts resulted has radically involved companies, supply-chain operators, maritime users and nations. Active disruption management strategies currently rely on mitigations, contingency actions, assurance strategy of continuity business process and recovery scenarios but few considerations on real-time management.

References

APTINDO (2003), Wheat flour users in Indonesia, accessed 23 September 2007 at http://www.bogasariflour.com/.

APTINDO (2007), Purchasing process of Indonesia flour import, accessed 23 September 2007 at http://www.bogasariflour.com/.

APTINDO (2008), Perkembangan harga terigu dan gandum impor 2007-2008 (US$/ton) accessed 8 March 2008 at www.aptindo.org.

ATSB (2007), Independent investigation into the grounding of the Panamanian registered bulk carrier Pasha Bulker on Nobbys Beach, Newcastle, New South Wales 8 June 2007, ATSB,

Barnes, P. and Oloruntoba, R. (2005), Assurance of security in maritime supply chains: Conceptual issues of vulnerability and crisis management, Journal of International Management, vol. 11, no. 4: pp. 519-540.

Bearing-Point and Hewlett-Packard (2005), Impacts of Major Disruptions On The Maritime Industry, http://www.bearingpoint.com/portal/site/bearingpoint/menuitem.5a42edeee4908885f7a4c810224041a0/?vgnextoid=1ea093beec68f010VgnVCM1000005816600aRCRD&vgnextchannel=7a2d4a9d0b0ce010VgnVCM1000003264a8c0RCRD, accessed 20 June 2008.

Blackhurst, J., Craighead, C. W., Elkins, D. and Handfield, R. B. (2005), An empirically derived agenda of critical research issues for managing supply-chain disruptions, International Journal of Production Research, vol. 43, no. 19: pp. 4067-4081.

Blackhurst, J., Wu, T. and O'Grady, P. (2004), Network-based approach to modelling uncertainty in a supply chain, International Journal of Production Research, vol. 42, no. 8: pp. 1639-1658.

Blas, J. (2007), 'Dry Bulk Freight costs at Record High', Financial Times, accessed 9 November 2007.

BMG (2007), Laporan maritim: Daerah gelombang tinggi, accessed 12 March 2007 - 20 July 2007 at http://www.bmg.go.id/daerah-gelombang-tinggi.bmg?Jenis=URL&IDS=5765606163342513107.

BMG (2008), Laporan maritim: Daerah gelombang tinggi, accessed 2 March 2008 - 2 September 2008 at http://maritim.bmg.go.id/bulletin1.php.

BULOG (2007), Konsep Angkutan Laut dan BULOG Masa Depan, Warta Intern BULOG, 12 July 2008, http://www.bulog.co.id/majalah.php#, accessed 12 July 2008.

Carafano, J. J. (2007), Small Boats, Big Worries: Thwarting Terrorist Attacks from the Sea, accessed 14th of June 2008 at http://209.85.141.104/search?q=cache:DI9smA_7Q_8J:www.heritage.org/Research/HomelandDefense/bg2041.cfm+maritime+disruptions&hl=en&ct=clnk&cd=1&gl=au.

Cavinato, J. (2004), Supply chain logistics risks; From the back room to the board room, International Journal of Physical Distribution and Logistics Management, vol. 34, no. 5: pp. 383.

Chang, S. E. (2000), Transportation performance, disaster vulnerability, and long-term effects of earthquakes, Proceedings of Second Euro Conference on Global Change and Catasrophe Risk Managament, Luxenburg, Austria, pp. 12-24.

Clarkson (2007), Drybulk Trade Outlook, Clarkson Research Service Limited, at http://www.crsl.com/, accessed 3 March 2008.

Coy, P., Foust, D., Woelert, L., Palmeri, C. and Reed, S. (2005), Katrina's Wake, BusinessWeek 7 September 2005, http://www.businessweek.com/bwdaily/dnflash/sep2005/nf2005091_6199_db094.htm, accessed 29 July 2008.

Craighead, C., Jennifer, B., Rungtusanatham, M. J. and Robert, B. H. (2007), The Severity of Supply Chain Disruptions: Design Characteristics and Mitigation Capabilities, Decision Sciences, vol. 38, no. 1: pp. 131.

CSIRO (2007), State of Logistics in Australia, CSIRO, Australia.

Cullinane, K. and song, D.-W. (1998), Container terminals in South Korea: problems and panaceas, Maritime Policy and Management, vol. 25, no. 1: pp. 63-80.

DFAT (2007), Indonesia-Australia FTA Feasibility Study: Background Paper for Industry Consultations, Department of Foreign Affairs and Trade, http://209.85.141.104/search?q=cache:NgXzXHHhF44J:www.dfat.gov.au/geo/indonesia/ia-fta_background_paper.html+Australia+wheat+export+to+Indonesia&hl=en&ct=clnk&cd=33&gl=au, accessed 5 July 2008,

Ditjenhubla (2006), Cetak Biru (Blue Print) Pembangunan Transportasi Laut, Direktorat Jenderal Perhubungan Laut, Jakarta, Indonesia,

Drewry (2007), Dry Bulk Market, Proceedings of DCI Drewry Shipping Consultants Limited Drewry Shipping Consultants Limited Hamburg.

Edijatmiko (2007), Giliran harga terigu yang naik, KOMPAS, http://209.85.173.104/search?q=cache:t11gMhNqMXIJ:www2.kompas.com/ver1/Ekonomi/0801/14/173836.htm+Harga+gandum,+desember+2007,+welirang,+aptindo&hl=en&ct=clnk&cd=2&gl=au, accessed 15 January 2008.

Elkins, D., Kulkarni, D. and Tew, J. (2008), Identifying and Assessing Supply Chain Risk in Supply Chain Risk Management: Minimizing Disruptions in Global Sourcing, ed. B.Handfield, R. & Mccormack, K., Auerbach Publications, New York, pp. 51-56.

FAO (2008), Food outlook: Global market analysis, Trade and Market Division of Food and Agriculture Organization, at http://www.fao.org/docrep/010/ai466e/ai466e03.htm, accessed 20 November 2008.

Fearnsearch (2007), Review 2006 and Dry Bulk Market, http://www.fearnleys.no/index.gan?id=306, accessed 19 August 2008.

Frittelli, J. (2005), Hurricane Katrina: Shipping Disruptions, accessed 27 July 2008 at http://digital.library.unt.edu/govdocs/crs//data/2005/upl-meta-crs-7619/RS22257_2005Sep13.pdf.

G-PORTS (2007), Port Congestion Index, Global Port Congestion Index, at http:/www.g-ports.com/downloads/congestionindexexample-250707.pdf, accessed 25 July 2007.

Gaonkar, R. S. and Viswanadham, N. (2007), Analytical Framework for the Management of Risk, in Supply Chains, IEEE Transactions on Automation Science & Engineering, vol. 4, no. 2: pp. 265-273.

Gupta, D. (2003), An Analysis of Disruptions in the U.S Apparel Manufacturing Industry and Identification of Continuity Planning Strategies, M.Sc, North Carolina State University, Raleigh, North Carolina, USA.

Gurning, S. (2008a), Dampak climate change di angkutan laut nasional, Media indonesia, 29 Juli 2008, http://www.mediaindonesia.com/index.php?ar_id=MTk2MTA=, accessed 29 Juli 2008.

Gurning, S. (2008b), Indonesian ports in 2008, what remain to be tackled? in Indonesia Economics Almanac 2008, ed. Ardi, Y. and Amri, R., Pustaka Bisnis Indonesia, pp. 105 -115.

Gurning, S. and Grewal, D. (2007), An analysis of implementing container transport for wheat cargoes between Australia and Indonesia, Proceedings of 5th International Conference on Supply Chain Management and Information Systems (SCMIS), Melbourne, Australia, pp. 75-85.

Handfield, B. and Mccormack, K. (eds) (2008), Supply Chain Risk Management: Minimizing Disruptions, in Global Sourcing.Auerbach Publications, New York.

Homan, A. C. (2007), The impact of MTSA on financial risk and volatility of marine firms, Maritime Policy and Management, vol. 34, no. 1: pp. 69 - 79.

Howick, S. and Eden, C. (2001), The Impact of Disruption and Delay When Compressing Large Projects: Going for Incentives?, The Journal of the Operational Research Society, vol. 52, no. 1: pp. 26-34.

Jones, R. M. and Towill, D. R. (1998), Shrinking the Supply Chain Uncertainty Circle, Control, vol. September no. pp. 17-22.

Jula, H., Chassiakos, A. and Ioannou, P. (2006), Port dynamic empty container reuse, Transportation Research Part E: Logistics and Transportation, vol. 42, no. 1: pp. 43-60.

Kleindorfer, P. R. and Saad, G. H. (2005), Managing Disruption Risks, in Supply Chains, Production and Operations Management, vol. 14, no. 1: pp. 53-68.

Lin, L. C. and Tseng, C. C. (2007), Operational performance evaluation of major container ports in the Asia-Pacific region, Maritime Policy and Management, vol. 34, no. 6: pp. 535-551.

Mahbub, H. (2007), Miliaran rupiah melayang di Merak-Bakauheni, Tempo interaktif, 3 September 2007, http://209.85.141.104/search?q=cache:hVB_IitUa9kJ:www.tempointeraktif.com/hg/ekbis/2007/09/03/brk,20070903-106750,id.html+kerugian+kemacetan+penyeberangan&hl=en&ct=clnk&cd=2, accessed 25 July 2008.

McCormack, K. (2008), Measuring and Managing Risk, in Supply Chain Risk Management: Minimizing Disruptions in Global Sourcing, ed. B.Handfield, R. and Mccormack, K., Auerbach Publications, New York, pp. 65-91.

Nathan, L. (2005), Union Networks and Global Unionism, in Maritime Shipping, Relations Industrielles, vol. 60, no. 1: pp. 88.

Patunru, Nanda Nurridzki and Rivayani (2007), Port Competitiveness in Indonesia: Cases of Surabaya and Semarang, Asian Development Bank, at www.adbi.org, accessed 12 January 2009.

Peck, H. (2006), Reconciling supply chain vulnerability, risk and supply chain management, International Journal of Logistics Research and Applications, vol. 9, no. 2: pp. 127 - 142.

Peter, H. (2007), WA growers fear fall in wheat price, at accessed 12 March 2008.

Peter, T. (2000), Contesting globalization on the waterfront, Politics and Society, vol. 28, no. 3: pp. 367.

Pettitt, T. (2007), Will a national port regulator solve congestion problems in Australian ports?, Maritime Policy and Management, vol. 34, no. 2: pp. 121-130.

Ray, B. (2007), Container growth up for wheat, The Australian, 15th September 2007, vol. no. pp.

Rodrigues, V. S., Stantchev, D., Potter, A., Naim, M. and Whiteing, A. (2008), Establishing a transport operation focused uncertainty model for the supply chain, International Journal of Physical Distribution and Logistics Management, vol. 38, no. 5: pp. 388-411.

Schultz, G. (2006), Hurricane disruptions, fuel costs add new distribution challenges, accessed 30 July 2008 at http://www.mhia.org/news/industry/7086/hurricane-disruptions--fuel-costs-add-new-distribution-challenges.

Sheffi, Y. (2001), Supply chain management under threat of international terrorism, International Journal of Logistics Management, vol. 12, no. 2: pp. 1-11.

Siagian, N. (2007), Harga Gandum Capai Rekor Tertinggi Sepanjang Sejarah (Wheat price reached the highest record along the history of global wheat trade), Sinar Harapan, http://209.85.173.104/search?q=cache:dlu1ENj8F_cJ:www.sinarharapan.co.id/berita/0709/08/eko01.html+Bisnis+Indonesia,+Welirang,+harga+gandum+melonjak,+2007&hl=en&ct=clnk&cd=7&gl=au, accessed 20 October 2007.

Tang, C. S. (2006), Robust strategies for mitigating supply chain disruptions, International Journal of Logistics Research and Applications, vol. 9, no. 1: pp. 33 - 45.

Tomlin, B. (2006), On the Value of Mitigation and Contingency Strategies for Managing Supply Chain Disruption Risks, Management Science, vol. 52, no. 5: pp. 639-657.

Van de Voort, M., O'Brien, K. A., Rahman, A. and Valeri, L. (2003), Seacurity: Improving the Security of the Global Sea-Container Shipping System, RAND Europe, vol. MR-1695-JRC, no. pp.

WEA (2007), Changes in Export Wheat Marketing, Australian Wheat Export Authority, www.wea.gov.au, accessed 16 Nopember 2007.

WEA (2008), 2006-2007 Statistics. Combined Non-Niche and Niche Application, Australian Wheat Export Authority at http://www.wea.gov.au/WheatExports/Forms.htm, accessed 6 October 2008.

Wright, R. (2008), Threat of disruption hangs over US ports, Financial Times, 3 July 2008, http://www.ft.com/cms/s/0/928fcf40-48e7-11dd-9a5f-000077b07658.html, accessed 28 July 2008.

Yank, L. (2007), Cyclone Gonu aims at strait of Hormuz - fill your tank, Daily Kos, 6 June 2007, http://www.dailykos.com/story/2007/6/6/22122/42975, accessed 29 July 2008.

Yu, G. and Qi, X. (2004), Disruption Management, edn, World Scientific Publishing Co. Pte. Ltd,

Zsidisin, G. A., Ellram, L. M., Carter, J. R. and Cavinato, J. L. (2004), An analysis of supply risk assessment techniques, International Journal of Physical Distribution and Logistics Management, vol. 34, no. 5: pp. 397-413.

Zsidisin, G. A. and Smith, M. E. (2005), Managing Supply Risk with Early Supplier Involvement: A Case Study and Research Propositions, Journal of Supply Chain Management: A Global Review of Purchasing and Supply, vol. 41, no. 4: pp. 44-57.

Comparative analysis of efficiency for major Northeast Asia airports

Hun-Koo Ha,[1] Yuichiro Yoshida[2] and Anming Zhang[3]*

[1] Asia-Pacific School of Logistics, Inha University, Incheon, South Korea.
E-mail: hkha@inha.ac.kr
[2] National Graduate Institute for Policy Studies, Tokyo 106-8677, Japan.
E-mail: yoshida@grips.ac.jp
[3] Sauder School of Business, University of British Columbia, Vancouver, BC, V6T 1Z2, Canada.
E-mail: anming.zhang@sauder.ubc.ca
* Corresponding author

1. Introduction

Northeast Asia has been the world's fastest growth region for the past several decades. Its three principal countries, namely, Japan, Korea and China, all have achieved episodes of rapid economic growth: Japan from the 1950s to 1970s, Korea from the mid-1960s to mid-1990s, and China from 1978 (Table 1).[1] Together they now represent close to 30% of the world income. This rapid economic growth has significantly increased the demand for air transport services, which in turn places enormous pressure on airport infrastructure. Table 2 shows major airports that have been opened in Northeast Asia (NEA) since 1991. In addition, major airport expansions have been, or are being, undertaken at major (both new and "old") NEA airports including Beijing Capital, Narita, and Haneda. Nevertheless, traffic growth outpaces airport capacity increase. As a result, there is an urgent need to improve productivity of airports, thereby relieving demand pressure. Need for productivity improvement also arises from the recent developments of airport corporatization/commercialization and privatization in the region. Productivity certainly matters to a private airport as it will affect the airport's profit. With airport corporatization/commercialization, however, even public airports have been under growing pressure from governments to be more financially self-sufficient and less reliant on government support. In other words, productivity and hence profits can also matter to a public airport that is subject to a certain degree of cost recovery (e.g., Poole, 2003; Zhang and Zhang, 2003; Zhang and Yuen, 2008; see also the discussion in Section 2 below).

Table 1: Average Annual GDP Growth in Northeast Asia, %

	1950s-60s	*1960s-70s*	*1980s*	*1990s*	*2000-07*
Japan	9.6[a]	7.7	3.7	1.7	1.8
Korea		8.0	7.7	6.3	5.2
Hong Kong		9.6	7.4	3.8	5.6
Taiwan		10.2	8.1	6.3	3.8
China		5.1	9.8	9.7	9.5

Notes: a. 1953-1973.
Source: World Development Indicators database; The Directorate General of Budget, Accounting and Statistics, Republic of China.

[1] In this paper, the word "China" refers to Mainland China. Thus unless specifically indicated, Hong Kong and Macau – the two Special Administrative Regions of China – and Taiwan are excluded. Similarly, the word "Korea" refers to South Korea.

Table 2: Major International Airports Opened in Northeast Asia Since 1991

Shenzhen (China)	Kansai (Japan)	Zhuhai (China)	Macau	Hong Kong	Shanghai Pudong (China)	Incheon (Korea)	Guangzhou (China)	Nagoya (Japan)
1991	September 1994	June 1995	November 1995	July 1998	September 1999	March 2001	August 2004	February 2005

Another major development is the airline network rationalization expected to occur in the Northeast Asia in the near future. While there have been substantial progresses in regulatory liberalization in both domestic and NEA markets over the last two decades, the NEA liberalization is largely bilateral in nature; multilateral agreements governing air services among the three economies have yet existed. Due partly to this lack of NEA integration, the structure of airline networks remains ad hoc and fragmented. For example, at the moment there are no real hub-and-spoke networks by airlines in NEA. Nevertheless, the continued air transport liberalization, both within and outside of the NEA region, will facilitate airlines' network reorganization, intensifying competition among major airports in their effort to become the region's preeminent hubs and inter-continental gateways. Part of such competition hinges on how efficient the airports are.

Taken together, these two observations (growing airport capacity constraint and attention to profit, and regional hub formation) suggest a need to study efficiency performance of airports in Northeast Asia. This paper empirically evaluates efficiency (productivity) of major NEA airports in terms of both the trends of efficiency change and differences of efficiency among the airports. We construct a data set consisting of major airports, namely, Tokyo (Narita), Osaka (Kansai), Seoul (Incheon and Gimpo), Beijing Capital, Shanghai Hongqiao, Guangzhou and Hong Kong. These are candidate airports to emerge as the mega-hubs in Northeast Asia. We apply data envelopment analysis (DEA) to this new panel data covering the 1994-2007 period, and examine efficiency scores for each airport. Our analysis will not only be useful for governments' policy formulation, but also be useful for airports – both the ones examined in this study, and those not examined here – and for airlines when they form optimal hub-and-spoke networks.

Our analysis suggests that the rising interest in NEA aviation liberalization and integration is a result of increasing trade and investment ties in NEA over the last two decades. It is also influenced by recent continentalization occurring in the world's airline industry. The analysis further indicates that a liberal and integrated airline market will be formed in Northeast Asia in the medium term, and that four or five mega airport hubs will likely emerge in an integrated NEA. However, the specific shape of the network under an NEA open-skies block will depend on strategic trade policies as well as airports'/airlines' position joggling over the next 5-10 years. An integrated part of this policy formation and position joggling is a comparative study of efficiency performance of airports. Our empirical results of major NEA airports reveal that productivity performance across airports and over time. No sample airports have experienced a persistent growth in overall productive efficiency, and the average efficiency level has risen only slightly over the sample period. To improve productivity, the NEA airports need not only technological revolutions at the industry level, but also the technical improvement, such as of airport management, operations and investment, at an airport level.

Assessment of airport productivity has become the focus of a large number of studies. Different methodologies have been used to measure the productivity of airports in different regions of the world (see, e.g., Oum, et al., 2009, for a review on methodology.). In particular, the DEA approach is widely used; Table 3 summarizes the literature on airport efficiency using DEA. For example, Gillen and Lall (1997) applied DEA to assess terminal and airside operations of 21 top US airports from 1989 to 1993. This is followed by a number of papers applying DEA to performance evaluation of airports in different countries (e.g., Parker, 1999; Sarkis, 2000; Vasigh and Hamzaee, 2000; Chin and Siong, 2001; Martin and Roman, 2001, 2006; Pels, et al., 2001, 2003; Abott and Wu, 2002; Fernandes and Pacheco, 2002; Bazargan and Vasigh, 2003; Sarkis and Talluri, 2004; Pestana Barros and Dieke, 2007).

Table 3: Literature on Airport Efficiency Using DEA Models

Paper	Sample	Input	Output
Gillen & Lall (1997)	21 US airports (1989-1993)	No. of runways, gates, employees, collection belts and parking spots; Length of runway; Airport and terminal areas	No. of passengers, carrier and passenger movements; Pounds of cargos
Parker (1999)	BAA airports before and after privatization	Amounts of labor, capital stock, non-labor and capital cost	No. of passengers and amount of cargo
Sarkis (2000) Sarkis & Talluri (2004)	44 US airports (1990-1994)	Operational cost; No. of employees, gates and runways	No. of passengers and aircraft movements; Amounts of operational revenue and cargo
Martin & Roman (2001)	27 Spanish airports (1997)	Expenditure on labor, capital and materials	No. of aircraft movements and passengers; Amount of cargo
Abbott and Wu (2002)	12 Australian airports (1990-2000)	No. of employees; Amount of capital stock; Length of runway	No. of passengers; Amount of cargo
Fernandes & Pacheco (2002)	35 Brazilian airports (1998)	Areas of apron, departure lounges and baggage claim; No. of check-in counters and vehicle parking spaces; Length of curb frontage	No. of passengers
Bazargan & Vasigh (2003)	45 US airports (1996-2000)	No. of runways and gates; Amount of operating and non-operating expenses	No. of passengers and aircraft movements; Amounts of aeronautic and non-aeronautic revenues; Percentage of on-time operations
Pels, et al (2003)	33 European airports (1995-1997)	No of runways, parking positions, check-in desks and baggage claims	No. of passengers and aircraft movements
Yoshida & Fujimoto (2004)	67 Japanese airports (2000)	No. of employees; Length of runway; Terminal area; Access cost	No. of passengers and aircraft movements; Amount of cargo
Martin & Roman (2006)	34 Spanish airports (1997)	Expenditures on labor, capital and materials	No. of passengers and aircraft movements; Amount of cargo
Pestana Barros & Dieke (2007)	31 Italian airports (2001-2003)	Expenditures on labor, capital and operations (excl. labor)	No. of planes, passengers, cargo; Aeronautical, handling and commercial revenues
Fung, et al. (2008) Zhang & Yuen (2008)	25 Chinese airports (1995-2004)	Runway length; Terminal size	No. of passengers and aircraft movements; Amount of cargo
Yu, et al. (2008)	4 Taiwanese airports (1993-1999)	Expenditures on labor, capital and operations (excl. labor)	Aeronautical and non-aeronautical revenues
Lam, et al. (2009)	11 Asia-Pacific airports (2001-2005)	Labor, terminal area, soft input (outsourcing cost) and country's international trade value	No. of passengers and aircraft movements; Amount of cargo

Source: Reproduced from Lam, et al. (2009) plus the authors' own updates.

As can be seen from Table 3, majority of the productivity studies deal with airports in North America and Europe. This is not surprising: in terms of number of passengers carried, these two regions together accounted for more than 60% of the world traffic. In addition, many major liberalization events concerning airlines and/or airports had involved countries in these two continents. While Northeast Asia is playing an increasingly important role in world aviation, research on airport efficiency comparison in NEA is lacking, owing partly to lack of data especially for Chinese airports. DEA studies on NEA airports were recently emerging, however: Yoshida and Fujimoto (2004) examined 67 Japanese airports using data from year 2000; Fung, et al. (2008) measured productivity for 25 major Chinese airports over the 1995-2004 period; and Lam, et al. (2009) assessed productive efficiency for 11 Asia-Pacific airports over the 2001-2005 period. For example, Fung, et al. (2008) investigated whether airport efficiency in China was improving over time and whether productivity among the airports from different regions was converging, and found positive answers to both questions. The present study aims to add to the stock of studies on NEA airports. Our study examines airports in Korea, Japan and China. This multi-country approach is also in contrast to most of the existing studies that looked at airports with a single country (Pels, et al., 2003, covered EU airports for the 1995-1997 period during which the EU had started the third Package of its common market, and so it was effectively a liberalized EU "domestic" market). In addition, we work on a penal data (in contrast to most cross-sectional data approach in the literature) and cover a longer period of time (1994-2007) during which major regulatory changes had occurred in the three countries. As a consequence, trends of airport efficiency during the period can be analyzed.

The paper is organized as follows. Section 2 describes recent policy changes in the NEA aviation market during the sample period to provide background information for the study. Section 3 examines aviation continentalization, regional liberalization and hub formation in Northeast Asia, and discusses their implications for airports. Section 4 discusses the methodology and data employed by the analysis. Section 5 reports the empirical results on airport efficiency levels and changes. Section 6 discusses areas for future research.

2. Aviation Policy Developments in Northeast Asia

Air transport is a significant sector in China, Japan and Korea: In 2007, the three countries ranked 2nd, 4th and 7th, respectively, in the world in terms of total tone-kilometers performed.[2] Considerable attention has therefore been paid to the formation of "optimal" aviation policies within each country. For the last 20 years, inspired partly by the deregulation and liberalization experiences of the United States and other countries, significant effort has been extended to deregulating domestic markets by each NEA country. In Korea, Asiana Airlines, a trunk carrier, was allowed to enter the industry in 1988 to compete against the incumbent monopoly Korean Air. The increased capacity and competition following Asiana's entry, together with the country's rapid economic growth, resulted in a large increase in air traffic volume. Ten new domestic routes were introduced between 1988 and 1993, as compared to only two routes between 1980 and 1987. On the international routes, the market share of Korean carriers was 48% in 1990 when KAL was a monopoly, but it rose to 67% in 1998 with the duopoly (Park, 2000). Major recent developments include the signing of an open-skies agreement with the U.S. in 1998, and the liberalization of fare setting for domestic routes in August 1999.

In Japan, Skymark Airlines and Air Do, the two low-cost carriers, entered in 1998 the Tokyo-Fukuoka route and the Tokyo-Sapporo route, respectively. Although limited in their scope, these were the first independent entries since the 1960s. The passage of the new "Civil Aeronautics Law" in 1999 represents a significant deregulatory step, as it substantially liberalized the operating license system, fare approval system and other regulatory provisions. The liberalization also allowed airlines to set fares freely

---

[2] According various traffic forecasts by, e.g., Boeing, Airbus, or IATA, the future traffic growth for the region is also robust.

beginning in 2000. In April 2007, the Japanese Government unveiled its "Asia Gateway Plan" aiming at removing restrictions on foreign airline access to its regional airports, boosting trade and tourism, and addressing the issue of increasing regional economic disparity. The government has opened up 23 regional airports to strengthen its gateway position for international traffic.

That airline liberalization, both domestically and internationally, has come relatively slowly to Japan may have to do with the capacity constraint at major airports, particularly landing slots at Haneda, Narita and Osaka (Itami). In the past Japan's Council for Transport Policy argued that because of the airport capacity constraints, an American style of deregulation does not suit the circumstances of Japan (Takahashi, 2003). While the overall national capacity has increased paralleling the deregulation, slot shortage at congested airports has not been resolved. Yamauchi (2000) points out that the slot limitation at the Narita airport has been the single most important barrier against expanding air services to and from Japan for a long time.[3] He attributes two factors to this capacity shortage. First, the Japanese government has rendered too much resource to airport construction in local areas upon strong political pressures. As a result, Japan is able to build a very dense nationwide airport network, which has nevertheless resulted in an airport-capacity shortage in the metropolitan areas.[4] Second, these physical barriers are caused, to a large extent, by financial barriers, since the cost for any capacity expansion in Tokyo area is extremely high. On the other hand, the budget for airport construction has been substantially chopped off by other expenditures such as the expenditure for noise pollution abatement.

In addition, there appeared problems with airport profitability. For instance, fourteen of the 22 airports managed by the Ministry of Land, Infrastructure and Transport lost money amounting to 33.3 billion yen in the 1999 fiscal year (Takahashi, 2003). To reduce government subsidies and improve airport management efficiency, Japan announced plans to privatize, via public share offerings, three major airports: Narita (New Tokyo International), Kansai International Airport, and Central Japan International Airport (in the Chubu/Nagoya region, which was opened in February 2005; see Table 2). The original plan would have privatized the three as a package and used the proceeds to pay off much of the debt of money-losing Kansai. But after considerable airline protests, the government backed off and announced that the New Tokyo International Airport Authority would be privatized on its own, after first being corporatized in 2004. Shares would be sold in tranches over the subsequent five years (Poole, 2003).

The Chinese market shifted from a monopoly to a more competitive market structure in the late 1980s, and China's international aviation policy appears to have shifted away somewhat from the previous restrictive approach – motivated primarily for protecting domestic state-owned airlines – to a proactive regime that views aviation as a facilitator of national trade, foreign direct investment, tourism and economic development (Zhang and Chen, 2003). As argued further by Zhang and Chen (2003), the liberalization efforts have contributed not only to a more competitive market place, but also to airline productivity growth as well as to the industry's dramatic traffic growth. As for airports, to improve their productivity, China embarked an airport localization program, in which airports are turned over to local governments. As a test run, operation of the Xiamen Airport and Shanghai Hongqiao International Airport (including all fixed and working capital and all personnel) was transferred to their municipal governments in 1988 and 1993, respectively. The Civil Aviation Administration of China (CAAC) – the industry's regulator – was, however, still heavily involved in airport business during the late 1980s and 1990s. The localization program regained momentum in the early 2000s and was completed by 2003,

[3] One consequence of the capacity shortage is the high airport charge at Narita. In effect, Narita and Kansai have been among the highest airport charges in the world.
[4] Nevertheless, analysis by Yamaguchi (2005) suggests that there is a significant productivity gain to the national economy from the improvement in air transport accessibility between 1995 and 2000, particularly for the agglomerated areas such as the Tokyo metropolitan region. See also Yoshida and Fujimoto (2004) for a related discussion.

when the CAAC transferred ownership and control of all its remaining airports, except the Beijing and Tibet Airports, to their respective local governments. While the localization program increased the initiatives for local and private investment into airport capacity expansion, airport productivity was expected to improve after the implementation of the localization program. As pointed out by Zhang and Yuen (2008), as opposed to the "soft budget" approach taken by the CAAC, the localization program made the airports more financially accountable and consequently improves their efficiency.[5] Furthermore, as the efficiency of airports has significant implications for local economies, local governments may have greater incentives to improve their airport efficiency than would the CAAC.

The second recent policy change that aimed to improve airport productivity is to allow Chinese airports to be listed on stock markets. Although attracting private funds was one rationale for airport listing, the principal objective was to improve airport efficiency (Zhang and Yuen, 2008). Since the initial public offering (IPO) of Xiamen Gaoqi International Airport, six Chinese airport companies have been listed on stock exchanges in Hong Kong, Shanghai, and Shenzhen. Zhang and Yuen (2008) investigated the effect of public listing on Chinese airport productivity, and found that the listed airports had higher efficiency scores than did unlisted airports, while the correlation between productivity growth and listing was statistically insignificant.

Despite the important progress made *within* each NEA country, the air transport market for the region as a whole is fragmented and restricted. This is due largely to the restrictions set in the bilateral air service agreements (ASAs), an array of laws and regulations and other barriers that prohibit free flow of people, goods and services. As a result, the existing air transport system appears to be ill-equipped for providing efficient air transport services in Northeast Asia.

For the past several years, however, the NEA countries have made significant efforts and progress for achieving overall regional economic cooperation and integration. Since the late 1990s there has been an increasing interest in economic cooperation among the three NEA countries, thanks to a number of important political, economic and social developments. For example, the 1997-1998 Asia financial crisis triggered a sense of the East Asian regional identity, leading to the creation of "ASEAN+3" – i.e., Association of Southeast Asian Nations (ASEAN) plus China, Japan and South Korea – as a formula for regional integration. In November 1999, China, Japan and South Korea held a tripartite summit on the sidelines of the ASEAN summit in Manila, and agreed to conduct joint research to seek ways of institutionalizing economic cooperation. In particular, the three countries would commission their research institutes to identify ten areas, including commerce, shipping, fisheries and customs, as the target sectors for cooperation. At the 2002 ASEAN+3 Cambodia summit, the leaders of the three countries reached an agreement on launching a joint effort to study the feasibility of establishing a Northeast Asian free trade area (FTA). Early in 2004, the first meeting attended by senior officials was held under this agreement, where views and opinions were exchanged on the work program for the joint research. It was believed that the research would pave the way for, and eventually bring about, an official negotiation for the NEA FTA. As discussed in next section, these efforts are certainly conducive to the regional liberalization and integration in air transport.

3. Continentalization, Regional Liberalization and Hub Formation

The rising interest in NEA aviation liberalization and integration is also influenced by recent continentalization in the world's airline industry. More specifically, air transport markets are being continentalized by the creation of the single aviation market of the European Union, the US-Canada open-skies bloc (and an increasing call for creating a single aviation market in North America), and the Trans-Tasmanian "single aviation market" (SAM) between Australia and New Zealand. The EU, US-Canada

[5] As part of the localization program, the central government began to phase out its subsidization of airports in 2006.

open-skies bloc and SAM experiences have demonstrated that significant benefits can be gained from the regional liberalization and integration in air transport markets. These, together with the on-going negotiations for an open aviation area (OAA) across the North Atlantic, have put pressures on Northeast Asia to respond.

In effect, there have been significant recent developments towards more liberal air transport arrangements in Northeast Asia. In June 2006, China and Korea signed an "open skies" agreement on the routes between Korea and China's Shandong Province and Hainan Province. This agreement removes capacity restrictions and pricing control, and allows multiple designation of airlines. As discussed in Lee (2008), the agreement has dramatically reduced airfares on the Shandong routes (from about $500 to $100+) while stimulating new demand on those routes. Furthermore, somewhat surprisingly, Chinese carriers, which were perceived as weaker carriers and hence likely "victims" of liberalization prior to the signing of the agreement, actually gained market shares at the expense of their Korean counterparts on the routes. The two countries also signed a Memorandum of Understanding, intending to expand "open skies" (mainly liberalization of 3^{rd}-/4^{th}-freedom traffic rights) to all regions of China by 2010 (Lee, 2008). In addition, in August 2007 Korea and Japan signed a "limited" open-skies agreement (liberalization of 3^{rd}-/4^{th}-freedom traffic rights) with the exception of the routes involving Tokyo area airports (Narita and Haneda). Moreover, the triangular air shuttle (charter) services among Shanghai's Hongqiao, Seoul's Gimpo and Tokyo's Haneda started in late 2006. These three airports are considered as the "domestic" airports (and hence only handling domestic traffic) in their respective countries.

An important source of gains from NEA liberalization and integration lies in its facilitation of low-cost carrier (LCC) development. LCCs in Northeast Asia have not been very successful so far. Air Do and Skynet Asia in Japan have been operating at a loss while Skymark made a profit only in 2004 (Murakami, 2008), whereas in Korea, both Hansung Airlines and Jeju Air have been operating at a loss (Lee, 2008). This is in contrast to the success enjoyed by major LCCs in North America and Europe, as discussed in Zhang, et al. (2008). One disadvantage of LCCs in NEA is the smaller geographic areas of domestic markets. China does have a large geographic area; nevertheless, as elaborated in Zhang, et al. (2008), its various regulatory barriers make entry and growth of LCCs difficult. In order to succeed and survive, therefore, the NEA LCCs need to expand their operation from domestic routes to intra-NEA regional routes.

Another important source of gains from NEA liberalization and integration lies in its facilitation of airline network rationalization. While there have been substantial progresses in regulatory liberalization in both domestic and NEA markets over the last two decades, the NEA liberalization includes mainly the adoption of more liberal bilateral policies; multilateral agreements governing air services among the three economies have yet existed. Due partly to this lack of NEA integration, the structure of airline networks remains ad hoc and fragmented. For example, at the moment there are no real hub-and-spoke networks by NEA airlines; in contrast, emergence of hub-and-spoke networks is a phenomenon observed in North America and the EU. While bilateral liberalization will continue, effort to negotiate an NEA "open skies" bloc will get its own momentum, especially in light of the fact that China, Japan, and Korea have all participated in the discussion with the ASEAN "common air agreement". It is highly likely that an open-skies bloc will be created in the medium term (e.g., five to ten years) in the region.

Such an open-skies bloc would induce major NEA airlines to set up hub-and-spoke networks in an effort to fully realize economies of traffic density and economies of network evidenced in airline business, as has occurred in the US and EU. When the dust is settled, there would be a very limited number of super-hubs emerging in NEA. Consider the three global airline alliances, namely, Star Alliance, OneWorld, and SkyTeam. Each alliance has airlines in the three main markets – North America, Europe, and Asia – as its senior members. These senior members set up mega-hubs in each continent as their major traffic collection and distribution centers. In the case of North America, these hubs are Chicago (for Star's

United Airlines), Dallas/Fort Worth (for OneWorld's American Airlines) and Atlanta (for SkyTeam's Delta Airlines), as well as to a lesser extent New York and Los Angeles. In the case of Europe, they are Frankfurt (for Star's Lufthansa Airlines), London Heathrow (for OneWorld's British Airways), Paris (for SkyTeam's Air France-KLM) and Amsterdam (for SkyTeam's Air France-KLM).
Recently Air China, which is the largest airline in China, and Shanghai Airlines joined Star Alliance; the other two NEA member carriers are All Nippon Airways (ANA) and Asiana. China Southern joined SkyTeam; the other NEA member carrier is Korean Air, a founding and senior member of SkyTeam. China Eastern and Hainan Airlines (No. 3 and No. 4 largest airlines in China) applied to join OneWorld; the existing NEA member carriers are Cathay Pacific (plus Dragonair) and Japan Air Lines (JAL).

Based on experiences of the US and Europe, four or five mega-airport hubs will likely emerge in an integrated Northeast Asia. Furthermore, it is likely that Hong Kong competes with Guangzhou (perhaps plus Shenzhen) for one such mega-hub, Shanghai competes with Taipei for one – note that Taipei is in the picture only after Mainland China and Taiwan fully implement the "Direct flights" policy), and Tokyo, Seoul, Beijing and Osaka compete among one another for two or three. With the large stake involved for national winners/losers, "strategic trade policy" will take place by governments to ensure their own hub airports to eventually emerge as the hubs for the region. Governments of the three NEA countries would, for instance, engage policies to channel traffic originating in the others' hinterland regions into their own hubs for onward carriage to Europe or North America, in an attempt to ensure their own hub airports to eventually emerge as the regional hubs. For example, the 1998 US-Korea open-skies accord in 1998 and new liberalized US-China ASAs could be viewed as measures to ensure trans-Pacific traffic to be channeled through respective hubs. To illustrate, the aggressive liberalization measures taken by Korea in recent years are particularly important for developing the new Incheon International Airport, opened in 2001, as a major hub airport in East Asia. Incheon's ambitions as a regional aviation and logistics hub are also strengthened by Korean Air's success in the cargo realm – World No. 1 in freight carried, by capacity constraints at Narita and cost constraints at Kansai, and by Korean Air's major participation in the SkyTeam Alliance.

In the meantime, competition among the carriers have been intensifying and changing in nature. Low cost carriers have displaced significant market share from traditional network carriers (about 12% in East Asia), and have forced the network carriers to re-consider their business models. For example, Korean Air established a low cost carrier, Jin Air, in July 2008 to compete with rivals from China and Southeast Asia as well as from the domestic market. In addition, there will be intense competition among major carriers in the region as they attempt to set up multiple-hub network for continental market coverage.

4. A Comparative Study of Airport Efficiency Using DEA

The analyses in Sections 2 and 3 indicate that a liberal and integrated airline market will be formed in Northeast Asia in the medium term, and that four or five mega airport hubs will likely emerge in an integrated NEA. However, the specific shape of the network under an NEA open-skies block will depend on strategic trade policies as well as airports'/airlines' position joggling over the next 5-10 years. Some of the main policy instruments can be those related to airport developments. Here, policies are designed, and investment is made, to ensure that sufficient capacities – both hard and soft capacities – are available for national hub airports, and that these airports are run efficiently. An integrated part of this policy formation is a comparative study of efficiency performance of airports.

4.1. Methodology

As indicated in the introduction, data envelopment analysis (DEA) has been widely used in measuring the performance of airports. As compared to other efficiency measurement techniques, the DEA approach does not require any assumption concerning either the technology or the behavior of actors (e.g., cost minimization) (Pels, et al., 2001), and it can be done without some detailed operating information (such as input costs).

We use DEA to evaluate airport efficiency. In general, DEA is a method that can be used to assess the technical efficiency of a firm. The technical (productive) efficiency of a firm is reflected by the relationship between the outputs the firm produces and the inputs the firm uses in a given period of time. One way to measure a firm's efficiency is to compare it with other firms in the same industry. In the simple case where firms in an industry produce a single output with a single input, Farrell (1957) measured efficiency in terms of potential *input* per unit of output. Here the most efficient firm is used to define the potential input/output. The efficiency measure for any firm in the industry is then defined as the ratio of the potential input to the actual input the firm is using to produce one unit of output. Alternatively, efficiency may be measured in terms of potential *output* per unit of input.

Empirical applications of such efficiency measurements later became feasible by a non-parametric technique known as DEA, developed by Charnes et al. (1978, 1981) based on Farrell's (1957) efficiency measurement. While Farrell's original concept is concerned with singular input and output, the DEA method of Charnes et al. (CCR) can deal with the case where firms use multiple inputs to produce multiple outputs. More specifically, in this more complicated case, similar efficiency measurements can still be obtained, with the most efficient firms forming an efficient frontier. A DEA model gives an efficiency score for each firm in a given industry. For the output-oriented model, the efficiency score has a value between zero and one. Firms with an efficiency score of unity are located on the frontier in the sense that their outputs cannot be further expanded without a corresponding increase in input. Firms with an efficiency score below one are inefficient. The DEA model defines the efficiency score of any firm as the fraction of the firm's outputs that can be produced for a firm on the efficient frontier with the same level of inputs. So it is an extreme point method and compares each producer with only the "best" producers.

We use the CCR method to assess the productive efficiency of major airports in NEA. More specifically, technical efficiency (TE) of decision-making unit (DMU) k is obtained by solving the following linear programming problem:

$$\text{Maximize } E_k = \sum_{r=1}^{s} y_{kr} u_r \tag{1}$$

$$\text{Subject to} \quad \begin{aligned} & \sum_{i=1}^{m} x_{ki} v_i = 1 \\ & \sum_{r=1}^{s} y_{jr} u_r - \sum_{i=1}^{m} x_{ji} v_i \le 0, \qquad j = 1,2,...,n \\ & v_i \ge 0, \qquad i = 1,2,\cdots,m \\ & u_r \ge 0, \qquad r = 1,2,\cdots,s \end{aligned}$$

where y_{kr} = the rth output, x_{ki} = the ith input, u_r = weight of the rth output, v_i = weight of the ith input, s = the number of outputs, m = the number of inputs, and n is the number of DMUs (airports). We use (1) to compute the efficiency score, E_k, where subscript k indicates the DMU (an airport) whose performance relative to the other units is under investigation. With a total of n airports being evaluated, v_i and u_r are the weights of the inputs and outputs respectively, defining the efficiency frontier, and solving the linear programming (1) gives the efficiency score for one airport. To estimate efficiency scores for all the airports, the linear programming must be solved n times (adjusting index k each time).

Although the CCR method is widely regarded as a fundamental model on which the majority of DEA techniques are based, various improvements and refinements have since been made.[6] In particular, the CCR method assumes a constant return to scale technology, which occurs when a two-fold increase in input factors is offset by a similar increase in output factors. To relax this assumption, Banker et al. (1984) decompose the CCR TE (technical efficiency) into "pure technical efficiency" (PTE) and "scale efficiency" (SE), where PTE measures the extent of efficient utilization of inputs while allowing for variable return to scale. Thus, the model of Banker et al. (BCC) allows for potential productivity gains from achieving optimal size of airport. By decomposing TE into PTE and SE, it may also allow one to see sources of productivity differences among airports. Specifically, the PTE of DMU k is obtained by solving the following BCC function:

$$\text{Maximize } E_k = \sum_{r=1}^{s} y_{kr}u_r - w_k \qquad (2)$$

$$\text{Subject to} \quad \begin{aligned} &\sum_{i=1}^{m} x_{ki}v_i = 1 \\ &\sum_{r=1}^{s} y_{jr}u_r - \sum_{i=1}^{m} x_{ji}v_i - w_k \le 0, \qquad j = 1,2,\ldots,n \\ &v_i \ge 0, \qquad i = 1,2,\cdots,m \\ &u_r \ge 0, \qquad r = 1,2,\cdots,s \end{aligned}$$

We shall apply both the CCR and BCC models to our airport data. Since the CCR model assumes constant returns to scale and the BCC model allows variable returns to scale, the efficiencies evaluated from these two models define the scale efficiency.

Both the CCR and BCC models are used to measure the static efficiency of a DMU in a certain year and so it may be difficult to infer from these models about airport efficiency trends, which is one of the present study's objectives. In order to capture the variations of efficiency over time, Charnes, et al. (1985) proposed a technique called "window analysis" in DEA. The DEA-window analysis assesses the performance of a DMU over time by treating it as a different entity in each time-period. DEA-window analysis works on the principle of moving averages, and measures the level of relative efficiency during a certain period (called a "window"). By using windows we can overcome the problem of (potentially) unstable efficiency indices produced by the standard, one-year DEA methods (CCR or BCC).

4.2. Sample Airports and Variable Construction

We construct a data set consisting of major airports – namely, Narita International (Tokyo), Kansai International (Osaka), Incheon International (including Seoul Gimpo International), Beijing Capital International, Shanghai Hongqiao International, Guangzhou Baiyun International, and Hong Kong International – in the period between 1994 and 2007.[7] Note that the Seoul Gimpo data were added up into the Incheon data, and that due to the opening of the new Guangzhou Baiyun Airport in August 2004 (while the closing of the old airport at the same month) we adjust the inputs of the two airports (the terminal size and runway length) on a pro-rata basis for that year.

To measure productive efficiency by DEA, one must first identify outputs that an airport produces and the inputs it uses in producing those outputs. The characteristics of the DEA input/output variables are

[6] Useful general references on DEA include Banker, et al. (1989), Seiford and Thrall (1990), Lovell (1993) and Seiford (1996).

[7] We started with a sample period of 1991-2007, which was later reduced to the current 1994-2007 period due to missing data for Chinese airports.

summarized in Table 4. On the input side, we consider two physical capital input measures – runway length and terminal size – along with the number of employees. Narita and Guangzhou expanded runway length in 2003 and 2005 (see Figure 1). On the output side, we consider passenger volume, air cargo volume, and the number of aircraft movement as outputs of airports. The number of passengers served is the most commonly used output for measuring airport productivity. As can be seen from Figure 2, the outputs of Kansai and Shanghai Hongqiao have been stable over time, while the outputs of the other airports have been increasing steadily.

Figure 1: DEA Input Variables

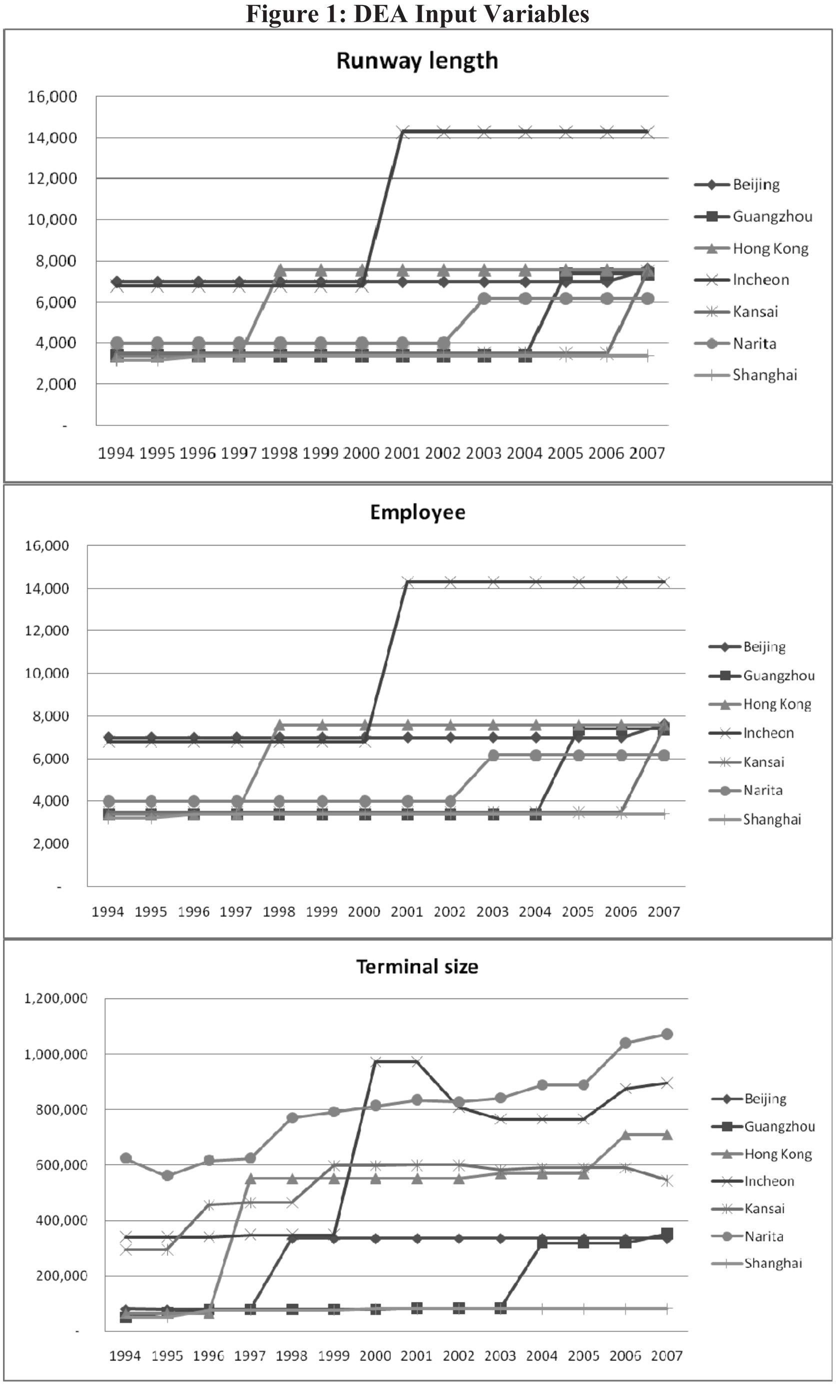


Figure 2: DEA Output Variables

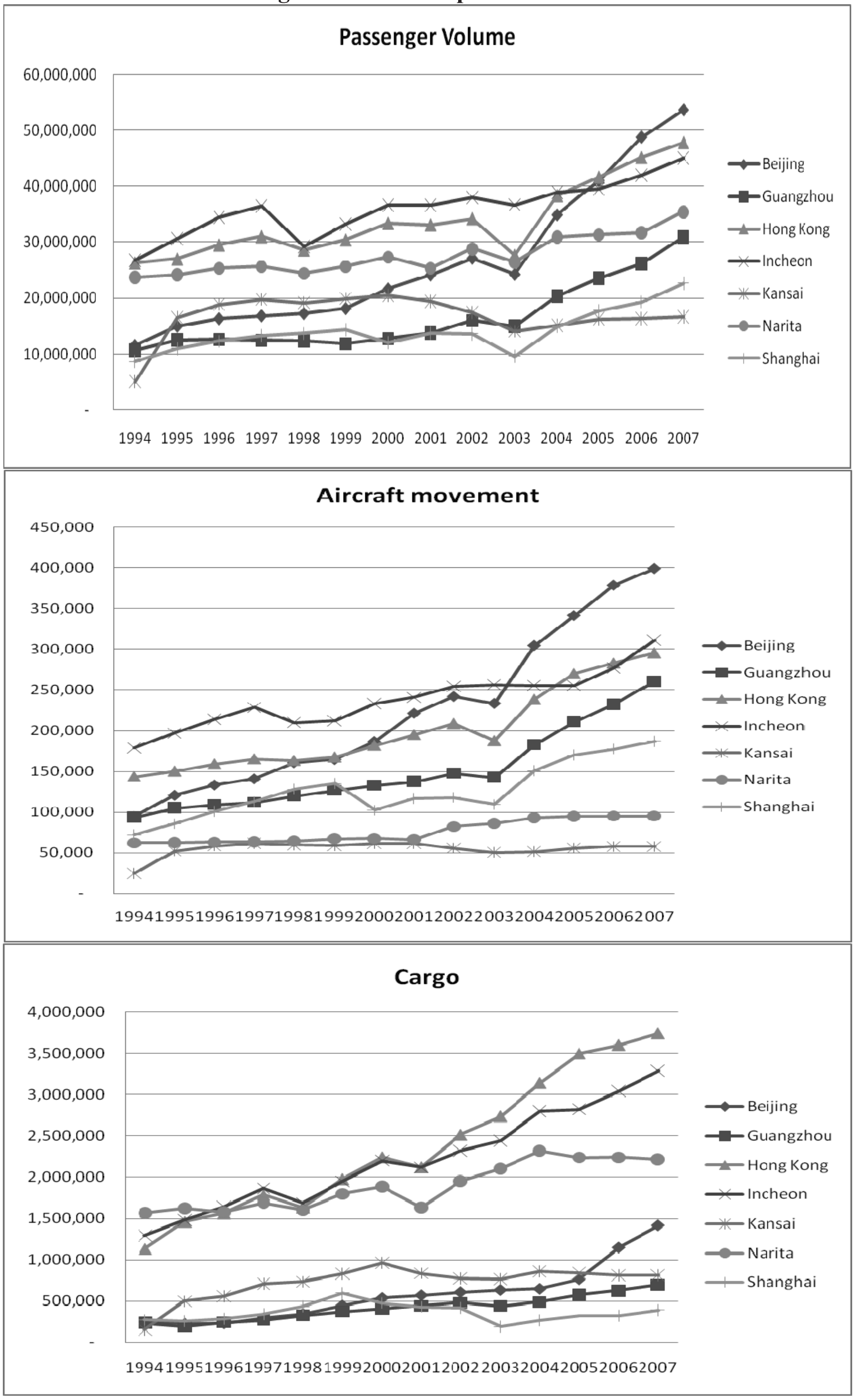


Table 4: Characteristics of DEA Input/Output Variables

Variable (units)	Definition	Mean	Standard deviation
Input			
Runway length (meter)	• The total sum of runway length	5,739	2,954
Terminal size (sq. meter)	• The total area of passenger and cargo terminal	397,413	292,705
Employee (persons)	• The total sum of full-time and part-time employee	2,361	1,953
Output			
Passenger (persons)	• The number of passengers who arrive at and depart from the airport	24,967,344	10,516,235
Cargo (tons)	• The number of cargo that arrive at and depart from the airport	1,412,000	936,698
Aircraft movement (flights)	• The number of planes that landed at and take off from the airport	159,782	82,442

The data are complied from various sources: the Japanese data are from Japan Aeronautic Association (Nippon Koku Kyoukai), 1995-2008; the Ministry of Land, Infrastructure, Transport and Tourism, 2008; and National Airport Building Association (Zenkoku Kuukou Buil Kyoukai), 1982-2007. The Chinese data are obtained from Statistical Data on Civil Aviation of China, 1995-2008 (for runway length, terminal size, passenger, cargo and aircraft movement); and company annual reports of listing airports (for the number of employees). Finally, the Korean data are from Statistical Data on Incheon International Airport Corporation; Statistical Data on Korea Airports Corporation (for runway length, terminal size, passenger, cargo and aircraft movement); and direct contact (for the number of employees).

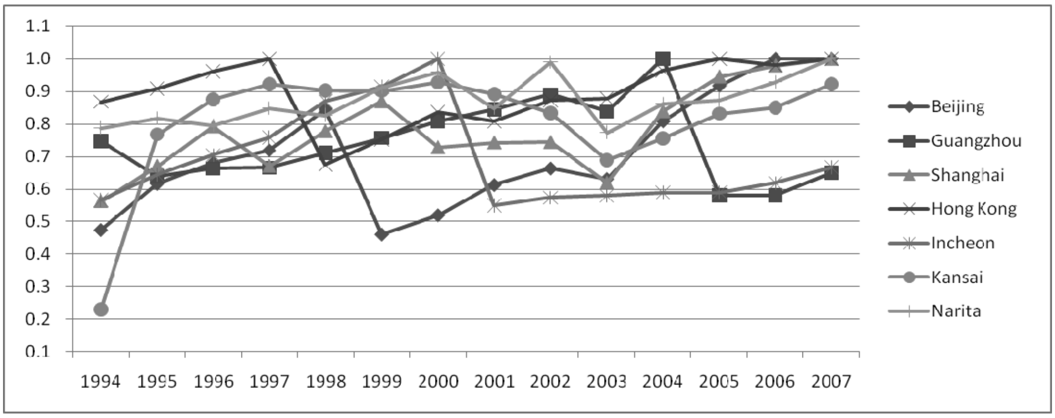
5. Empirical Results

5.1. Efficiency Scores

The results of the DEA technical efficiency (TE) scores of the sample airports in the period between 1994 and 2007 are shown in Table 5. Guangzhou airport had a sharp decline of TE in 2005, owing to a huge increase of inputs in 2004 when the airport moved to a brand new location. Similarly, Incheon had a big drop in TE in 2001 when Incheon International opened. Beijing saw a sharp TE decline in 1999 but has enjoyed a steady increase in TE since that year. Hong Kong has gradually increased TE since 1998, the year when its new airport opened. Over time, the average TE scores have shown an upward trend, with the exception of 2003. The sharp decline in 2003 is caused largely by the SARS breakout in Northeast Asia that had a severe negative impact on air travel demand in the region.

Table 5: DEA Technical Efficiency (TE) of Airports, 1994-2007

	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Beijing	0.47	0.62	0.68	0.72	0.85	0.46	0.52	0.61	0.66	0.63	0.81	0.92	1.00	1.00
Guangzhou	0.75	0.64	0.66	0.67	0.71	0.76	0.81	0.84	0.89	0.84	1.00	0.58	0.58	0.65
Shanghai	0.56	0.67	0.79	0.67	0.78	0.87	0.73	0.74	0.74	0.62	0.84	0.95	0.98	1.00
Hong Kong	0.87	0.91	0.96	1.00	0.67	0.75	0.84	0.81	0.87	0.88	0.96	1.00	0.98	1.00
Incheon	0.57	0.64	0.70	0.76	0.87	0.91	1.00	0.55	0.57	0.58	0.59	0.59	0.62	0.67
Kansai	0.23	0.77	0.87	0.92	0.90	0.90	0.93	0.89	0.83	0.69	0.75	0.83	0.85	0.92
Narita	0.79	0.82	0.80	0.85	0.83	0.91	0.96	0.85	0.99	0.77	0.86	0.87	0.93	1.00



The technical efficiency is then decomposed into pure technical efficiency (PTE) and scale efficiency (SE), with the latter two efficiency scores being given in Tables 6 and 7 respectively. Like the case of TE, Guangzhou had a sharp decline of PTE in 2005 due to the opening of the new Baiyun airport in 2004, whereas Beijing had a sudden decrease of PTE in 1999, but had since seen a stead rise in PTE (especially since 2004). On the other hand, Shanghai and Kansai have kept a very high level of PTE during the entire sample period. On SE, Incheon has a sharp decline in 2001 due to a sudden increase in inputs, whereas the scores for Kansai and Shanghai fell in 2003 because of the declines in passenger traffic. On the other hand, Hong Kong and Narita have more or less kept a high level of SE over the period. Other highlights include a sharp increase of SE for Guangzhou since 2004 owing to a large reduction of employees, and Shanghai's significant increase of SE since 2004 owing to output expansion.

We have also conducted a DEA-Window analysis on the data, with Table 8 providing average efficiency scores for both DEA (discussed above) and DEA-Window. As can be seen from the table, DEA-Window, while stabilizing the efficiency scores, shows a very similar trend of efficiency scores as the DEA analysis. Furthermore, the time-series patterns of TE, PTE and SE under the DEA-Window analysis are drawn in Figure 3. As can be seen, when the sample NEA airports are taken as a whole, both TE and SE have shown an upward trend. On the other hand, PTE has exhibited significant fluctuation over time, with no noticeable upward trend observed, suggesting that more efforts be needed to improve PTE.

Table 6: DEA Pure Technical Efficiency (PTE) of Airports, 1994-2007

	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Beijing	0.69	0.71	0.75	0.77	0.83	0.48	0.50	0.58	0.63	0.61	0.80	0.90	1.00	1.00
Guangzhou	1.00	0.96	0.96	0.96	0.97	0.97	0.97	0.98	0.98	0.98	1.00	0.55	0.60	0.67
Shanghai	1.00	1.00	1.00	0.96	0.96	0.97	0.95	0.96	0.96	0.95	0.98	0.99	0.99	1.00
Hong Kong	1.00	1.00	1.00	1.00	0.71	0.77	0.82	0.80	0.86	0.91	0.97	1.00	0.96	1.00
Incheon	0.56	0.62	0.66	0.70	0.88	0.97	1.00	0.55	0.59	0.60	0.59	0.59	0.64	0.73
Kansai	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00
Narita	0.85	0.85	0.85	0.87	0.85	0.90	0.93	0.86	0.95	0.83	0.89	0.89	0.96	1.00

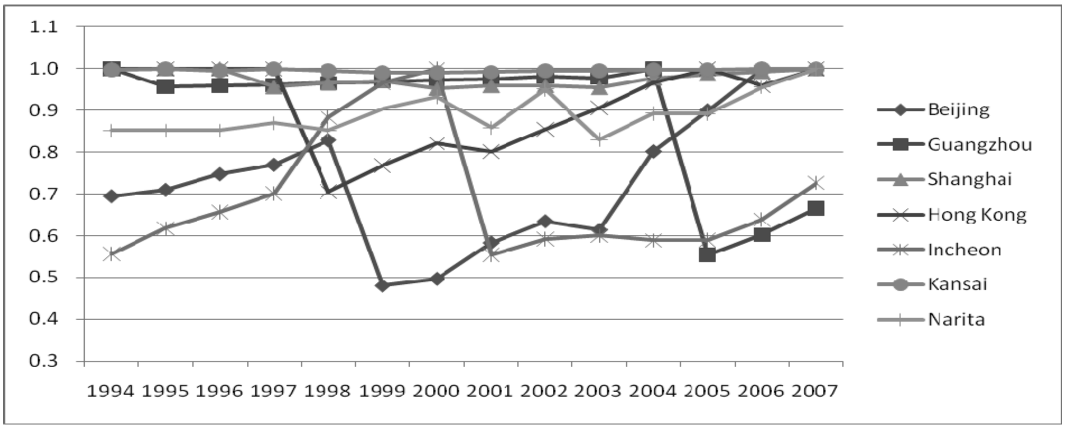


Table 7: DEA Scale Efficiency (SE) of Airports, 1994-2007

	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Beijing	0.68	0.87	0.91	0.93	0.99	0.90	0.99	1.00	1.00	1.00	0.99	0.99	1.00	0.98
Guangzhou	0.75	0.66	0.68	0.64	0.68	0.72	0.75	0.77	0.82	0.80	1.00	0.95	0.96	0.98
Shanghai	0.56	0.67	0.79	0.65	0.74	0.77	0.60	0.67	0.67	0.62	0.84	0.93	0.96	1.00
Hong Kong	0.87	0.91	0.96	1.00	0.95	0.93	0.97	0.97	0.95	0.92	0.98	1.00	1.00	1.00
Incheon	0.93	0.95	0.99	0.99	0.94	0.95	1.00	0.99	0.97	0.96	0.98	0.98	0.97	0.92
Kansai	0.23	0.76	0.86	0.90	0.87	0.87	0.89	0.86	0.80	0.67	0.73	0.81	0.84	0.92
Narita	0.88	0.91	0.89	0.93	0.90	0.96	0.97	0.92	0.99	0.84	0.91	0.93	0.96	1.00

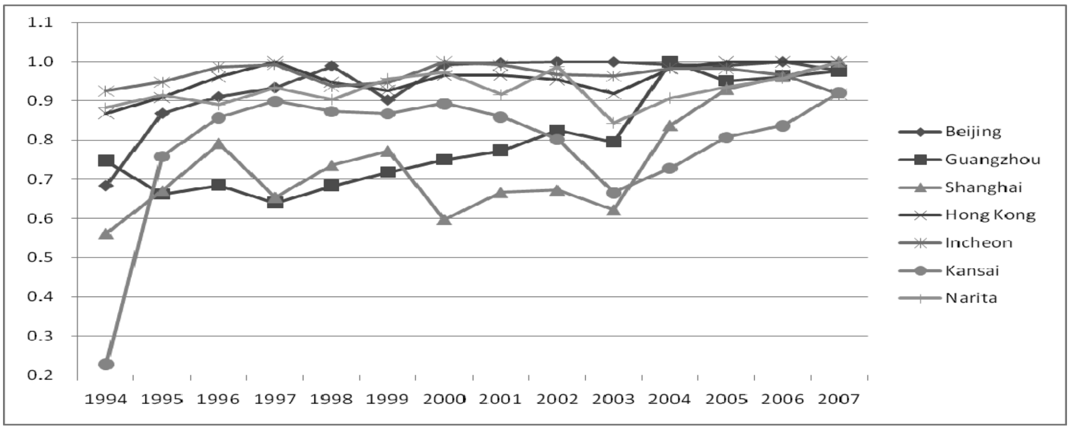
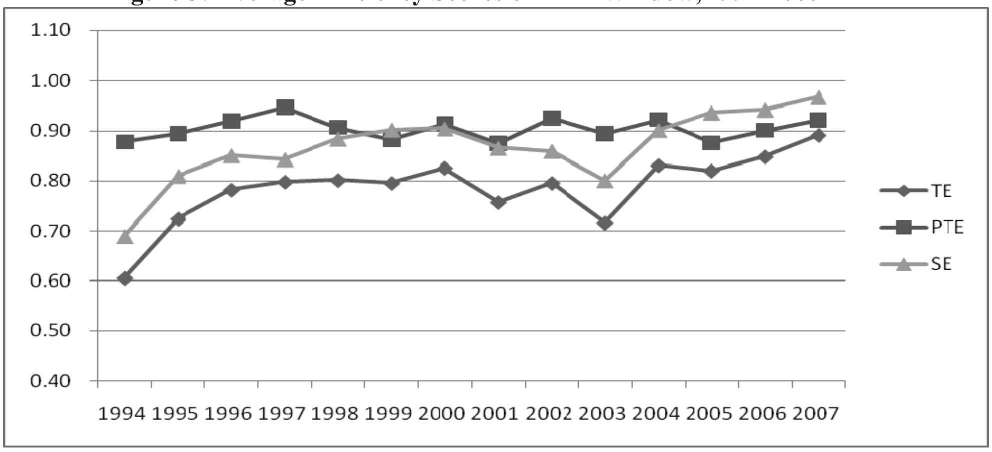


Table 8: Average Efficiency Scores of DEA vs. DEA-Window, 1994-2007

	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
TE: DEA	0.59	0.71	0.76	0.77	0.76	0.75	0.77	0.71	0.74	0.68	0.81	0.79	0.84	0.89
TE: DEA-Window	0.60	0.72	0.78	0.80	0.80	0.80	0.83	0.76	0.80	0.72	0.83	0.82	0.85	0.89
PTE: DEA	0.87	0.88	0.89	0.89	0.89	0.86	0.88	0.82	0.85	0.84	0.89	0.85	0.88	0.91
PTE: DEA-Window	0.88	0.89	0.92	0.95	0.91	0.88	0.91	0.87	0.93	0.89	0.92	0.88	0.90	0.92
SE: DEA	0.70	0.82	0.87	0.86	0.87	0.87	0.88	0.88	0.89	0.83	0.92	0.94	0.96	0.97
SE:	0.71	0.82	0.85	0.85	0.89	0.91	0.91	0.87	0.86	0.81	0.90	0.93	0.94	0.96

Notes: TE = Technical efficiency; PTE = Pure technical efficiency; SE = Scale efficiency

Figure 3: Average Efficiency Scores of DEA-Window, 1994-2007



Notes: TE = Technical efficiency; PTE = Pure technical efficiency; SE = Scale efficiency

5.2. Productivity-growth Analysis Using Malmquist Productivity Index

Having examined levels of productivity for the sample airports, we will now turn to the changes in levels of productivity. This examination is useful in that, if the low level of productivity at some airports is due to their low starting point, then a faster growth rate in productivity could reduce and eliminate the gap. For instance, Fung, et al. (2008) found that productivity among Chinese airports from different regions was converging.

The Malmquist productivity index (MPI) is used to measure changes in the overall productivity of each DMU over time:

$$M_k^{t+1}(x^t, y^t, x^{t+1}, y^{t+1}) = \frac{D_k^{t+1}(x^{t+1}, y^{t+1})}{D_k^t(x^t, y^t)} \times \left[\frac{D_k^t(x^{t+1}, y^{t+1})}{D_k^{t+1}(x^{t+1}, y^{t+1})} \times \frac{D_k^t(x^t, y^t)}{D_k^{t+1}(x^t, y^t)} \right]^{\frac{1}{2}} \tag{3}$$

where D_k is an output distance function of airport *k*. Thus MPI is defined with use of the distance function, referring to the distance between the observed point and point on the production frontier. DEA methods are generally used for the computation of distance functions; and here, the distance function is the inverse of the output-oriented efficiency score calculated above. The superscripts on D_k indicate the time periods within which the efficiency scores are calculated. The superscripts on *x* and *y* indicate the time periods of the data used in the calculation of the efficiency scores. As a measure of the "total factor productivity change" (TFPC), an MPI M_k^{t+1} greater (less) than unity indicates that the overall efficiency of airport *k* has increased (declined) from period *t* to period *t+1*.

The Malmquist productivity changes can be decomposed into two useful components: Note that equation (3) also represents a decomposition of efficiency change from period *t* to period *t*+1. The ratio outside the bracket on the right-hand side of (3) measures the "technical efficiency change" (TEC) of airport *k* from period *t* to period *t*+1. Greater (smaller) than unity implies that the technical efficiency has improved (declined) in reference to the production frontier from period *t* to period *t*+1. The bracketed term represents the geometric mean of the shift in production frontier. When the value of this term is greater (less) than unity, it implies that the technology of the industry has progressed (regressed) from period *t* to period *t*+1; this value measures the extent of airport business innovations and is termed as TC (for "technology change").

Tables 9, 10 and 11 present the MPI results, that is, TFPC, TEC and TC, respectively. Kansai experienced the highest growth rate of TFPC (of all TFPC growth rates in Table 9) in 1994-95. On the other hand, the overall productivity of Hong Kong dropped by 78% in 1997-98, whereas Beijing saw a sharp decline of TFPC in 1998-99. All these TFPC changes are driven by changes in its TEC component. In contrast, Beijing experienced the highest growth of TFPC in 2005-06, but this change is driven by the TC component rather than the TEC component. Furthermore, Table 10 shows that in 1997-98, except for Hong Kong that saw its TEC decreased by 72%, most of the sample airports recorded high growth rates in TEC. In comparison, in 1997-98 all the airports recorded a large fall in TC (Table 11). The average TC increased by 46% in 2003-04, with every airport experienced an increase in TC (Beijing recorded 72%, the highest growth in 2004).

Table 9: MPI Total Factor Productivity Changes (TFPC) by Airports

	95	96	97	98	99	00	01	02	03	04	05	06	07	Average
Beijing	1.29	1.10	1.06	1.15	0.51	1.13	1.15	0.94	0.83	1.41	1.25	2.08	0.98	1.15
Guangzhou	0.98	1.04	0.95	1.05	1.06	1.16	1.01	1.12	1.06	1.69	0.53	1.08	1.08	1.06
Hong Kong	1.15	1.06	1.10	0.22	1.24	1.15	1.20	1.11	0.93	1.28	1.15	0.92	1.06	1.04
Incheon	1.13	1.09	1.08	1.58	1.13	1.12	0.36	1.05	1.07	1.04	1.00	1.02	1.09	1.06
Kansai	5.69	1.41	1.23	0.92	0.97	1.01	0.95	0.91	0.79	1.07	1.10	1.08	1.09	1.40
Narita	1.17	0.83	1.30	1.04	1.03	1.03	0.87	1.16	0.76	1.13	1.12	1.31	1.21	1.07
Shanghai	1.19	1.15	0.92	1.15	1.16	0.81	1.00	0.94	0.84	1.37	1.27	1.01	1.06	1.07
Average	1.80	1.10	1.09	1.02	1.01	1.06	0.93	1.03	0.90	1.28	1.06	1.21	1.08	1.12

Table 10: MPI Technical Efficiency Changes (TEC) by Airports

	95	96	97	98	99	00	01	02	03	04	05	06	07	Average
Beijing	1.23	1.04	1.02	2.63	0.50	1.03	1.54	0.85	0.89	0.82	1.84	1.73	0.95	1.24
Guangzhou	0.93	0.98	0.91	2.00	1.00	1.34	0.92	1.06	1.26	1.13	0.60	0.53	1.11	1.06
Hong Kong	0.99	0.96	1.01	0.28	1.00	1.01	2.00	1.02	0.97	1.07	1.11	0.91	1.00	1.03
Incheon	1.08	1.03	1.03	2.72	1.02	1.01	0.43	0.99	1.19	0.80	0.90	1.04	1.05	1.10
Kansai	5.59	1.01	1.00	1.09	0.92	0.83	1.26	0.94	0.99	0.69	1.03	1.11	0.81	1.33
Narita	0.88	0.63	0.94	2.03	0.99	0.97	1.00	1.03	0.88	0.78	0.95	1.32	1.02	1.03
Shanghai	1.14	1.09	0.88	2.42	1.07	0.86	1.02	0.85	0.89	0.89	1.91	0.97	1.01	1.15
Average	1.69	0.96	0.97	1.88	0.93	1.01	1.17	0.96	1.01	0.88	1.19	1.09	0.99	1.13

Table 11: MPI Technological Changes (TC) by Airports

	95	96	97	98	99	00	01	02	03	04	05	06	07	Average
Beijing	**1.05**	**1.06**	**1.04**	**0.44**	**1.03**	**1.10**	**0.74**	**1.10**	**0.94**	**1.72**	**0.68**	**1.20**	**1.03**	**1.01**
Guangzhou	**1.05**	**1.06**	**1.04**	**0.52**	**1.06**	**0.87**	**1.09**	**1.05**	**0.84**	**1.50**	**0.88**	**2.03**	**0.98**	**1.07**
Hong Kong	**1.16**	**1.09**	**1.09**	**0.76**	**1.24**	**1.14**	**0.60**	**1.08**	**0.96**	**1.20**	**1.04**	**1.01**	**1.05**	**1.03**
Incheon	**1.05**	**1.06**	**1.04**	**0.58**	**1.11**	**1.11**	**0.84**	**1.06**	**0.90**	**1.29**	**1.11**	**0.99**	**1.04**	**1.01**
Kansai	**1.02**	**1.39**	**1.23**	**0.84**	**1.05**	**1.22**	**0.75**	**0.97**	**0.80**	**1.54**	**1.07**	**0.96**	**1.35**	**1.09**
Narita	**1.34**	**1.32**	**1.39**	**0.51**	**1.04**	**1.07**	**0.87**	**1.13**	**0.87**	**1.46**	**1.18**	**0.99**	**1.18**	**1.10**
Shanghai	**1.05**	**1.06**	**1.04**	**0.48**	**1.08**	**0.94**	**0.98**	**1.11**	**0.95**	**1.54**	**0.67**	**1.04**	**1.05**	**1.00**
Average	**1.10**	**1.15**	**1.12**	**0.59**	**1.09**	**1.06**	**0.84**	**1.07**	**0.89**	**1.46**	**0.95**	**1.17**	**1.10**	**1.05**

Other noticeable points include:

- In 2005-06, Beijing experienced the highest growth of overall efficiency (TFPC), with TEC having a more impact on this growth than TC. The increase of TEC was influenced by the significant drop in employee number;
- In 1997-98, Hong Kong's TFPC dropped by 78%, with TEC having a greater impact on this decline than TC. The drop in TEC (by 72%) is due to the expansion of inputs: the new Chek Lap Kok airport opened in July 1998, which replaced the old Hong Kong airport;
- The productivity rate of Guangzhou airport fell in 2004-05 suddenly, with TEC having a more impact on the decrease than TC. In September 2004, Guangzhou Baiyun International Airport opened, replacing the old Guangzhou airport;
- Except for 2001, the productivity of Incheon was trending upwards. In 2001, both TEC and TC recorded a sharp drop: The inputs increased dramatically due to the opening of the Incheon International Airport in March 2001;
- Kansai airport experienced huge growth of TFPC in 1994-95. The sharp rise of TFPC was affected predominantly by TEC, as the outputs increased greatly in 1995. Note that Kansai International Airport opened in September 1994;
- The average (over yaers) growth rates for individual airports are given in the last columns of Tables 9, 10 and 11. The indices of TFPC, TEC and TC are larger than unity, and so all the sample airports recorded an increase in productivity. The rise of productivity is especially high for Beijing. TFPC was affected by TEC and TE at the same time, but the graphs of average productivity changes showed that TC has had a greater effect on TFPC than TEC;
- The average (over airports) growth rates for individual years are given in the last rows of Tables 9, 10 and 11. Except 2000-01 and 2002-03, TFPC had increased gradually. High growth rates of TEC were recorded in 1994-95 and 1997-98, while TC decreased suddenly in 1997-98 and also experienced a large degree of up-and-down fluctuations.

Implications of our MPI analysis may include: First, TFPC of most sample airports grew at a similar rate as TEC except for Narita. In view of the results, while TFPC is affected by both TEC and TE, the technical efficiency of individual airports appears to have a stronger effect on the overall productive efficiency than the level of technological innovations in the industry. Second, no sample airports have experienced a persistent growth in productivity, and the average efficiency level has risen only slightly over the sample period. Alternative plans are needed to improve productivity of airports in NEA. Finally, to improve productivity, the major airports in NEA need not only technological revolutions and effective policies at the industry level, but also the improvement in airport management, operations and investment so that technical efficiency is improved at an individual airport.

6. Future Work

6.1. Non-major Airports

Unlike the above study where a few major airports in Northeast Asia are analyzed, it will also be important to study a large number of non-major airports in the region. There are several reasons for why such a study is important and timely. First, as discussed in Zhang, et al. (2008), the low-cost carrier (LCC) model, while being a much more recent phenomenon and being so far less successful in NEA than in North America and Europe, will grow and prosper in NEA in the near future.[8] Unlike full-service airlines (FSAs), LCCs in North America and Europe de-emphasize hub-and-spoke networks; instead, they tend to provide point-to-point services linking non-major cities, or using secondary airports in a metro area (Tretheway, 2004; O'Connell and Williams, 2005; Zhang, et al., 2009). This strategy will likely be followed by NEA LCCs especially in a more liberal NEA airline market, thereby suggesting the importance of studying efficiency performance of non-major airports.

Second, as mentioned in Section 3, one of the major gains from the NEA liberalization and integration is from better facilitation of LCC development. The main disadvantage of the LCCs in NEA is the smaller geographic areas of domestic economies while confronting an array of regulatory constraints and barriers in intra-NEA markets. Here, a liberal and integrated NEA aviation market is critical for the development and growth of regional LCCs, since it will open up numerous secondary city-pair markets between Korea, Japan and China.[9] These are the markets where FSAs tend not to operate, but LCCs can thrive with their business model (regional jets, one fare-class, "no frill" services, less congestion at secondary airports and hence fast aircraft turnaround, etc.). Furthermore, there is unused capacity available in Japan's secondary airports as a result of over-development in the past, a consequence of the government airport policy and strong local political pressures (see the discussion in Section 2). Capacity now becomes available at Korea's secondary airports as well, owing to stagnant/declining domestic traffic following the operation of high-speed passenger rail; and will likely become available in China as the government invests further in secondary airports as part of its "developing the west" strategic plan. Third, given the potentially numerous city-pair markets and hence massive airline markets to be tapped, major NEA airlines will, in addition to strengthening their main hubs, need to develop their secondary hubs in order to effectively serve the whole NEA market. These observations again point to the importance of studying efficiency performance of non-major airports.

We will in a follow-up study evaluate the performance of airports in NEA, with about 10 airports from Korea, 25-40 airports from China and 20 airports from Japan. Because of a large number of airports involved, it is not likely that we will have data for the number of employees, various airport soft costs, or concession revenues, although other input and output categories will still be available. Due to such data deficiency, it is difficult to assess airport "productivity"; rather, we will just focus on the capital (or capacity) efficiency of these airports. On the other hand, methodology issues remain more or less the same as those in this study. Earlier studies by Yoshida and Fujimoto (2004) on 67 Japanese airports and Zhang and Yuen (2008) on 25 Chinese airports will also be heavily relied upon in such a follow-up study.

[8] The low-cost model, represented by Southwest Airlines, has been evolving in the United States since its deregulation. This model is repeated in Canada and Australia, as well as in Europe since the mid-1990s when the EU liberalized and integrated its internal market. In 1998 the two low-cost carriers, Skymark Airlines and Air Do, entered, respectively, the Tokyo-Fukuoka route and the Tokyo-Sapporo route. In Korea, the LCCs, Hansung Air and Jeju Air, entered the domestic market in 2005 and 2006 respectively, and two more LCCs entered the market in 2008. Similar developments have occurred in China, Hong Kong and Macau (Zhang, et al., 2008).

[9] As shown in Table 12, there are 52 cities in China, 12 cities in Japan, and 8 cities in Korea with an urban area population over 1 million. In comparison, the figures for the US, Canada and EU-27 are 51, 6 and 42 respectively (69 for Europe as a whole).

Table 12: World Cities by Urban Area Population, in million

China		Japan		Korea		United States		Canada		EU-27	
Shanghai	15.550	Tokyo	8.718	Seoul	9.820	New York	18.816	Toronto	5.113	London	13.220
Beijing	13.133	Yokohama	3.644	Busan	3.524	Los Angeles	12.876	Montreal	3.636	Paris	11.886
Guangzhou	11.046	Osaka	2.650	Incheon	2.531	Chicago	9.525	Vancouver	2.117	Madrid	6.321
Shenzhen	8.464	Nagoya	2.244	Daegu	2.465	Dallas	6.145	Ottawa	1.131	Barcelona	4.972
Dongguan	6.450	Sapporo	1.897	Daejeon	1.443	Philadelphia	5.828	Calgary	1.079	Milan	4.340
Tianjin	5.190	Kobe	1.532	Gwangju	1.418	Houston	5.628	Edmonton	1.035	Berlin	4.056
Chongqing	5.087	Kyoto	1.468	Ulsan	1.049	Miami	5.413			Rome	3.940
Wuhan	4.890	Fukuoka	1.435	Suweon	1.044	Washington	5.307			Athens	3.853
Harbin	4.755	Kawasaki	1.385			Atlanta	5.279			Naples	3.841
Shenyang	4.420	Saitama	1.197			Boston	4.483			Hamburg	3.264
Chengdu	3.750	Hiroshima	1.165			Detroit	4.468			Frankfurt	3.144
Zhengzhou	3.500	Sendai	1.030			San Francisco	4.204			Lisbon	2.638
Qingdao	3.200					Phoenix	4.179			Budapest	2.574
Shantou	3.200					Riverside-Ontario	4.081			Stuttgart	2.332
Nanjing	3.110					Seattle	3.309			Munich	2.308
Kunming	3.055					Minneapolis	3.208			Warsaw	2.256
Huizhou	2.900					San Diego	2.975			Brussels	2.202
Zibo	2.900					St. Louis	2.809			Bucharest	2.178
Fuzhou	2.600					Tampa	2.724			Vienna	2.124
Changsha	2.520					Baltimore	2.668			Stockholm	1.965
Nanchang	2.440					Denver	2.465			Lyon	1.815
Wuxi	2.400					Pittsburgh	2.356			Turin	1.698
Suzhou	2.400					Portland	2.175			Marseille	1.635
Hangzhou	2.350					Cincinnati	2.134			Glasgow	1.629
Changchun	2.290					Cleveland	2.096			Bielefeld	1.444
Ningbo	2.182					Sacramento	2.091			Prague	1.395
Guiyang	2.180					Orlando	2.032			Valencia	1.385
Dalian	2.090					San Antonio	1.991			Seville	1.344
Jinan	2.090					Kansas City	1.985			Porto	1.311
Shijiazhuang	2.090					Las Vegas	1.836			Helsinki	1.276
Lanzhou	2.060					San Jose	1.804			Nuremberg	1.227
Yantai	2.000					Columbus	1.754			Sofia	1.205
Xi'an	1.960					Indianapolis	1.695			Bremen	1.188
Jilin City	1.900					Virginia Bench	1.659			Antwerp	1.157
Xuzhou	1.900					Charlotte	1.652			Toulouse	1.153
Urumqi	1.830					Providence	1.601			Bilbao	1.144
Xinyang	1.732					Austin	1.598			Hanover	1.116
Tangshan	1.700					Milwaukee	1.544			Dresden	1.077
Luoyang	1.500					Nashville	1.521			Dublin	1.064
Nantong	1.500					Jacksonville	1.301			Nottingham	1.056
Qiqihar	1.438					Memphis	1.281			Bordeaux	1.018
Datong	1.426					Louisville	1.234			Palermo	1.007
Nanning	1.400					Richmond	1.213				
Baotou	1.400					Oklahoma City	1.189				
Handan	1.390					Hartford	1.189				
Fushun	1.384					Buffalo	1.128				

China		Japan		Korea		United States		Canada		EU-27	
Hohhot	1.300					Birmingham	1.108				
Anshan	1.286					Salt Lake City	1.010				
Yangzhou	1.125					Raleigh	1.048				
Hefei	1.100					Rochester	1.030				
Xi'ning	1.029					New Orleans	1.030				
Zhenjiang	1.014										

Notes: Year of the data: China (2007), Japan (2008, projected), Korea (2005), United States (2007), Canada (2006), and EU-27 (the latest year available)
Source: http://wikipedia.org; The city population figures for Korea's cities are from the *National Statistical Office of the Republic of Korea*

6.2. Other Issues

The study has also raised a number of other issues and avenues for future research. First, we have used the DEA-Window method to analyze airport efficiency trends. In order to obtain a "smoothed surface" frontier over a multi-year period, one may further introduce the DEA windows into the MPI (Malmquist productivity index) analysis which, in the present paper, is based only on a one-year change.

Second, due to data limitation, we only considered three inputs and three outputs when estimating efficiency scores by DEA methods. In practice, different airports are very different in their operating characteristics and in their services provided. Several other measures have been considered in the literature (see Table 3). Exclusion of some inputs and outputs may yield biases in measuring airport productivity. For example, if the so-called "soft cost input," which is measured by all expenses not directly related to capital and personnel, is not included, the efficiency measurement may favor the airports that outsourced most of their services. Similarly, non-aeronautical businesses – i.e., concessions, retailing, car parking, advertising and other commercial services at an airport – are getting more important for airports around the world including major NEA airports (e.g., Zhang and Zhang, 1997, 2003; Yu, et al., 2008). The proportion of non-aeronautical revenue to the total revenue may vary significantly among airports: In general, the portion of non-aeronautical revenue for airports in China is still relatively low as compared to airports in Japan and Korea. For example, in 2006 the non-aeronautical revenue only contributed 27.3 percent to the total revenue of Beijing Capital International Airport; and this figure rose to 47.3 percent of total revenue at Guangzhou airport. Exclusion of non-aeronautical revenue as an output may bias productivity against the airports that generate more revenues from non-aeronautical businesses. Consequently, further research may need to consider extending our analysis by using a larger set of inputs and outputs in measuring the airports' efficiency.

Finally, it would be interesting, and important, to investigate factors that might influence the efficiency performance of airports in Northeast Asia. For example, how have country-specific factors affect airport productive efficiency? How various aviation liberalization policies – e.g., airport localization and privatization programs as well as domestic/international airline deregulation measures – affected airport productivity? Our new dataset might provide a basis for investigating the question. To achieve this goal, one might run regressions to examine the effects of, e.g., country-specific and policy liberalization factors on the efficiency scores while controlling for airport characteristics and other shifting variables. While such regression, either OLS, the Tobit model (Tobin, 1958) or other models, have been used extensively (see, e.g., some of the studies listed in Table 3 for airport applications, and Ali and Flinn (1989) and Kalirajan (1990) for applications in other industries), recent econometric work by Simar and Wilson (2007) suggests that strict application of regressions on DEA efficiency scores may not be apt. In our future work we shall follow Simar and Wilson (2007)'s procedure and conduct necessary econometric adjustments prior to the regression analysis.

References

Abbott, M. and Wu, S. (2002) Total Factor Productivity and Efficiency of Australian Airports, Australian Economic Review, 35, 244-260.

Ali, M. and Flinn, J. C. (1989) Profit Efficiency among Basmati Rice Producers in Pakistan Punjab, American Journal of Agricultural Economics, 71(2), 303-310.

Banker, R.D., Charnes, A. and Cooper, W.W. (1984) Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, Management Science, 30, 1078-1092.

Banker, R.D., Charnes, A., Cooper, W.W., Swarts, J. and Thomas, D.A. (1989) An Introduction to Data Envelopment Analysis with some of its Models and their Uses, in: J.L. Chan (Ed.), Research in Government and Nonprofit Accounting, 5, pp. 125-164, Greenwich, CT: JAI Press.

Bazargan, M. and Vasigh, B. (2003) Size versus Efficiency: A Case Study of US Commercial Airports, Journal of Air Transport Management, 9, 187-193.

Charnes, A., Cooper, W.W., Golany, B., Seiford, L. and Stutz, J. (1985) Foundation of Data Envelopment Analysis for Pareto-Koopmans Efficient Empirical Production Functions, Journal of Econometrics, 30, 91-107.

Charnes, A., Cooper, W.W. and Rhodes, E.L. (1978) Measuring the Efficiency of Decision-making Units, European Journal of Operation Research, 2, 429-444.

Charnes, A., Cooper, W.W. and Rhodes, E.L. (1981) Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow through, Management Science, 27, 668-697.

Chin, A.T.H. and Siong, L.E. (2001) Airport Performance: A Comparative Study between Changi Airport and Airports in the New York-New Jersey Metropolitan Area, Paper presented at the World Conference on Transport Research (WCTR), Seoul, Korea.

Farrell, M.J. (1957) The Measurement of Productive Efficiency, Journal of the Royal Statistics Society, Series A: General, 120, 253-281.

Fernandes, E. and Pacheco, R.R. (2002) Efficient Use of Airport Capacity, Transportation Research Part A, 36(3), 225-238.

Fung, M., Wan, K., Hui, Y.V. and Law, J. (2008) Productivity Changes of Chinese Airports 1995–2004, Transportation Research, E (Logistics), 44(3), 521-542.

Gillen, D. and Lall, A. (1997) Developing Measures of Airport Productivity and Performance: An Application of Data Envelopment Analysis, Transportation Research, E (Logistics), 33(4), 261-273.

Kalirajan, K.P. (1990) On Measuring Economic Efficiency, Journal of Applied Econometrics, 5 (1), 75-85.

Lam, S.W., Low, J.M.W. and Tang, L.C. (2009) Operational Efficiencies across Asia Pacific Airports, Transportation Research, E (Logistics), forthcoming.

Lee, Y.H. (2008) Open Sky and Current Issues in Korean Air Transport Policy, A paper presented at the International Forum on Shipping, Ports and Airports held at the Hong Kong Polytechnic University, May 2008, Hong Kong.

Lovell, C.A.K. (1993) Production Frontiers and Productive Efficiency, in H.O. Fried, C.A.K. Lovell and S.S. Schmidt (eds.), The Measurement of Productive Efficiency: Techniques and Applications, 3-67, New York: Oxford University Press.

Martin, J.C. and Roman, C. (2001) An Application of DEA to Measure the Efficiency of Spanish Airports prior to Privatization, Journal of Air Transport Management, 7, 149-157.

Martin, J.C. and Roman, C. (2006) A Benchmarking Analysis of Spanish Commercial Airports: A Comparison between SMOP and DEA Ranking Methods, Networks and Spatial Economics, 6, 111-134.

Murakami, H. (2008) An Empirical Analysis of Inter-firm Rivalry between Japanese Full-service and Low-cost Carriers, Pacific Economic Review, forthcoming.

O'Connell, J. F. and Williams, G. (2005) Passengers' Perceptions of Low Cost Airlines and Full Service Carriers: A Case Study Involving Ryanair, Aer Lingus, Air Asia and Malaysia Airlines, Journal of Air Transport Management, 11, 259-272.

Oum, T.H., Yamaguchi, H. and Yoshida, Y. (2009) Efficient Measurement Theory and its Application to Airport Benchmarking, in: A. de Palma, R. Lindsey, E. Quinet and R. Vickman (eds.), Handbook of Transport Economics, Edward Edgar, forthcoming.

Park, Y.H., (2000) Reducing Legal and Institutional Barriers to Liberalize the Air Transportation Market: Korea's Perspective, A paper presented at the EWC/KOTI Conference on Northeast Asia Transportation, August 2000, Hawaii.

Parker, D. (1999) The Performance of BAA before and after Privatization, Journal of Transport Economics and Policy, 33, 133-145.

Pels, E., Nijkamp, P. and Rietveld, P. (2001) Relative Efficiency of European Airports, Transport Policy, 8(3), 183-192.

Pels, E., Nijkamp, P. and Rietveld, P. (2003) Inefficiencies and Scale Economies of European Airport Operations, Transportation Research Part E, 39, 341-361.

Pestana Barros, C. and Dieke, P.U.C. (2007) Performance Evaluation of Italian Airports: Data Envelopment Analysis, Journal of Air Transport Management, 13, 184-191.

Poole, R.W., (2003), Global Airport Privatization Regains Altitude, Report, Reason Public Policy Institute, Los Angeles.

Sarkis, J. (2000) An Analysis of the Operational Efficiency of Major Airports in the United States, Journal of Operations Management, 18, 335-351.

Sarkis, J. and Talluri, S. (2004) Performance-based Clustering for Benchmarking of US airports, Transportation Research Part A, 38, 329-346.

Seiford, L.M. and Thrall, R.M. (1990) Recent Developments in DEA: The Mathematical Programming Approach to Frontier Analysis, Journal of Econometrics, 46, 7-38.

Seiford, L.M. (1996) Data Envelopment Analysis: The Evolution of the State of the Art (1978-1995), Journal of Productivity Analysis, 7, 99-137.

Simar, L. and Wilson, P.W. (2007) Estimation and Inference in Two-Stage, Semi-parametric Models of Production Processes, Journal of Econometrics, 136, 31-64.

Tobin, J. (1958) Estimating the Relationship for Limited Dependent Variable, Econometrica, 26, 24-36.

Takahashi, N. (2003) Airline Deregulation in Japan: Its Economic Effects and Issues, Kansai University Review of Business and Commerce, 5, 47-63.

Tretheway, M.W. (2004) Distortions of Airline Revenues: Why the Network Airline Business Model is Broken, Journal of Air Transport Management, 10, 3-14.

Vasigh, B. and Hamzaee, R. G. (2000) Airport Efficiency: An Empirical Analysis of the US Commercial Airports, Paper presented at the Fourth Annual Conference of the Air Transport Research Group, Amsterdam.

Yamaguchi, K. (2005) Inter-regional Air Transport Accessibility and Macro Economic Performance in Japan, A paper presented at the 9th Air Transport Research Society (ATRS) Conference, July 2005, Rio de Janeiro.

Yamauchi, H. (2000) Air Transport Policy in Japan: Policy Change and Market Competition, A paper presented at the EWC/KOTI Conference on Northeast Asia Transportation, August 2000, Hawaii.

Yoshida, Y. and Fujimoto, H. (2004) Japanese-airport Benchmarking with the DEA and Endogenous-weight TFP Methods: Testing the Criticism of Overinvestment in Japanese Regional Airports, Transportation Research Part E, 40, 533-546.

Yu, M.M., Hsu, S.H., Chang, C.C. and Lee, D.-H. (2008) Productivity Growth of Taiwan's Major Domestic Airports in the Presence of Aircraft Noise, Transportation Research Part E, 44, 543-554.

Zhang, A. (1998) Industrial Reform and Air Transport Development in China, Journal of Air Transport Management, 4, 155-164.

Zhang, A. and Chen, H. (2003) Evolution of China's Air Transport Development and Policy towards International Liberalization, Transportation Journal, 42, 31-49.

Zhang, A. and Yuen, A. (2008), Airport Policy and Performance in Mainland China and Hong Kong, in: Clifford Winston and Gines de Rus (eds.), Aviation Infrastructure Performance: A Study in Comparative Political Economy, pp. 159-192, Washington, D.C.: Brookings Institution Press.

Zhang, A. and Zhang, Y. (1997), Concession Revenue and Optimal Airport Pricing, Transportation Research, E (Logistics), 33, 287-296.

Zhang, A. and Zhang, Y. (2003) Airport Charges and Capacity Expansion: Effects of Concessions and Privatization, Journal of Urban Economics, 53, 54-75.

Zhang, A., Hanaoka, S., Inamura, H. and Ishikura, T. (2008) Low Cost Carriers in Asia: Deregulation, Regional Liberalization and Secondary Airports, Research in Transport Economics, 24, 36-50.

Zhang, A., Zhang, Y. and Clougherty, J.A. (2009) Competition and Regulation in Air Transport, in: A. de Palma, R. Lindsey, E. Quinet and R. Vickman (eds.), Handbook of Transport Economics, Edward Edgar, forthcoming.

Collective strategy and resilience in supply chains: opportunities for port managers

Hilary Haugstetter and Dr Stephen Cahoon

Department of Maritime and Logistics Management, Australian Maritime College, Tasmania 7250, Australia.

Abstract

Strategic collaborations in supply chain networks, of which ports are a significant node, reduce risks for all firms in the network, increase firm's ability to access efficiencies such as economies of scale and create opportunities for learning. The knowledge learnt from collaborators is enhanced by managing the key relationships and forming a strategic intent to learn for mutual benefit. Mutual benefits may include market information or innovation. As Awad and Ghaziri (2004 p. 17) comment 'Beyond efficiency and productivity, the real benefit of collaboration is innovation.' The more central the port is in this network, the greater are the opportunities for learning and resource sharing.

Three key factors are critical for port managers to integrate knowledge from networks; firstly, there are levels of common knowledge that enable effective communication, including concepts, behavioural norms, language and experience. Secondly integrative efficiency requires organisational routines which foster knowledge sharing through frequent, coordinated activities that are repetitive. Thirdly efficient structures within the firm help maintain effective and efficient communication to minimise the knowledge loss on integration.

Collective strategy provides opportunities for port managers to steer the critical path between maintaining their current sources of competitive advantage and developing future resilience through innovation opportunities provided by learning from their trading networks. Port managers, by developing skills in reading the collective mind and interpreting information available in trading networks, can foster resilience into the future.

Key words: strategic information system, collaboration, ports, innovation, knowledge sharing

1. Introduction

Ports have a significant role in today's networked business environment. In fact, they are increasingly being regarded as hubs that are part of various logistics systems (Meersman et al., 2005). They are effectively becoming part of complex value chains, with value 'created as much outside a company as within' (Gratton 2006, p.2). These networks enable value creation.

The value of any network for each individual organisation is dependent upon the network's size; a bigger network creates more positive externalities, which benefits all members (McGee and Sammut Bonnici, 2002; Shapiro and Varian, 1999). Networks facilitate the development of key relationships, for example strategic alliances, partnerships and collaborations: the significance of key relationships and a market orientation for firms' success have been highlighted (Jaworski and Kohli, 1993). A network of collaborative relationships is an intangible strategic asset in which business sustainability and growth are increasingly found (Doyle 1995; Johnson 1999; McCormack et al., 2003). Collaborations, then, are clearly an integral component of extended supply chains and occur both strategically and operationally, providing the potential for the development of long-term, high trust relationships (Sawney and Zabin, 2002). In a port environment, these relationships can be with strategic alliance partners and other key stakeholders, such as suppliers, competitors and community groups.

To develop intellectual capital, knowledge and skills can be transferred from collaborations and key relationships, such as with major customers and suppliers. According to Tranfield et al. (2004, p. 375) successful companies are those that “exploit knowledge, skills, and creativity.” Competitive advantage is no longer based on positioning and responses to market needs, but also strategically exploiting the firm’s intellectual capital (Appelbaum and Gallagher, 2000; McCormack et al., 2003; Sharkie, 2003; Von Krogh et al., 2001). Port organisations are well positioned to learn from multiple stakeholders. For a port though part of the paradox lies in whether to continue providing the current services and facilities or to innovate for new opportunities learnt from their multiple stakeholders (Stacey, 1996); innovation offers a port organisation resilience into the future. This paradox, between continuing exploitation of existing advantages and exploring for new ones through innovation, adds to the strategic complexity inherent in the management of a port.

Exploration is much less predictable, being riskier in terms of potential outcomes but is ‘the only way to finish first’ (Levinthal and March 1994 p. 106 in (Powell et al., 1996, p. 118). Yet as Powell et al. (1996, p.118) argue, ‘the neat distinctions of theory’ are often blurred by ‘the messy world of practice’. They continue that ‘Exploitation and exploration, and calculation and community are intertwined. Organizational learning is both a function of access to knowledge and the capabilities for utilizing and building on such knowledge’ (Powell et al., 1996, p.118). Ports, in their capacity as hubs, have many opportunities for collaborating and thinking strategically across boundaries. Capturing the knowledge and learning from these collaborations and opportunities for collective strategy and integrating knowledge from their networks into their strategic knowledge system is critical to port organisation’s sustainable competitive advantage and success.

2. Strategic Advantage

Increasingly the literature on supply chain management considers the supply chain as an integrative framework for strategic planning, partly driven by the dynamism of modern business (Sahay, 2003) and developments in information technology (Bowersox et al. 2007; Simatupang and Sridharan, 2004). Managers in supply chains that is, logistics networks, tend to focus on integration and analysis for decision making, with collaboration a key element in effectiveness and efficiency gains (Barratt 2004; Kampstra et al., 2006). Complex decisions need to be made in a changing environment on activities to continue to reap these benefits and ways to strategically innovate to ensure sustainability into the future. For port managers whose port may be a nexus in many different supply chains (Meersman et al., 2005), there is thus a need to strategically plan for both continuity and future changes.

Sustainability requires innovation for a firm to survive and prosper (Miles et al., 2005; Moccario Li Destri and Dagnino, 2004). By collaborating with supply chain members for innovations, the possibility of a supply chain retaining competitive advantage in the long term can be realised (Kim, 2005; Sahay, 2003)(Miles et al., 2005). Innovation provides the potential for a new competitive advantage to develop. Hamel and Prahalad (1989, p.69) suggest that ‘the essence of strategy lies in creating tomorrow’s competitive advantages faster than competitors mimic the ones you possess today.’ They argue that leveraging resources is crucial to strategic success, provided they are consistently applied towards achieving ‘seemingly unattainable goals’ which are the cornerstone of their strategic intent across the business at all levels and in all its functions (1989, p.65). In fact it is the continuous dialogue between different tiers of management within an organisation that enables strategy to evolve (Pascale, 1994). Middle managers then have a role to play in ports as part of the dialogic process.

The role of senior managers within this strategy process is changing. Formerly their role was controlling, but nowadays they create opportunities to challenge other managers and staff within the organisation to develop new capabilities by creating a ‘misfit between resources and ambitions’; ultimately these result in taking ‘the initiative away from better positioned players’ by constructing new advantages (Hamel and Prahalad 1989, p.67). Clear communication of the strategic intent through milestones and review procedures, providing staff with relevant training and the engagement of employees in meeting the challenge helps clarify the strategic direction.

Strategic intent, or 'vision', is fundamental to the strategic choices that firms make: it 'is a desired leadership position' and as such relates to the future position of the firm once it achieves its goals (Stacey, 1996, p.114). Unlike a firm's mission which relates to the firm's values, core purpose and current strategies, strategic intent is about future outcomes. It is the 'obsession with winning at all levels of the organization', but is greater than solely obsession – it '*captures the essence of winning*', is '*stable over time*' and sets a target that '*deserves personal effort and commitment*', motivating managers and employees throughout the organisation (Hamel and Prahalad 1989, p.64-66)(original emphasis). Clear communication of that intent is crucial in the strategy process to enable port managers throughout the organisation to achieve the port's strategic intent.

Fundamental to communication throughout the strategy process is a strategic information system; it not only gathers and disseminates information on the external environment and internal skills and resources, it also integrates the information across the organisation (Viljoen and Dann, 2003). Challenging senior managers today is the ability to create middle managers who have 'a working understanding of the concepts as well as the activities and processes of strategic management' (Viljoen and Dann, 2003, p.57).

Strategic thinking, 'the 'understanding or grasp of strategy creates a mental pattern, or framework, for dealing with strategic issues' is far more important than strategic planning, but the two are mutually dependent to some extent (Viljoen and Dann, 2003, p.57). Thinking is central to the process of transforming information into knowledge via insights learned from experience, effectively applying mental models (Senge, 1990). Hence managers and their ability to think are a key element in an organisation's learning system and its capacity for innovation. Learning can be considered vital to any creative activity, such as solving managerial problems and innovating in supply chains. It extends the 'capacity to create, to be part of the generative process of life' (Senge, 1990, p. 14). Levels of common knowledge, learning systems, strategic thinking skills and concepts enable effective communication and the achievement of strategic intent in port organisations.

Over time, network evolution changes the nature of the interactions, leading to a process of learning and systematising actions for each and every organisation within that network (Charan 1991). These organisational routines contribute to the integration of the knowledge learnt; each network participant has to manage the strategic issue of ensuring that these processes capture the network evolution that leads to positive outcomes (Draulans et al., 2003). Building capabilities and routines provides strategic advantage (Powell et al., 1996). Strategically managing routines and their role in the network so that knowledge and future opportunities can be effectively integrated is vital for a port organisation's development and growth. Effectively a superior system of learning permits the product/service innovation cycle to continue, providing a sustainable competitive advantage: it constitutes a dynamic capability (Helfat and Raubitschek 2000). The strategic information system within a port organisation supports the learning system.

Port organisations are not strategising in isolation. Mintzberg (1987, p.21) argues 'strategy is not just a notion of how to deal with an enemy or a set of competitors or a market' but '[I]t also draws us into some of the most fundamental issues about organizations as instruments for collective perception and action.' Taking the view that strategy is a perspective then 'intriguing questions about intention and behavior in a collective context' are raised (Mintzberg, 1987, p.21). As such collaborations in trading networks maybe considered with a collective mind. The collective mind unites strategic thinking and behaviour. However a critical issue it raises is 'how to read that collective mind-to understand how intentions diffuse through the system called organization to become shared and how actions come to be exercised on a collective yet consistent basis' (Mintzberg, 1987, p.17). Collaborations in logistics networks provide a collective context with many challenges and benefits to port managers; these are explored next.

3. Collaborations

Collaborating with fellow participants, whether formally in a partnership, or informally, is a key part of networked business in modern supply chains: in fact, some authors argue that supply chain management is the management of relationships (Håkansson and Persson 2004), and that it is 'ultimately about influencing behaviour' (Storey et al., 2006, p.754).

Collaboration itself is a broad term that appears to encompass a range of meanings, particularly in the context of supply chain management (Barratt 2004). Mentzer et al. (2000, p.53) found support among supply chain practitioners that collaboration means 'all companies [in the supply chain] are actively working together as one toward common objectives.' Central to collaboration is a focus on sharing.

Within a collaborative supply chain, knowledge is broadly distributed among diverse interorganisational relationships: the locus of innovation is then in this network. Firms with greater capacity to learn develop more advantages from collaborating – effectively such firms gain synergistic advantages over and above the more publicised benefits of collaborating (Powell et al., 1996). Collaboration then can be considered an underpinning tenet of knowledge management – knowledge sharing is dependent on effective and continuous collaboration (Laycock 2005). In a complex port environment this provides challenges to management.

Participant firms are increasing their learning and integrating capability, for example by collaborating for innovations. Such collaborative entrepreneurship requires member firms making investments in collaborative capability within the network, individual firms and often in society too (Miles et al., 2005) . The ability of a port to manage its collaborative network is thus a core competence (Ritter, 1999; Ritter et al., 2004; Song, 2003). As Awad and Ghaziri (2004, p. 17) comment 'Beyond efficiency and productivity, the real benefit of collaboration is innovation.' Within the innovation literature for example, there is a body of work on external knowledge acquisition; this research is mainly quantitative and 'rarely examines the nature of relationships developed in these networks and is unable to assess the managerial motivation and strategic intent which animate network formation' (Lane and Probert 2007, p.6). Quite clearly the role of strategic intent in the management of these relationships is critical for innovation.

For it to be a sustainable venture, though, Barratt (2004, p.39) indicates that some key strategic elements need to be present, particularly 'a shared understanding of what supply chain partners are collaborating over, clearly defined processes, and a clear understanding of the information required to populate such process.' This in turn requires processes and technology to support the collaboration and resources and commitment assigned to the collaboration. Importantly, it also needs corporate focus (Ireland and Bruce 2000; Sabath and Fontanella 2002). Given the significance of sustainability to continuity, Barratt (2004) calls for a better understanding of the interdependencies of these critical collaborative elements. To arrive at a collaborative culture, change both intra- and inter-firms may be required, as processes need to be aligned, relevant metrics developed and conflict resolution processes need to be developed (Barratt 2004). Common levels of understanding of the necessary structures to effect efficient communication within the port will reduce incremental losses in the quality of the information on its integration.

4. Knowledge Sharing

An organisation can be regarded as a knowledge system (Tsoukas and Mylonopoulos, 2004b) with new knowledge 'created out of human interactions' (Tsoukas and Mylonopoulos, 2004a, p. 14). Considering a firm as a knowledge system recognises that all of its work involves knowledge which resides in employees as 'a constantly developing set of generalizations, collective understandings and experiences' (Tsoukas and Mylonopoulos 2004a, p. 11). This recognises 'the crucial role of human interpretation, communication, and skills in generating effective organizational action' (Tsoukas and Mylonopoulos 2004a, p. 13). Port managers then are pivotal in ensuring that actions occur in line with the port's strategic intent, at all levels of management.

Firms have developed as integrators of knowledge for effective and efficient management of production processes, creating efficiencies through specialisation. One tenet from the strategic management literature of particular relevance to ports and their logistics networks is that firms with the ability to effectively and efficiently integrate distributed knowledge are more likely to be sustainable (Martin 2006). The firm is thus an institution for knowledge integration 'based upon close integration between organizational members, implying stability, propinquity and social relationships, but it does not readily yield a [sic] precision definition of the firm and its boundaries' (Grant 1996, p. 377). In today's networked environment firms boundaries are becoming ever more blurred. Port management has a range of sources for knowledge sharing; integrating that knowledge can be a challenge.

Knowledge integration, rather than the knowledge itself, is considered to be the source of the competitive advantage: individuals who hold knowledge can easily transfer to other firms and are more likely to accumulate the economic rents accruing from it (Grant 1996). Grant (1996) considers there are three characteristics of knowledge integration that contribute to the firm's ability to develop competitive advantage namely the efficiency of integration, the scope of integration and the flexibility. Efficiency of integration is 'the extent to which the capability accesses and utilizes the specialist knowledge held by individual organizational members' (Grant 1996, p. 380). Its scope is 'the breadth of specialized knowledge the organization draws upon' (Grant 1996, p. 380). And the flexibility relates to 'the extent to which a capability can access additional knowledge and configure existing knowledge' (Grant 1996, p. 380). These three key characteristics of effective knowledge integration are critical to a port organisation's ability to learn from its networks.

The firm's ability to integrate this specialised knowledge for the creation of customer value is considered by Grant (1996, p. 377) to be 'the essence of *organizational capability*, defined as a firm's ability to perform repeatedly a productive task, which relates either directly or indirectly to a firm's capacity to create value through effecting the transformation of inputs into outputs.' (Original emphasis.) For example in ports logistics value is created by instigating processes that reduce waiting times.

Knowledge as a firm's resource, can be considered a key contributor to a firm's potential for growth: it functions as a 'stepping stone(s) to further expansion' (Wernerfelt, 1984). It can be integrated into the firm through two distinct approaches. Firstly it can be integrated by direction, such as a firm's operating manual for the order processing system or by policies and procedures for maintenance. Secondly organisational routines provide a second approach that beneficially results in less knowledge loss, as it is unnecessary to specifically communicate the knowledge. An organisational routine develops 'sequential patterns of interaction' and as such does not need any inherent specialised knowledge to be made explicit (Grant 1996, p. 379). However identifying resources can provide practical difficulties, knowing the extent to which one can, 'in practice combine capabilities across operating divisions, or about how one can set up a structure and systems' to help execute strategy (Wernerfelt, 1984, p. 180). In a multi-party environment such as a port organisation this can be difficult.

The organic nature of knowledge, both in use and in its creation, means that it 'thrives when it is cultivated, cross-pollinated, contaminated, and [sic] fertilized' (Tsoukas and Mylonopoulos 2004a, p. 13). As such, knowledge is part of a dynamic process, which occurs in the context of social interaction in firms. This intangible resource is fundamental to sustainable competitive advantage and raises three crucial issues for its management (Tsoukas and Mylonopoulos 2004a). In a port these are, firstly, the ways that managers can share their knowledge of both the port and its environment, secondly, the methods of organising the port's structure to facilitate the development of knowledge and skills and thirdly 'how, and the extent to which, individuals are willing to re-arrange and re-order what they know' to enable 'the emergence of new knowledge and innovation' (Tsoukas and Mylonopoulos 2004a, p. 13). Strategically planning directions and routines to facilitate knowledge sharing and creation through effective management of patterns of social interaction, encouraging learning and the nurturing of mental flexibility are critical to sustainable competitive advantage for port organisations.

Hansen et al. (1999, p. 114) argue that 'competitive strategy must drive knowledge management strategy' to be effective. Effective knowledge management provides the means to identify and develop the necessary people and systems: organisations can then be agile and responsive (Awad and Ghaziri 2004). By accumulating knowledge firms can effectively compete (Appelbaum and Gallagher 2000; Sharkie, 2003). Managing this knowledge requires some formal process that provides easy-access to the organisational memory, or intellectual capital, and simple methods to add to the knowledge bank (Zikmund, 2003). And to use the acquired knowledge within the firm's networks, it needs to be able to transfer knowledge both between and within firms. Knowledge sharing becomes a critical capability for a port in today's global business environment.

Knowledge sharing and the contribution that knowledge management can make towards innovation and improved levels of performance and productivity has increased management's focus on this resource. Knowledge sharing creates 'exponential benefits from the knowledge as people learn from it' (Awad and Ghaziri 2004, p. 10). It is an evolving, social resource that managers need to strategically analyse and manage (Von Krogh et al., 2001). By its evolution, new knowledge is created: knowledge creation then is vital to innovation.

Von Krogh et al., (2001) consider that knowledge creation is a human process, both social and individual, which is based on experience and may involve feelings and belief systems. Given that a firm's networks comprise social relationships, knowledge resides therein. With the geographical range of trading networks, complexity is added to knowledge transfer, management and creation between internal and external networks (Lindsay, Chadee, Mattsson, Johnston and Millett 2003). Close relationships enable effective knowledge transfer to occur (Cavusgil et al., 2003), further endorsing the need for relationship management for knowledge creation.

Relationships are based on social ties; through these ties information and resources can flow (Tsai and Ghoshal, 1998). In interfirm relationships, organisations gain access to other members of the network's resources and tacit knowledge and the potential to combine resources more effectively. The more central the actor is in this network, the greater are the opportunities for learning and resource sharing. And by having access to multiple ways of exchanging and combining resources in a business network, new sources of value can be generated (Sahay, 2003; Sawney and Zabin, 2002; Tsai, 2002; Tsai and Ghoshal, 1998; Zairi, 1999). Ports are central nodes in many diverse supply chains and are therefore uniquely placed to create new value and resilience into the future.

From a senior manager's perspective, employees having access to wider networks across more permeable boundaries is an efficient way to access more inputs into decision-making processes, such as strategic innovation (Ashkenas 1999). These wider networks enable more opportunities for value creation and innovation, depending on the strength, reach and sustainability of the inherent relationships (Gratton 2006). For example, many multinational firms recognise the value of concentrating on core competencies and utilising an extended value network through outsourcing and other collaborative ventures, such as those utilised by large pharmaceutical companies Pfizer and GlaxoSmithKline where R & D networks are extensive both for innovation and for clinical trials. The formation of these networks are frequently driven by cost-savings, but to become effective they also 'need to be well-managed, collaborative, and evolutionary' (Ashkenas, 2003, p.4). And finally, accessibility through technology has made it possible to communicate with anyone, anytime and anywhere. But this ability to communicate needs to be effectively managed – sending wrong information is not constructive, neither is information overload. Mechanisms are required so that knowledge can be captured and leveraged effectively across the firm to share learning and best practices (Ashkenas 1999). Leadership effectively becomes distributed, with firms 'including strategic education for all employees at all levels' (Ashkenas 1999, p.7). This is where strategic intent needs to be communicated effectively – as part of the leadership function in permeable firms.

Equally important to knowledge creation is support from senior management through the creation of microcommunities for effective knowledge sharing. Central to these microcommunities are five knowledge enablers, comprising instilling a vision, management of conversations, mobilizing activists, creating an appropriate context and globalizing local knowledge (Von Krogh, Ichijo & Nonaka 2000). These facilitate the sharing of knowledge across an organisation and cannot occur without a caring atmosphere, as the good relations required for care break down communication barriers such as distrust and fear (Von Krogh et al., 2000).

Each firm has a unique stock of specialised knowledge and history; there are three key factors for efficient integration of new knowledge (Grant 1996). Firstly, the level of common knowledge between different specialists is fundamental to their ability to communicate effectively, including concepts, behavioural norms, language and experience. For collaborators/boundary spanners in a firm who will have less commonality with their counterparts in other firms this can be a point of inefficient knowledge integration and possibly high knowledge loss (Grant 1996). By maintaining social networks with peers in other firms this risk of loss can be reduced as communication increases.

Secondly integrative efficiency requires organisational routines which foster knowledge sharing through frequent, coordinated activities that are repetitive (Grant 1996). This develops communication skills and the ability to interpret incoming messages. Thirdly Grant (1996) contends that efficient structures within the firm are necessary to maintain effective and efficient communication to minimise the knowledge loss on integration.

5. Conclusions

Strategic collaborations in supply chain networks provide opportunities for learning that can be used strategically for growth and innovation. Port organisations, with a central role in many supply chains and a wide range of stakeholders are ideally situated to access knowledge that is strategically useful for knowledge creation.

To successfully integrate that knowledge into their strategic planning system, port organisations can ensure that there are levels of common knowledge, behavioural norms, language and experience, including strategic thinking skills, at all management levels throughout the organisation. Additionally establishing routines that foster knowledge sharing through coordinated activities across multi-function teams may assist knowledge integration. Finally efficient communication systems at all levels, including the strategic information system, may contribute to the effective and efficient flow of key information from the port organisation's trading network.

Collaborating provides many opportunities for growth and collective strategy can be a key source of innovative opportunities. It is one additional strategic element from which port managers can steer the critical path between maintaining their current sources of competitive advantage and developing future resilience through innovation opportunities provided by learning from their trading networks. Port managers, by developing skills in reading the collective mind and interpreting information available in trading networks, can foster resilience into the future.

References

Appelbaum, S.H. and Gallagher, J. (2000), The competitive advantage of organizational learning, Journal of Workplace Learning, 12(2): 40-56.

Argyris, C. (1993), On organizational learning, Blackwell Publishers, Cambridge, Ma., USA.

Ashkenas, R. (1999), Creating the boundaryless organization, Business Horizons, 42(5): 5-10.---- (2003), Crossing boundaries, Executive Excellence, 20(1): 3-4.

Awad, E.M. and Ghaziri, H.M. (2004), Knowledge management Pearson Prentice Hall, Upper Saddle River, N.J.

Barratt, M. (2004), Understanding the meaning of collaboration in the supply chain, Supply Chain Management, 9(1): 30-42.

Bowersox, D. J., Closs, D.J. and Cooper, M.B. (2007), Supply chain logistics management, 2nd edn, McGraw-Hill/Irwin, New York.

Cavusgil, S.T., Calantone, R.J. and Zhao, Y. (2003), Tacit knowledge transfer and firm innovation capability, The Journal of Business and Industrial Marketing, 18(1): 6-21.

Charan, R. (1991), How networks reshape organizations - for results, Harvard Business Review, 69(5): 104-115.

Doyle, P. (1995), Marketing in the new millennium, European Journal of Marketing, 29(13): 23-41.

Draulans, J., deMan, A.-P. and Volberda, H.W. (2003), Building alliance capability: management techniques for superior alliance performance, Long Range Planning, 36(2): 151-166.

Grant, R.M. (1996), Prospering in dynamically-competitive environments: organizational capability as knowledge integration, Organization Science, 7(4): 375-387.

Gratton, L. (2006), Connections and conversations provide the fuel for innovation, Financial Times, Friday March 31, pp. 2-3.

Håkansson, H. and Persson, G. (2004), Supply chain management: the logic of supply chains and networks, The International Journal of Logistics Management, 15(1): 11-26.

Hamel, G. and Prahalad, C.K. (1989), Strategic intent, Harvard Business Review, 67(3): 63-76.

Hansen, M.T., Nohria, N. and Tierney, T. (1999), What's your strategy for managing knowledge? Harvard Business Review, 77(2): 106-116.

Helfat, C.E. and Raubitschek, R.S. (2000), Product sequencing: co-evolution of knowledge, capabilities and products, Strategic Management Journal, 21(10/11): 961-979.

Ireland, R. and Bruce, R. (2000), Only the beginning of collaboration, Supply Chain Management Review, 4(4): 80-88.

Jaworski, B. and Kohli, A. (1993), Market orientation: antecedents and consequences, Journal of Marketing, 57(3): 53-70.

Johnson, J.L. (1999), Strategic integration in industrial distribution channels: managing the interfirm relationship as a strategic asset, Academy of Marketing Science, 27(1): 4-18.

Kampstra, R.P., Ashayeri, J. and Gattorna, J. (2006), Realities of supply chain collaboration, Tilburg University.

Lane, C. and Probert, J. (2007), The external sourcing of technological knowledge by US pharmaceutical companies: strategic goals and inter-organizational relationships, Industry and Innovation, 14(1): 5-25.

Laycock, M. (2005), Collaborating to compete: achieving effective knowledge sharing in organizations, The Learning Organization, 12(6): 523-538.

Martin, G. (2006), Managing people and organizations in changing contexts, Elsevier Ltd, Oxford.

McCormack, K., Johnson, W. and Walker, T. (2003), Supply Chain Networks and Business Process Orientation Advanced Strategies and Best Practices, CRC Press LLC, Boca Raton, Florida.

McGee, J. and Sammut Bonnici, T.A. (2002), Network Industries in the New Economy, European Business Journal, 14(3): 116-132.

Meersman, H., Van de Voorde, E. and Vaneslander, T. (2005), Ports as hubs in the logistics chain. In H. Leggate, J. McConville and A. Morvillo (eds), International maritime transport: Perspectives, Routledge, Abingdon, Oxon, pp.137-144.

Mentzer, J.T., Foggin, J.H. and Golicic, S.L. (2000), Collaboration the enablers, impediments, and benefits, Supply Chain Management Review, 4(4): 52-58.

Miles, R.E., Miles, G. and Snow, C.C. (2005), Collaborative entrepreneurship: how communities of networked firms use continuous innovation to create economic wealth, Stanford University Press, Stanford, California.

Mintzberg, H. (1987), The strategy concept I: five Ps for strategy, California Management Review, 30(1): 11-24.

Moccario Li Destri, A. and Dagnino, G.B. (2004), Time and strategy: towards a multitemporal view of the firm, Journal of Managerial Psychology, 19(8): 776-794.

Pascale, R.T. (1994), Perspective's on strategy: the real story behind Honda's success, California Management Review, 26(3): 47-72.

Powell, W.W., Koput, K.W. and Smith-Doerr, L. (1996), Interorganizational collaboration and the locus of innovation: networks of learning in biotechnology, Administrative Science Quarterly, 41: 116-145.

Ritter, T. (1999), The networking company antecedents for coping with relationships and networks effectively, Industrial Marketing Management, 28: 467-479.

Ritter, T., Wilkinson, I.F. and Johnston, W.J. (2004), Managing in complex business networks, Industrial Marketing Management, 33: 175-183.

Sabath, R.E. and Fontanella, J. (2002), The unfulfilled promise of supply chain collaboration, Supply Chain Management Review, 6(4): 24-30.

Sahay, B.S. (2003), Supply chain collaboration: The key to value creation, Work Study, 52(2/3): 76.

Sawney, M. and Zabin, J. (2002), 'Managing and measuring relational equity in the network economy', Journal of the Academy of Marketing Science, 30(4): 313-332.

Senge, P.M., (1990), The fifth discipline: the art and practice of the learning organization, Doubleday, New York, NY.

Shapiro, C. and Varian, H.R. (1999), Information Rules: a Strategic Guide to the Network Economy Harvard Business School Press, Boston, MA.

Sharkie, R. (2003), Knowledge creation and its place in the development of sustainable competitive advantage, Journal of Knowledge Management, 17(1): 20-31.

Simatupang, T.M. and Sridharan, R. (2004), A benchmarking scheme for supply chain collaboration, Benchmarking, 11(1): 9-30.

Song, D. (2003), Port co-opetition in concept and practice, Maritime Policy & Management,30(1): 29-44.

Stacey, R.D. (1996), Strategic management and organisational dynamics, 2nd edn, Pitman Publishing, London.

Storey, J., Emberson, C., Godsell, J. and Harrison, A. (2006), Supply chain management: theory, practice and future challenges, International Journal of Operations & Production Management, 26(7): 754-774.

Tranfield, D., Denyer, D., Marco, J. and Burr, M. (2004), Co-producing management knowledge, Management Decision, 42(3/4): 375-386.

Tsai, W. (2002), Social structure of "Coopetition" within a multiunit organization: coordination, competition, and intraorganizational knowledge sharing, Organization Science, 13(2): 179-190.

Tsai, W. and Ghoshal, S. (1998), Social capital and value creation: the role of intrafirm networks, Academy of Management Journal, 41(4): 464-476.

Tsoukas, H. and Mylonopoulos, N. (2004a), Introduction: what does it mean to view organizations as knowledge systems? In H. Tsoukas and N. Mylonopoulos (eds), Organizations as knowledge systems knowledge, learning and dynamic capabilities, Palgrave Macmillan, Basingstoke, UK, pp. 1-26.

---- (eds) (2004b), Organizations as knowledge systems knowledge, learning and dynamic capabilities, Palgrave Macmillan, Basingstoke, UK.

Viljoen, J. and Dann, S. (2003), Strategic management, 4th edn, Pearson Education Australia, Frenchs Forest, NSW, Australia.

Von Krogh, G., Ichijo, K. and Nonaka, I. (2001) Enabling knowledge creation: how to unlock the mystery of tacit knowledge and release the power of innovation, Oxford University Press, Oxford.

Wernerfelt, B. (1984), A resource-based view of the firm, Strategic Management Journal, 5(2): 171-180.

Zairi, M. (1999), Total supply chain for innovation, In M. Zairi (ed.), Best practice process innovation management, Butterworth-Heinemann, Oxford, pp. 235-282.

Zikmund, W.G. (2003), Business Research Methods, 7th edn, Thomson South-Western, Mason, Ohio.

Addressing modern piracy and maritime terrorism in Southeast Asia: a legal perspective

Nong Hong

The China Institute, University of Alberta, Canada.

Abstract

The frequent Somali pirates have instigated a global discussion on how to confront and tackle this problem militarily, politically and legally. The military and political measures are well known, but the legal ones are not, because there is a certain amount of uncertainty and vagueness. This paper tries to answer the question: what can international law do in excising its role to prevent, control and punish piracy and maritime terrorism, particularly in the Southeast Asia waters?

In the aftermath of 9.11 terrorist attacks, serious concern was raised in the shipping community in respect of the prospect of terrorist attacks against ships or against targets such as port facilities by using ships as terrorist weapons, pretty much in the same way the planes were used as weapons against the Twin Towers. In other words, "terrorism goes to the sea". (Luft & Korin, 2004). In addition, as a traditional problem, piracy has never been eliminated; on the contrary, piracy on the high seas nowadays is becoming a key tactic of terrorist groups. Many of today's pirates are maritime terrorists with an ideological bent and a broad political agenda. This intertwining of piracy and terrorism poses great dangerous for energy markets because most of the world's oil and gas is shipped through the world's most piracy-infested water. Maritime security becomes an important agenda for most of maritime states whose national economy might get seriously affected by the insecure sea-lanes threatened by modern piracy and maritime terrorism. Among the maritime security agenda, combating piracy and maritime terrorism has become an urgent mission for both the littoral states and the user states.

Given the complex nature of piracy and maritime terrorism in the Southeast Asia, there is no simple solution to this problem. It is a complex, multifaceted problem that requires a multifaceted solution. This paper addresses this problem from a legal perspective and tries to identify the challenges and put forward some suggestions on how to solve this globally concerned threat to maritime security.

1. Modern Piracy and Maritime Terrorism in the Southeast Asia Seas

Piracy has existed for nearly as long as people have sailed the ocean. It remains a serious threat to international commerce and safety in modern time, especially in Southeast Asian archipelago where piracy has been a nagging problem for centuries. Commercial ships in this region have always been particularly vulnerable to piracy due to the narrow waterways and countless small islands that define the region's geography. The situation gets even worse now in the Malacca Strait and Singapore Strait due to its location as a busy and significant sea lane. Every day, a quarter of world trade, including half of all sea shipments of oil bound for eastern Asia and two-thirds of global shipments of liquefied natural gas, passes through this strait. Roughly 600 freighters loaded with everything from Japanese nuclear waste bound for reprocessing facilities in Europe to raw materials for china's booming economy traverse this chokepoint daily. Any disruption of shipping in the South China Sea would harm not only the economies of China, Japan, South Korea, Taiwan, and Hong Kong, but that of the United States as well.

There is a wide speculation that some of the piratical incidents which occurred in Southeast Asia involved organized criminals since the hijacking of a whole ship and resale of its cargo requires huge resources and detailed planning (Zou, 2005). Individual pirates do not have these resources. Luft and Korin (2004) claim that the string of maritime attacks perpetrated in recent years demonstrates that terror has indeed gone to sea.

2000.01	Al Qaeda attempted to ram a boat loaded with explosives into the USS The Sullivans in Yemen
2000.10	Al Qaeda suicide bombers in a speedboat packed with explosives blew a hole in the USS Cole, killing 17 sailors.
2001.10	Tamil Tiger separatists carried out a coordinated suicide attack by five boats on an oil tanker off northern Sri Lanka.
2002.06	Moroccan government arrested a group of al Qaeda operatives suspected of plotting raids on British and US tankers passing through the Strait of Gibraltar.
2002.10	An explosives-laden boat hit the French oil tanker Limbrug off the coast of Yemen
2004.02	The southern Philippines-based Abu Sayyaf claimed responsibility of an explosion on a large ferry that killed at least 100 people.

Terrorist groups such as Hezbollah, Jemaah Islamiyah, the Popular Front for the Liberation of Palestine-General Command, and SriLanka's Tamil Tigers have long sought to develop a maritime capability. Intelligence agencies estimate that Al Qaeda and its affiliates now own dozens of phantom ships-hijacked vessels that have been repainted and renamed and operate under false documentation, manned by crews with fake passports and forged competency certificates. Security experts warn that the terrorists may create a hazard by bumping a vessel containing dynamite or even mass destructive weapons into a port or a harbor. Such attacks will halt the international commerce and lead to huge loss of billions of US dollars. It is well known that the most efficient way for terrorists to intervene the global economy is to attack the oil supply. It is relatively easier to protect the attacked targets in the land; however, the oil supply through maritime transport is mostly vulnerable to maritime terrorists. Sixty percent of the world's oil supply is through about 4000 oil tankers which are old and slow. They become the targets for maritime terrorists. If a single tanker were attacked on the high seas, the impact on the energy market would be marginal. But geography forces the tankers to pass through strategic chokepoints, many of which are located in areas where terrorists with maritime capabilities are active. These channels-major points of vulnerability for the world economy-are so narrow that a single burning supertanker and its spreading oil slick could block the route for other vessels. After 9.11, al-Qaedais has moved some of their bases onto the sea and set up a "terrorism fleet" composed of 20 vessels, scattering at the Arab Sea and Indian Ocean. An expert on terrorism said at a maritime security meeting in Singapore on November 29 2005, al-Qaedais is found to be developing underwater attack technology. In fact, since September 11, 2001, strikes on oil targets have become almost a routine. In October 2001, Tamil Tiger separatists carried out a coordinated suicide attack by five boats on an oil tanker off northern SriLanka. Oil facilities in Nigeria, the united State's fifth-largest oil supplier, have undergone numerous attacks (Luft and Korin, 2004). Ominously, there have been cases of terrorist pirates hijacking tankers in order to practice steering them through straits and crowded sea lanes-the maritime equivalent of the September 11 hijacker's training in Florida flight schools (Luft and Korin, 2004).

Most disturbingly, the scourges of piracy and terrorism are increasingly intertwined: piracy on the high seas is becoming a key tactic of terrorist groups. Unlike the pirates of old, whose sole objective was quick commercial gain, many of today's pirates are maritime terrorists with an ideological bent and a broad political agenda. They are often trained fighters abroad speedboats equipped with satellite phones and global positioning systems and armed with automatic weapons, antitank missiles, and grenades.

Pirates and Islamist terrorist groups have long operated in the same areas, including the Arabian Sea, the South China Sea, and in waters off the coast of western Africa. Since the international community has worked hard to freeze the capital of terrorist's group, they tend to get fund through activities of pirates. The interrelation between piracy and terrorisms poses great threat to the energy market since oil and natural gas transportation are mostly through the areas where piracy happened most frequently. This appeal is particularly apparent in the Strait of Malacca, the 500-mile corridor separating

Indonesia and Malaysia, where 42 percent of pirate attacks took place in 2003. Detained senior members of Jemaah Islamiyah, the al Qaeda-linked Indonesian terrorist group, have admitted that the group has considered launching attacks on Malacca shipping (Luft and Korin, 2004). Uniformed members of the Free Aceh Movement, and Indonesian separatist group that is also one of the most radical Islamist movements in the world, have been hijacking vessels and taking their crews hostage at an increasing rate. In some cases, the Free Aceh Movement has demanded the release of members detained by the government in exchange for hostages (Abuza, 2003)

With the crackdown on Middle Eastern funding mechanisms, especially the financial centuries in Abu Dhabi and other parts of the United Arab emirates, Al Qaeda has increasingly relied on Southeast Asia to move its money and hide its assets (MacCartnery and Cameron-Moore, 2003). Although the war on terror has continued apace in Southeast Asia since the September 11 attacks on the United States, little has been done to disrupt the terrorist financial networks in Southeast Asia. Weak domestic legislation, resource-strapped financial investigative agencies, poor enforcement capacity, and a lack of political will have hampered this important front in the war on terror (Abuza). To that end, Southeast Asia likely remains an important financial hub for the Al Qaeda organization and governments must be more proactive in their investigations and oversight.

2. Legal Approach: International Law on Piracy and Maritime Terrorism

What is the role of international law in the battle against piracy and maritime terrorism in terms of providing legal protection to shipping? Such legal protection of shipping seems to result from the rules of international law applicable to sea piracy, as contained in Articles 15-19 of the 1958 High Seas Convention (HSC) which was the first legal instrument to codify such rules, and the 1982 Convention on the Law of the Sea, which reproduces the same regime in its Articles 100-107 and the 1988 SUA Convention which regulates, as amongst states parties, unlawful acts against the safety of maritime navigation (Jesus, 2003). Besides, the 2005 SUA Amendment, the PSI Interdiction Principles, and the U.S. Ship Boarding Agreement have been developed after 911 as efforts against maritime terrorism. This section is going to analyze the contribution and limits of these mentioned conventions or agreements.

2.1. The United Nations Convention on the Law of the Sea (UNCLOS)

International law has established an obligation on States to cooperate in suppression of piracy and grants States certain rights to seize pirate ships and criminal. UNCLOS is a major anti-piracy treaty in contemporary ere with the following relevant provisions. Article 100 provides that “All States shall cooperate to the fullest possible extent in the repression of piracy on the high seas or in any other place outside the jurisdiction of any State”. Article 105 provides that “on the high seas, or in any other place outside the jurisdiction of any State, every State may seize a pirate ship or aircraft, or a ship or aircraft taken by pirate and under the control of pirates, and arrest the persons and seize the property on board”. The rules of piracy provide an exception to the principle of exclusive jurisdiction of the flag state on the high sea (Beckman, 2002). They give warships of all states the right, on the high seas or in an EEZ, to seize a pirate ship, to arrest the pirates, and to seize the property on board the pirate ship. The rules on piracy also give the state whose warship has seized the pirate ship the right to prosecute the pirates in its courts under its national law.

According to the UNCLOS definition on piracy, it consists of five elements: (1) the acts complained against must be crimes of violence such as robbery, murder, assault or rape; (2) committed on the high seas beyond the land territory or territorial sea, or other territorial jurisdiction, of any State; (3) by a private ship, or a public ship which through mutiny or otherwise is no longer under the discipline and effective control of the State which owns it; (4) for private ends; and (5) from one ship to another so that two ships at least are involved (Goldie, 1988).

However, the definition provided for in the UNCLOS has limitation in respect to the phenomenon of piracy. First of all, for an illegal act of violence or detention or any act of depredation against a ship to be considered an act of piracy it also has to meet the “private ends” requirement. That is to say, UNLOCS defines “piracy” as only for “private ends”, though it is argued that such wordings could be given a wider interpretation. Therefore, the terrorist acts at sea for political ends are generally excluded (Zou, 2005). This requirement seems to exclude sheer politically motivated acts directed at ships or their crew from the definition of piracy. Secondly, according to the above definition, piracy *juris gentium* presupposes that a criminal act be exercised by passengers or the crew of a ship against another ship or persons or property on its board. The two-vessel requirement is an ingredient of the crime of piracy, unless a criminal act occurs in *terra nullius*. Thus “internal seizure” within the ship is hardly regarded as “act of piracy” under the definition of the LOS Convention. Jesus (2003) argues that the piracy definition does not and was not supposed to contemplate the one-ship situation (2003). Thirdly, since the above definition is only applicable to the acts of piracy in the high seas (traditional vessel-specific exceptions of exclusive flag states' jurisdiction or non intervention of free navigation) or places outside jurisdiction of States, it has a geographic limitation and could not cover the whole practical situation in Southeast Asia. This high sea limitation renders international obligations to combat piracy unenforceable once the pirates have moved into the jurisdiction of any coastal state. Similarly, the said international obligation does not expressly compel any country to crack down on suspected pirates who move within the territorial waters of the country. The limitation also opens a back door for countries to shy away from any blame laid against them for their inefficiency in controlling piracy such as in their territorial seas or areas subject to disputed jurisdiction. None of the attacks against ships in the Straits of Malacca and Singapore were piracy because they did not take place on the high seas or in the EEZ. Almost all attacks on moving ships in Southeast Asia are against ships exercising rights of passage in the territorial sea or in archipelagic waters. Very few of the incidents in Southeast Asia are "piracy" as defined in international law because they took place in waters under the sovereignty of a coastal state. Such attacks are offenses under the laws of the costal states. Therefore, the law of piracy in the UNCLOS appears to be a weak tool for preventing and suppressing attacks on ships in Southeast Asia.

If the same piratical act takes place in the internal or territorial waters of a state, such an act is not covered by the international rules on piracy. As a matter of fact it may not even be considered an act of piracy, under the coastal state domestic law. Therefore, foreign warships are barred, under the UNCLOS piracy provisions, from exercising jurisdiction over a ship aboard which this act is being committed. As absurd as it is, under the existing piracy regime, the offended state or any other state, whatever the gravity of the offence might be, even if it involves killing of crew members or passengers, has no legal means, because it lacks jurisdiction, to take police action against a foreign-flagged ship and secure prosecution and adequate punishment of offenders, unless authorized by the coastal state in whose territorial waters the act is taking place (Jesus, 2003). In this case, whether or not and to what extent such an act will be prosecuted or not totally depends on the coastal state or the flag state, as the “extradite or prosecute” clause does not apply to piracy (Jesus, 2003).

Besides the problem of definition, lack of effective law enforcement is another severe problem in anti-piracy in Southeast Asian seas. First of all, after the entry into force of UNCLOS, the water areas under national jurisdiction have been greatly expanded. Such expansion gives coastal States additional sovereignty or sovereign rights over their respective jurisdictional waters, but on the other hand, it also brings difficulties in enforcement within these areas, particularly in regard to piracy. It even poses a big problem for some small countries which own a vast size of water areas but are lack an effective enforcement mechanism (e.g. Archipelagic waters of Indonesia). As is pointed out, pirate attacks often occur in areas where the law enforcement response is either non-existent or negligible (Ellen, 1986). In addition, UNCLOS created the EEZ and Continental Shelf regimes. Accordingly, the high seas are shrunk upon the expansion of territorial seas and EEZs, and the free mobility area in the high seas to control piracy is getting smaller. The question then arises whether the patrol vessels can freely enter into the EEZ areas of other States. Although the provisions in UNCLOS regarding piracy are applicable in the EEZs (Article 58 (2) of UNCLOS), the coastal States may not be very happy to see warships or governmental vessels of other countries to pursue and arrest piracy vessels in their EEZs where they have sovereign rights and jurisdiction. The above zoning provisions of the UNCLOS may thus complicate the enforcement of the law of piracy (Zou, 2005).

Secondly, unresolved maritime delimitation among Southeast Asian seas, including the multi-overlapping claims of maritime jurisdiction in the South China Sea makes the work on anti-piracy even more complex. Perhaps the main obstacle to the regional cooperation is the overlapping territorial claims for the islands in the South China Sea. The Spratly Islands are claimed by five adjacent countries, i.e., Brunei, China including Taiwan, Malaysia, the Philippines, and Vietnam, and the territorial dispute has not yet been resolved. Even if the territorial dispute had been solved, there are still boundary delimitation issues in the South China Sea. As is pointed out, disputes over maritime boundaries make accurate delineation of enforcement responsibility difficult, if not impossible (Clingan, 1989). In addition, effective law enforcement is difficult in the South China Sea, because of its vastness and due to the fact that it is dotted with numerous uninhabited islands to which pirates can easily retreat (IMO report, June 17, 1998).

The third problem is on hot pursuit. Luft and Korin (2004) argues that navies of foreign countries are normally forbidden to chase pirates across national boundaries, in what is known as the “right of hot pursuit”. This is of particular concern in areas such as the Strait of Malacca, where pirates often rapidly escape from one country’s territorial waters to another’s, leaving frustrated security forces in their wake. The view by Brittin (1989) may be insightful. He once said that "if a pirate is chased on the open sea and flees into the territorial maritime belt, the pursuers may follow, attack and arrest the pirate there; but they must give him up to the authorities of the littoral state" (1989) In other words, foreign warships have the right of hot pursuit within the EEZ of a coastal State and the right to arrest the piratical vessel there, but the coastal State may have the right to request the State which have exercised the rights in respect of suppression of piracy to hand over the pirates for trial in the coastal State, if that State intends to do so.

To conclude at this point, as Judge Jesus (2003) holds, the piracy regime contained in the UNCLOS only deals with the “powers, rights and duties of the different states *inter se*, leaving to each state the decision how and how far through its own law it will exercise its own powers and rights” (Jesus, 2003) It does not thus impose on the state any obligation to prosecute and punish the offenders and dispose of the properties (Jesus, 2003).

2.2. Convention on the Suppression of Unlawful Acts against the Safety of Maritime Navigation

When UNCLOS was drafted more than two decades ago, the most important criminal activities at sea included piracy, armed robbery against ships, narcotic drugs and illegal dumping and discharge of pollutants. Since 1982, and especially after the 9.11 attacks, the importance of other crimes at sea, like terrorism or transportation of WMD, rose dramatically. It was felt that in order to deal effectively with future cases of maritime terrorism from a judicial point of view; a specific international regulation was needed to secure the prosecution and punishment of the offenders, since the piracy laws seemed to be inadequate to that end.

The first international legal instrument on a specific legal regime covering sea terrorist acts, though without mentioning terrorism, came about only in 1988 with adoption of the IMO Convention on the Suppression of Unlawful Acts against the Safety of Maritime Navigation (SUA). The main goal of the SUA is to punish any person who commits an offense by unlawfully and intentionally seizing or exercising control over a ship by force or threat thereof; or performing an act of violence against a person on board a ship if that act is likely to endanger the safe navigation of that ship; or destroying a ship or causing damage to a ship or to its cargo which is likely to endanger the safe navigation of that ship (McDorman, 2005). It covers the unlawful acts no matter whether they are for political ends or for private ends.

The fundamental purpose of SUA Convention, along the lines of other anti-terrorist Conventions from which it drew inspiration, is the adoption of the *“extradite or prosecute”* clause, imposing an international obligation on all states parties in which the offenders may be present to either prosecute them in their own court system, whether or not the offence was committed in their territory, or to extradite the offenders to one of that states that has jurisdiction under the Convention (McDorman, 2005). Indeed, the SUA Convention only establishes a mechanism to secure the punishment through judicial means of those involved in maritime terrorism, by imposing a legal obligation on a state party to activate extradition of the offender if he is present in the state's territory or, if failing to do so for whatever reason, to prosecute him in the state's own court system.

The 1988 Protocol for the Suppression of Unlawful Acts against the Safety of Fixed Platforms Located on the Continental Shelf (the SUA Protocol) contains similar provisions. Both SUA Convention and Protocol can be regarded as complimentary as anti-piracy legal measures. However, the scope of the territorial application between the UNCLOS and the SUA Convention is different: while the former applies only to the high seas and the EEZ, the latter applies not only the waters beyond, but also waters within national jurisdiction. This enlarged territorial scope responds to the need to combat maritime terrorism in all areas of the ocean.

Beckman (2002) claims the SUA Convention could be an important tool for combating major criminal hijacks, the most serious type of attacks against ships in Southeast Asia. The convention would apply to such attacks whether they were committed in port, in the territorial sea, or in maritime zones outside the jurisdiction of the coastal states (Backman, 2002). If all the states in Southeast Asia were parties to the convention, persons who committed offences under the convention in Southeast Asian waters would become "international criminals". If they entered the territory of any state party to the convention, that state would be under a legal obligation to take them into custody and either prosecute them themselves or extradite them to another state for the purpose of prosecution. By making such persons "international criminals" among states parties to the convention, it would help ensure that offenders had nowhere to hide.

However, though the United States and other maritime powers are pressing other countries to ratify the 1988 SUA Convention, as of 31 January 2001, only 52 states were parties to the SUA Convention. The states parties from Asia include China, Japan, India, Pakistan, and Sri Lanka (Young and Valencia, 2003).Even though Southeast Asia is one of the regions with the highest incidents of piracy and armed robbery against ships, no ASEAN countries are parties to the SUA Convention.

Some Southeast Asian nations fear that the obligations under the SUA Convention could compromise their national sovereignty ant that the Convention could eventually be expanded to even allow maritime forces of other nations to pursue terrorists, pirates, and maritime criminals in general into their waters. In addition to that, some Southeast Asian states feel the Rome Convention only makes sense for those countries with already established maritime dominance or unchallenged maritime boundaries. For countries with a recent colonial history and relatively newly won independence, as well as disputed or porous maritime boundaries, the SUA Convention could be a serious compromise to both their national pride and domestic support for their government. However, if piracy and terrorism are fused into a general threat to maritime security, developing countries may find outside 'help' easier to accept and to 'sell' to their domestic polity. So it may be in the "United States' interest to conflate piracy and terrorism to persuade reluctant developing countries to assist maritime powers to pursue pirates and terrorist in their territorial and archipelagic waters" (Young and Valencia, 2003).

Besides the lack of membership as analyzed above, the limitation of the SUA Convention lies on its lack of preventive approach.(Jesus, 2003) As Jesus (2003) argues in order to effectively prevent acts of sea terrorism from happening and address terrorist attack against ships and other targets, states should be able to enjoy not only a judicial jurisdiction over offenders by claiming that they be prosecuted or by prosecuting themselves, but also a police jurisdiction that will allow them to prevent and stop terrorist ships from making terrorist attacks against other ships or against other targets such as port and pipeline facilities, platform structures, or that may be directed at blocking traits used for international navigation or causing major marine environment damage. Valencia holds that the SUA Convention may not be the appropriate instrument to combat piracy (Young and Valencia, 2003). The enumeration of offences under its Article three, even if interpreted broadly, will clearly cove only the serious but admittedly less common incidents of vessel hijackings and not the most common forms of piracy and armed robbery at sea, specifically in the Southeast Asian region. Thus more than 95 percent of the piracy and sea robbery incidents reported thus far would not be covered by its application. There is indeed a need for standardized international law that will facilitate the prevention and prosecution of piracy, but the SUA Convention may not be it (Young and Valencia, 2003).

2.3. The 2005 SUA Amendments

The Amendments to the Convention for the Suppression of Unlawful Acts Against the Safety of Maritime Navigation 1988 (2005 SUA Amendments) was adopted in October 2005. It substantially expands the scope of the original 1988 SUA Convention. According to McDorman (2005), for instance, the 1988 SUA Convention does not touch upon the possibility of a State Party boarding a vessel of another State Party to prevent a SUA offence, while article 8bis of 2005 SUA Amendments has created the possibility for the boarding of Sate Party vessel on the high sea. The 2005 SUA Amendments considerably broadens the covered offences beyond concerns of vessels and navigation safety to deal directly with: 1) the use of a vessel as an instrument of or platform for terrorist activity; 2) the transport of certain suspect materials or biological, chemical or nuclear weapons; and 3) the transport of a person who has committed an act that is an offence under any of nine terrorist conventions.

However, the breadth and nature of the new SUA offences may have made some States nervous about supporting a robust boarding regime (McDorman, 2005). First, the transport of "any equipment, materials or software or related technology that significantly contributes to the design, manufacturing or delivery" of a biological, chemical or nuclear weapon is impressively broad and could bring almost any commercial vessel under suspicion. Second, the international political climate is currently one of skepticism respecting the true motives of States seeking developments in the international law and practice for dealing with terrorist activity. Third, the IMO, as the global shipping arena which are composed of member States officials from transport departments and those engaged directly with shipping matters, tend to be supportive of flag State interest. The 2005 SUA Amendments do not fit the normal area of IMO expertise since the Amendments are primarily about criminal law, the exception being the boarding provisions. Fourth, the real accomplishment of the SUA Amendments is the expansion of the offences. Adopting a robust barding regime may have been seen as having the consequence of discouraging States from becoming a party to the Amendments and, thus, the boarding regime may have been "weaken" to avoid this possibility.

Hence, the 2005 SUA Amendments contains a very conservative approach to flag State consent. In the situation of a State Party having reasonable grounds to believe a vessel (or someone on board) is, has or may commit a SUA offence, it is necessary to have direct consent from the flat State to board, inspect or take other actions. More generally, the issue of the consent to boarding in the 2005 SUA Amendments has about it "an air of unreality." (McDorman, 2005) First, while the precise wording of the Amendments leaves consent to a request to board and inspect (and take other action respecting the vessel, cargo and persons on board) in the hands of the flag State, the realities are such that when confronted with a request to board and inspect a suspect vessel, unless a flag State is in a position to take its own direct action, it is highly unlikely that a flag State would withhold authorization since such a withholding would be tantamount to an admission of complicity in the activities of the vessel. Second, it is highly unlikely that "states of proliferation concern" (the wording from the PSI Interdiction Principles) are going to become a party to the 2005 SUA Amendments and its boarding regime only applies to those States that become a party to the SUA Amendments.

2.4. The PSI Interdiction Principles

The frustrating experience with the North Korean vessel in November 2002 (Syrigos, 2006), led US President George W. Bush to announce on 31 May 2003 in Krakow, Poland, a new initiative against shipments of WMD and missile-related equipment in transit via air, land, and sea named as 'Proliferation Security Initiative' (PSI), which was described by US as "an activity, not an organization" with the goal being enhanced cooperation amongst participating States respecting the existing framework of nonproliferation and control of weapons of mass destruction. The United States claims that more than sixty States have indicated their support for the PSI (Garvey, 2005). The PSI Interdiction principles were released in September 2003 following several meetings of the then PSI participants (Australia, France, Germany, Italy, Japan, the Netherlands, Poland, Portugal, Spain, the United Kingdom and the United States) that refined the wording and content (McDorman, 2005). The PSI Interdiction Principles are directed at "States or non-state actors of proliferation concern" which are engaged in proliferation through (1) efforts to develop or acquire chemical, biological, or nuclear weapons and associated delivery systems; or (2) transfers (selling, receiving, or facilitating) of WMD, their delivery systems, or related materials.

The PSI Interdiction Principles do not create legal obligations or responsibilities and do not, on their own, create or provide an international legal justification for the boarding of foreign flag vessels or any other interference action. There is no formal process by which States indicate they are willing to accept, apply or act on the PSI Interdiction Principles. Rather, the Principles indicate that the PSI participants (in September 2003) are committed to the Principles and call o other States to make a similar commitment to take "specific actions", including to board and search their own flagged vessels that are suspect; to stop and search suspect vessels "in their internal waters, territorial seas, or contiguous zones" and seize cargo; and to enforce condition on vessels "entering or leaving their ports, internal waters or territorial seas" such as that suspect vessels will "be subject to boarding, search, and seizure of…cargoes priory to entry".

These specific actions generally accord with the existing international legal framework respecting boarding and inspection since a flag State has authority over its own vessel and the "geographic exception" covers the listed interferences with foreign flag vessels. However, the action of a coastal State to stop and search suspect commercial vessels engaged in passage through territorial waters or to enforce conditions on the entry of a vessel into a State's territorial seas where the vessel is not heading to a port or internal waters may be questioned since in either case there may e an interference with the vessel's innocent passage rights. In other words, these specific actions may not be included within the "geographic exceptions".

Syrigos (2006) also questions the whether PSI was consistent with international law. The founders of the PSI claimed that their initiative was based on the Statement made by the President of the Security Council on 31 January 1992, at the conclusion of the 3046th meeting of the Security Council, in connection with the item entitled: "The responsibility of the Security Council in the maintenance of international peace and security"[1] Nevertheless this statement was quite vague as to the exact actions that could be undertaken for its implementation. Furthermore, Presidential Statements do not enjoy the same status as Security Council Resolutions. Thus the statement could not legitimize actions of pre-emptive interdiction of vessels on the high seas on the ground of shipment of WMD.

2.5. The U.S. Ship Boarding Agreement

The limitation to interdict ships on the high seas led the States parties to the PSI to rely on bilateral agreements with flag States. The idea was to stop and search ships (identified as being involved in WMD trading) on the high seas after getting permission from the country that issued the ship's flag and not from the master of the ship (Syrigos, 2006).

Commencing in February 2004, the United States has entered into six agreements referred to as "Proliferation Security Initiative Ship Boarding Agreement" with Liberia, Panama, the Marshall Islands, Croatia, Cyprus, and Belize (McDorman, 2005).All of the Ship Boarding Agreements respect flag State consent regarding boarding of foreign flagged vessels in international waters ("all parts of the sea not included in the territorial sea, internal waters and archipelagic waters of a State") by providing that if one Party encounters "a suspect vessel" claiming nationality in the other Party or flying the flag of the other Party, that Party may request the flag State to consent to a boarding and

[1] According to the statement: "The proliferation of all weapons of mass destruction constitutes a threat to international peace and security. The members of the Council commit themselves to working to prevent the spread of technology related to the research for or production of such weapons and to take appropriate action to that end"

inspection of the vessel. None of the six Ship Boarding Agreements provide an automatic consent by the flag State for any boarding, inspection or other action respecting the vessel, cargo or persons on board. All of the agreements provide a wide range of options respecting the responses the flag State can give to a request. In particular, in addition to boarding and inspection, flag State authorization can be given to detain the vessel, cargo or persons on board where evidence of "proliferation" is found pending further instructions.

Four of the six Ship Boarding Agreements provide that, following a flag State's acknowledgement of receipt of a request for consent to board and inspect, the flag State's not responding will be read as "The requesting Party will be deemed to have been authorized to board the suspect vessel for the purposes of inspecting the vessel's documents, questioning the person's on board, and searching the vessel to determine if it is engaged in proliferation at sea." (McDorman, 2005).

All of the Ship Boarding Agreements make it clear that "the primary right to exercise jurisdiction over a detained vessel, cargo or other items and persons on board (including seizure, forfeiture, arrest and prosecution)" rests with the flag State. Five of the agreements note that the flag State may "waive" primary jurisdiction and authorize the boarding State to apply its law against the vessel, cargo and person on board.

We may note that even in the face of maritime terrorism, the recent developments-the 2005 SUA Amendments, the PSI Interdiction Principles and the U.S. Ship Boarding Agreements-all respect the "traditional" international legal framework of requiring direct flag State consent before boarding and inspection of a foreign flagged vessel can be undertaken on the high sea. They make some progress in terms of measures to counter maritime terrorism regarding boarding of vessels of flag states, but the problem of consent from flag states still remains a problem.

3. Concluding Remarks

The frequent Somali pirates have instigated a global discussion on how to confront and tackle this problem militarily, politically and legally. The military and political measures are well know and obvious, but the legal ones are not, because there is a certain amount of uncertainty and vagueness. This paper tries to answer the question: what can international law do in excising its role to prevent, control and punish piracy and maritime terrorism? The following points can be concluded from the above discussion. 1. to promote the emergence of agreed rules to combat crime at sea and from the sea, applicable not only to the high seas and to the EEZ, but to all the ocean spaces, including territorial waters, based on a system of coordination and Conventional consent of coastal states, without jeopardizing their sovereignty over such waters, in other words, to extend the international regulations to cover situations that are not covered by the existing international rules; 2. Clarify certain ambiguities in the definition of piracy that have led to different approaches and interpretations by different states and practitioners; 3. Devise a cooperative mechanism to secure effective enforcement of the regulations, assisting the coastal state building up capacity; 4. States in the region should review their domestic legislation on piracy to make certain that they would have jurisdiction over piracy should their navel forces arrest and seize the pirates. 5. Efforts should be made to persuade all governments in the region to become parties to the SUA Convention and to implement its provisions in their national laws. Efforts should be made by governments in the region to harmonize their laws implementing the convention.

References

Beckman, R.C. (2002), Combating Piracy and Armed Robbery Against Ships in Southeast Asia: The Way Forward", Ocean Development & International Law, Vol. 33, No. 3-4, pp 317 – 341.

Brittin, B.H. (1989), The Law of Piracy: Does It Meet the Present and Potential Challenges? in Ellen (ed.), *Piracy at Sea* (Paris: ICC Publishing SA).

Clingan, T.A.Jr (1989), The Law of Piracy, in Eric Ellen (ed.), *Piracy at Sea* (Paris: ICC Publishing SA).

Ellen, E.F. (1986), Piracy, in Eric F. Ellen (ed.), Violence at Sea (Paris: ICC Publishing S.A.).

Goldie, L.F.E. (1988), Terrorism, Piracy and the Nyon Agreement, in Yoram Dinstein (ed.), International Law at a Time of Perplexity: Essays in Honor of Shabtai Rosenne (Dordrecht: Martinus Nijhoff, 1988).

Jesus, J.L. (2003), Protection of Foreign Ships against Piracy and Terrorism at Sea: Legal Aspects", The International Journal of Marine and Coastal Law, Vol. 18, No. 3, pp 363-400.

IMO report, June 17, 1998.

Luft, G. and Korin, A. (2004), Terrorism Goes to Sea, Foreign Affairs, pp 61-71.

MacCartnery, Jane and Simon Cameron-Moore, "U.S. to Freeze 'Terror' Funds in SE Asia-Sources", Reuters, 13 march 2003.

Syrigos, A.M. (2006), "Developments on the Interdiction of Vessels on the High Sea", in Anastasia Strati, Maria Gavouneli and Nikolaos Skourtos (ed), *Unresolved Issues and New Challenges to the Law of the Sea: Time Before and Time After* (Leiden/Boston: Martinus Nijhoff Publishers, 2006), pp 149-202.

McDorman, Ted L, "Maritime Terrorism and the International Law of Boarding of Vessels at Sea: Addressing the New Development", paper presented at the Conference on Maritime Security in the South China Sea, December 2005, China.

Young, A.J. and Valencia, M.J. (2003) Conflation of Piracy and Terrorism in Southeast Asia: Rectitude and Utility, Contemporary Southeast Asia, Vol. 25, pp 252-269.

Zou, K. (2005), "Crackdown on Piracy in Southeast Asian Seas: Need a More Effective Legal Regime?" paper presented at the Conference on Maritime Security in the South China Sea, December 2005, China.

A port capacity study: planning models and evaluation

Dušanka Jovović, *Branislav Dragović*, *Vesna Dragović-Radinović* and *Maja Škurić*

Maritime Faculty, University of Montenegro, Dobrota 36, 85330 Kotor, Montenegro.
E-mail: bdragovic@cg.yu (D Jovović)
E-mail: branod@cg.ac.yu (B Dragović)
E-mail: majask@t-com.me (V Dragović-Radinović)
E-mail dusankaj@cg.yu (M Škurić)

Abstract

This study addressed the optimal capacity calculation problem for Port of Bar (Montenegro). Unlike previous studies, this investigation is based on a detailed statistical model of handling activities in different years. We proposed models for determining the optimal port performance indicators. The objective of models was to maximize throughput per year and in the determination of maximum level of port utilisation factor. Finally, as it can be concluded from the output results, that existing capacity of Port of Bar may be significantly increased.

Ports must be planned and developed to provide an optimum port capacity for increasing volume of trade. Port capacity depends upon many parameters such as the type of cargo, the randomness of ship arrivals and sizes, the percentage of congestion in port, queue discipline, the degree of berth occupancy, the service time discipline, the optimal determination number of berths and cranes in port, the optimum cost combination, the mutual quay cranes interference exponent, the optimal combination of berths/terminal and quay cranes/berth, etc. Therefore, a new solution approach with a tactical focus should be developed in relation of port capacity. This approach should already have the basis for Port of Bar regarding aspects of capacity.

Keywords: Port capacity, Port investment model, Maximal Practical Capacity, Sustainable Practical Capacity

1. Introduction

The paper has analysed in detail this topic both through theoretical and practical studies, highlighting the effects of different forms of port capacity for Port of Bar. Looking at Figures and Table, it is clear that real and potential capacity only modelled the cargo flow and capacity constraints, as in the generic statistical analysis. In addition, other restrictions are not taken into account. It may be interesting to research the integration of more practical restrictions of capacity and, consequentially, a new solution algorithm for this problem.

Planning of port capacity should address the following six questions: 1) what is the expected demand for services in terms of types and volumes of the transport flows, 2) what is the required supply of capacity in terms of physical characteristics (sizes and numbers) and service characteristics (tariffs and productivities), 3) what is the utilization rate and equilibrium demand, 4) what are investment cost and service price, 5) what are the economic benefits, and 6) what is the overall viability of the port investment project.

The review shows that good solution approaches for real and potential capacity for Port of Bar are already developed, but nevertheless there are still many open questions. Therefore, a new solution approach with a tactical focus should be developed in relation of port capacity. This approach should already have the basis for Port of Bar regarding aspects of capacity.

An overview of the overall port operation can be seen in many Figures and Table illustrated below. This growth in cargo volumes is expected to continue and will require additional capacity on the freight transportation network and through ports in particular. Some of this additional capacity may be acquired through the increased use of new operation technologies. In contrast to terminal expansion and information technology deployments, new operation technologies are a low cost method to increase capacity. Although, new operation technologies on its own may not solve the capacity problem, it can be more quickly implemented than other solutions, and can be used to complement other strategies.

Most studies and papers focus their attention on a port capacity study and the optimum port capacity. The determination of optimum number and capacity of berths, and the optimum port capacity has been treated both theoretically and practically in many studies. All these studies refer to solve the problems, for example, in port planning, development, investment and design Gos (1967), Frankel (1979 and 1987), Agerschou et al. (2004), Dekker (2005), in port modeling Frankel and Chang (1977), Dragović (2000), Dragović et al. (2006) and optimization of the port processes and subsystems Wanhill (1974), Weille and Ray (1974), Noritake and Kimura (1983 and 1990), TranSystems Corporation – TSC (2003), Fonteijin et al. (2006) and Talley (2006). Mainly, mentioned papers present only few examples which have been conducted to the determination of the optimal capacity of port systems in different environment within various points of view and in heterogeneous cases.

The rest of this paper is organized as follows. In Section 2 we give port investment model with port investment strategy. Section 3 shows capacity model development related to annual throughput volumes achieved for Port of Bar. In addition, analysis of maximal practical capacity and sustainable practical capacity is given. Section 4 presents projected cargo throughput volumes for the Port of Bar in coming years. The final Section provides concluding remarks.

2. Port Investment Model

In view of the fact that an investment model must be built in relation to transportation demand and supply, DeNeufville (1973) developed a model for evaluating modal competition. It is noticed in (Frankel, 1979 and 1987) that this approach can be used to develop an analogous model for port investment that can be used to provide overall guidance for planning optimal strategies for investment in port, consisting of three subsystems. A typical strategic question is designed to answer is, should more specialized berths, such as container terminals rather than general cargo berths be built?

The accent of this model is on supply. We are indicating about how costs increase as the port traffic in the port increases. We calculate the cost of offering total throughput per year. To assume this latter cost we would have to make assumptions on how demand responds to other levels of service besides total cost such as waiting time of ships. It is also assumed that no port congestion occurs, and hence that direct cost of offering a unit cargo is independent of the actual total volume handled.
This model explicitly uses the following demand factors:

g - the growth rate of the supply of the subsystem,
V_0 - the initial volume of expected port demand at any subsystems.

The following result is given in (Frankel, 1987); also see (Meyer, 1970). We give here a complete proof of below expression.

Proposition. The total present cost P_j of port supply capacity of subsystem *j* over *t* years is given as

$$P_j = I_j + k_j V' + \frac{a_j}{r}(1 - e^{-rT}), \tag{1}$$

or

$$P_j = I_j + k_j \frac{V_0}{r-g}(1 - e^{-(r-g)T}) + \frac{a_j}{r}(1 - e^{-rT}) \quad (2)$$

where
I_j - *investment*,
k_j - *cost per unit* ($j = 1,2,3$),
$V^{'}$ - *discounted volume*,
a_j - *annual overhead*,
r - *instantaneous rate of return of capital and*
T - *planning period that is the same for every three considered subsystems,*

all associated with the *j*-th type of subsystem in a port (dry-bulk terminal; liquid-bulk terminal; and general cargo berths, respectively).

Proof. The above expression is given in (Frankel, 1987); also see (Meyer, 1970), where the second term of the above expression is given in another form. As noticed in (Frankel, 1987) the investment costs I_j for a container terminal are expensive because of the large scale cargo-handling equipment and specialized berths requirements. The operating costs, though, are usually small because of the high degree of mechanization. In general, cargo handling of the situation is quite different. Here, we deduce the expressions for the second and third term from Eq. 1.

$$dV^{'} = V_0 e^{-(r-g)t} dt, \quad (3)$$

whence, after integrating the above formula from $t = 0$ to T, we obtain

$$V^{'} = \frac{V_0}{r-g}(1 - e^{-(r-g)T}) . \quad (4)$$

Similarly, the third term in Eq. 1 expresses overhead costs, and hence

$$a_j dt = e^{-rt} \quad (5)$$

whence, after integrating the above formula from $t = 0$ to T, we obtain that it is equal to

$$\frac{a_j}{r}(1 - e^{-rT}) . \quad (6)$$

This completes the proof of Eq. 1. Substituting the expression for V' given by Eq. 3 immediately follows Eq. 2.

2.1. Investment strategy

For a fixed planning period *T*, from Eq. 1, we see that each P_j, $j = 1,2,3$, may be considered as linear function on V'; namely, if we set $b_j = I_j + \frac{a_j}{r}(1 - e^{-rT})$, $j = 1,2,3$, then Eq. 1 can be written as

$$P_j = k_j V' + b_j\,,\ j = 1,2,3 \tag{7}$$

3. Capacity Model Development

The capacity model development process began with the study establishing a detailed understanding of the Port of Bar facilities current operations. By using the break-bulk, dry-bulk and liquid-bulk trend functions developed here, we analyzed the current terminal operations and its current cargo handling capacity. The information contained in this study is designed to greater basic Port of Bar operating parameters and procedures.

This information was then supplemented its interviews of port staff as well as terminal operators and shippers. For the current analysis, all collected information obtained via the questionnaires and interviews was reviewed and confirmed with the appropriate operating staff.

From the above, it is easy to see that capacity of the Port of Bar was based on the actual cargo types. While an overall increase in the percentage of different cargoes handled by the Port of Bar in previous years could increase the overall total tonnage, it would follow that the Port of Bar now has more available capacity. The capacity model would need to be adjusted to the cargo split adequately determine the utilization of the available capacity of Port of Bar.

The models provide output in the form of the *Maximal Practical Capacity* (MPC) for each terminal. This MPC refers to estimated annual throughput volumes that represent the high end of a realistic operating scenario. In practice, operating at a level equivalent to the MPC for any significant period of time is typically considered impractical and uneconomical. For practical purposes, the throughput capacity of a Port of Bar is more reasonably approximated at 80% of the port's MPC. This fact, referred to as *Sustainable Practical Capacity* (SPC), is the practical throughput capacity; a facility can reasonably be expected to operate at over a sustained period of time. For planning purpose, a reasonable approach is to develop facilities to meet SPC.

Consequently, sea ports, as a general rule, do not operate at the MPC level most of the time. Therefore, the SPC level is used for planning and evaluation task.

With this in mind, the capacity model results for the Port of Bar existing facilities presented in this study are the SPC throughput capacity volumes, equaling 80% of the MPC.

3.1. Achieved Annual Throughput Volumes for Port of Bar

It is clear evident that Figures 1 and 2 show total throughput of different types of cargo from 1983 to 1992 and from 1993 to 2008 in Port of Bar. These results imply the following characteristics: for period from 1983 to 1992, the biggest throughput of general cargo was 860000 tons in 1991; also, for these ten years, the biggest throughput of bulk cargo was 1515000 tons in 1989; and, last, the biggest throughput of liquid cargo was 678000 tons in 1992 in Port of Bar (Milosevic, 1995; Škurić et al., 2008).

Considering the period from 1993 to 2008, the biggest throughput for general cargo was 1190000 tons in 2007, while the biggest throughput for bulk cargo was 1040000 tons in 2005 in Port of Bar. Finally, in viewing throughput of liquid cargo, the result is 759000 tons in 1998. In this context, the best year for all types of cargo was 1989, and reached 2744000 tons in Port of Bar (ARPAB, 2009; Milosevic, 1995; Škurić et al., 2008).

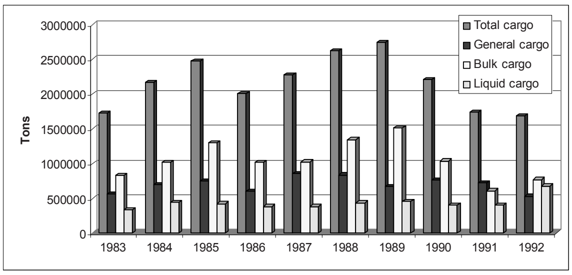


Figure 1: Annual throughput volumes for different cargo types from 1983 to 1992

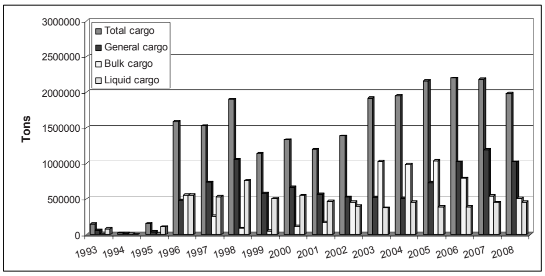


Figure 2: Annual throughput volumes for different cargo types from 1993 to 2008

In Figure 3 are presented the performances of throughput per different year in Port of Bar. This Figure shows the obtained results of throughput even if the capacity in Port of Bar was forecasted to be 5 million tons per year. As it can be shown, this capacity was never reached.

The main characteristics of this period are: total 26-year throughput was 44433900 tons; average throughput per year for each of 26 years was 1708996 tons; average throughput per month in this period was 142416 tons. On the other hand, it is obvious that the Port of Bar has to implement some adjustments to improve its environmental performance. A descriptive statistic approach shows that very interesting period from 1993 to 1995. During this period the Port of Bar was isolated system because UN established special sanctions as a consequence of destroyed former SFRJ.

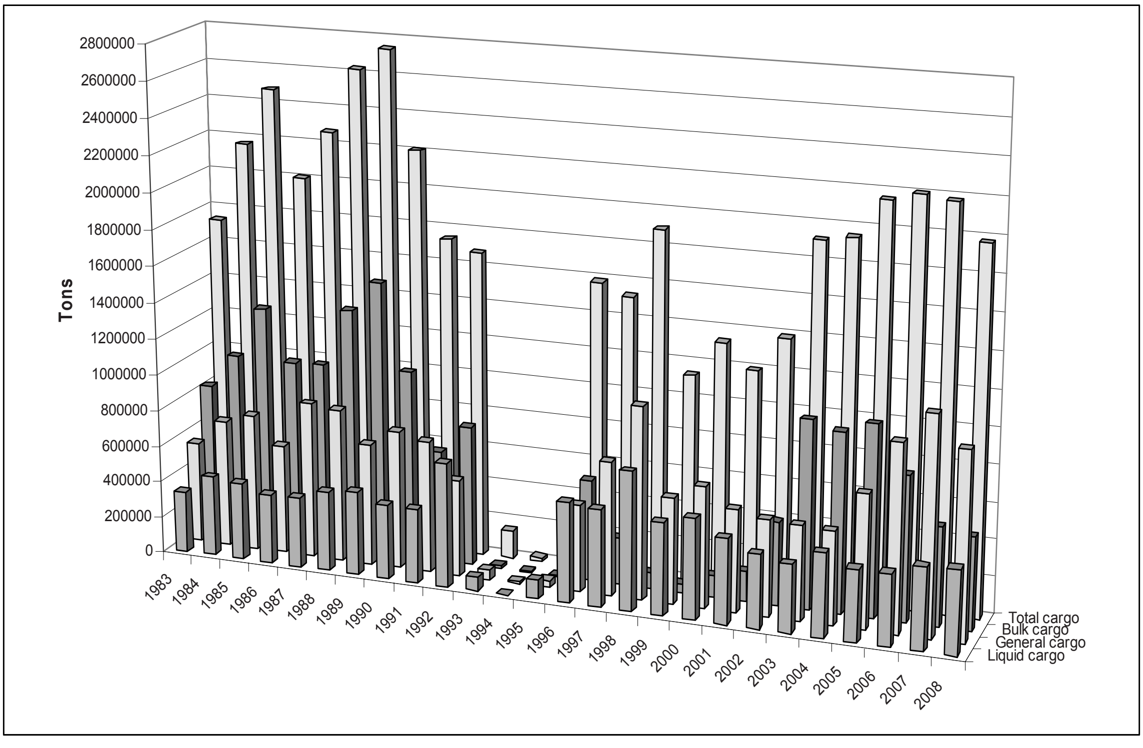


Figure 3: Trend function of annual throughput volumes for different cargo types from 1983 to 2008

When the status of 1989 in Port of Bar has been analysed, it suggests that the simmilar throughput was during all twelve months. Analysing the 1992, it can be noticed that just after July, it begins period of sanctions that last until 1995. Although, the throughput was drastically decreasing (see Figures 4 and 5, (Milosevic, 1995; Škurić et al., 2008)).

Last two years (2007 and 2008) are practically the same, but bigger throughput was in 2007. On the other hand, only general cargo reached bigger level in 2008, all the rest types of cargo, including total cargo was obtained in 2007. These trends are expecting to continue in future (see Figures 6 and 7).

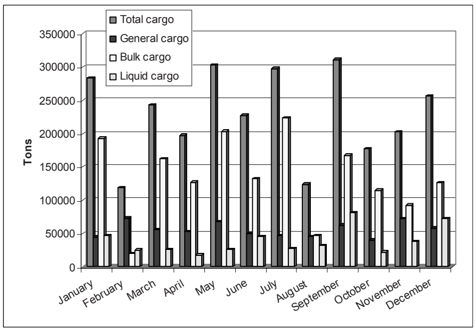


Figure 4: Monthly throughput volumes for different cargo types in 1989

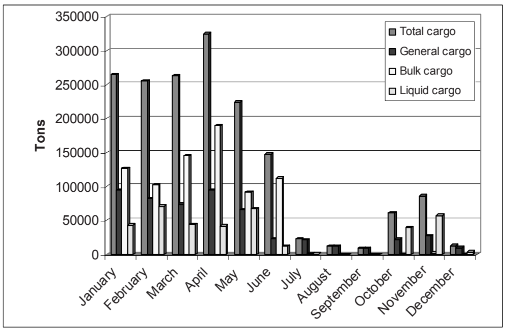


Figure 5: Monthly throughput volumes for different cargo types in 1992

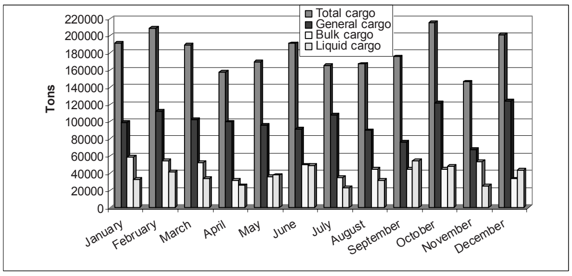


Figure 6: Monthly throughput volumes for different cargo types in 2007 (ARPAB, 2009)

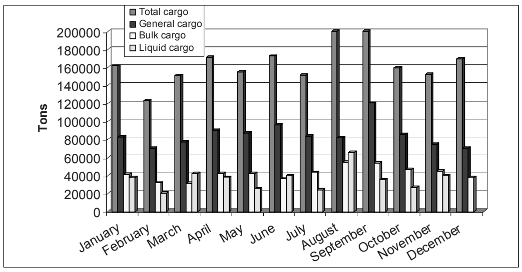


Figure 7: Monthly throughput volumes for different cargo types in 2008 (ARPAB, 2009)

3.2. Analysis of Maximal Practical Capacity and Sustainable Practical Capacity

In this Subsection we present the real possibilities of installed handling capacities and infrastructure objects of Port of Bar, i.e. Maximal Practical Capacity for the Port of Bar. It is now evident that the biggest port capacity from 1983 to 2008 has been in 1989. Total handling capacity was 2744000 tons. This capacity is expressed as C_{P1}.

Now, we consider the following factors: the biggest throughput of bulk cargo was 1615296 tons in 1989, 759000 tons (liquid cargo) in 1998 and 1190000 tons of general cargo in 2007, respectively. Supposing that all of these values are realized in one year, we have C_{P2}.

Therefore, the biggest throughput per month was 325003 tons in April, 1992. Similarly, assuming that the biggest throughput per month was realized every month in one year, we obtain C_{P3}.

Hence, we consider the following factors: the biggest throughput of bulk cargo was 224012 tons in July 1989, 81104 tons (liquid cargo) in September 1989 and 124100 tons of general cargo in December 2007, respectively. In this manner, the total handling capacity per twelve months would be equal C_{P4}.

In order to analyze realized capacities of Port of Bar, it is necessary to express theoretical capacity in the form (Agerschou et al., 2004; Frankel, 1979)

$$C_T = Q \cdot \frac{Laa}{L+k} \cdot \frac{d \cdot h}{u} \quad \text{tons per year,} \tag{8}$$

where: C_T - theoretical capacity in tons per year, Q - average ship size in dwt, L_{aa} - length of berths in meters, L - average length of ship in meters, k - safety space between ships (from 10 to 20 m), d - number of working days per year, h - daily working hours, u - average number of working hours. By substituting the corresponding values for Port of Bar in the above expression, we get theoretical capacity C_{P5}.

Then the real capacity C_{P6} can be expressed as $C_{P6} = \alpha \cdot \beta \cdot C_{P5}$ (where: α – efficiency coefficient of port handling capacity (0.5-0.9) and β – nonuniformity coefficient of port traffic flow (0.5-0.9).

Consequently, by using the above obtained results for C_{Pi} (i = 1, 2, ...,6), we can give in Table 1 proportions between these C_{Pi} (maximal practical capacity) and port handling capacity achieved per year from 1983 to 2008 (denoted in the first column as $C_{P'j}$). In particular, from Table 1 we have

$$\eta_{PT'83} = \frac{1}{6} \sum_{i=1}^{6} \frac{C_{P'83}}{C_{Pi}} = 39\ \%, \tag{9}$$

where $\eta_{PT'83}$ is the arithmetical mean of the proportions for $C_{P'83}$ in relation to C_{Pi} (i = 1, 2, ..., 6), respectively. Similarly, we obtain the corresponding values for $\eta_{PT'84}, \eta_{PT'85}, \ldots, \eta_{PT'07}$, respectively. Finally, the arithmetic mean of all these values including $\eta_{PT'83}$ is equal

$$\eta'_{PT} = \frac{39+48+55+45+51+50+61+49+39+37+3+5+4+36+34+42+25+29+27+31+43+44+48+49+48+44}{26} = 38\ \%. \tag{10}$$

In view of this, the value η'_{PT} = 38 % represents in fact the sustainable practical capacity for port of Bar in previous considering period of time.

Table 1: Maximal practical capacity (MPC) and Port handling capacity achieved (PHCA) for Port of Bar from 1983 to 2008

Port handling capacity achieved \ Maximal practical capacity in tons		C_{P1}	C_{P2}	C_{P3}	C_{P4}	C_{P5}	C_{P6}
		2744000	3564296	3900036	5150592	11728800	6568128
$C_{P'83}$	1731000	63 %	49 %	44 %	34 %	15 %	26 %
$C_{P'84}$	2166000	79 %	61 %	56 %	42 %	18 %	33 %
$C_{P'85}$	2479000	90 %	70 %	64 %	48 %	21 %	38 %
$C_{P'86}$	2005000	73 %	56 %	51 %	39 %	17 %	31 %
$C_{P'87}$	2276000	83 %	64 %	58 %	44 %	19 %	35 %
$C_{P'88}$	2265121	83 %	64 %	58 %	44 %	19 %	34 %
$C_{P'89}$	2744000	100 %	77 %	70 %	53 %	23 %	42 %
$C_{P'90}$	2212728	81 %	62 %	57 %	43 %	19 %	34 %
$C_{P'91}$	1743249	64 %	49 %	45 %	34 %	15 %	27 %
$C_{P'92}$	1683383	61 %	47 %	43 %	33 %	14 %	26 %
$C_{P'93}$	149457	5 %	4 %	4 %	3 %	1 %	3 %
$C_{P'94}$	220000	8 %	6 %	6 %	4 %	2 %	4 %
$C_{P'95}$	153000	6 %	4 %	4 %	3 %	1 %	3 %
$C_{P'96}$	1587000	58 %	45 %	41 %	31 %	14 %	24 %
$C_{P'97}$	1526000	56 %	43 %	39 %	30 %	13 %	23 %
$C_{P'98}$	1902000	69 %	53 %	49 %	37 %	16 %	29 %
$C_{P'99}$	1136000	41 %	32 %	29 %	22 %	10 %	17 %
$C_{P'00}$	1329000	48 %	37 %	34 %	26 %	11 %	20 %
$C_{P'01}$	1196000	44 %	34 %	31 %	23 %	10 %	18 %
$C_{P'02}$	1386000	51 %	39 %	36 %	27 %	12 %	21 %
$C_{P'03}$	1921000	70 %	54 %	49 %	37 %	16 %	29 %
$C_{P'04}$	1949000	71 %	55 %	50 %	38 %	17 %	30 %
$C_{P'05}$	2160000	79 %	61 %	55 %	42 %	18 %	33 %
$C_{P'06}$	2200000	80 %	62 %	56 %	43 %	19 %	33 %
$C_{P'07}$	2180000	79 %	61 %	56 %	42 %	19 %	33 %
$C_{P'08}$	1980600	72 %	56 %	51 %	38 %	17 %	30 %

4. Projected Cargo Throughput Volumes

Projected cargo throughput volumes identified in the previous Subsection and based on port investment model (Section 3) would be represent in Figure 8. The cargo volumes listed below include the volume handled, average maximal practical capacity, average port handling capacity achieved, average sustainable practical capacity I, average sustainable practical capacity II and sustainable practical capacity range. Recently, predominantly is handled break-bulk cargo (see Figure 7).

Consequently, the market assessment portion of our study investigated the possibility of containerized cargo as well as Ro-Ro units. Accordingly to previous experience with maritime opportunities, we can foresee the possibility of container and Ro-Ro markets developing, for instance, through the ferry and feeder Adriatic service supported by aggressive marketing techniques, it is within the scope of future analysis to do a market assessment to accurately define a projected forecast volume for such a cargo.

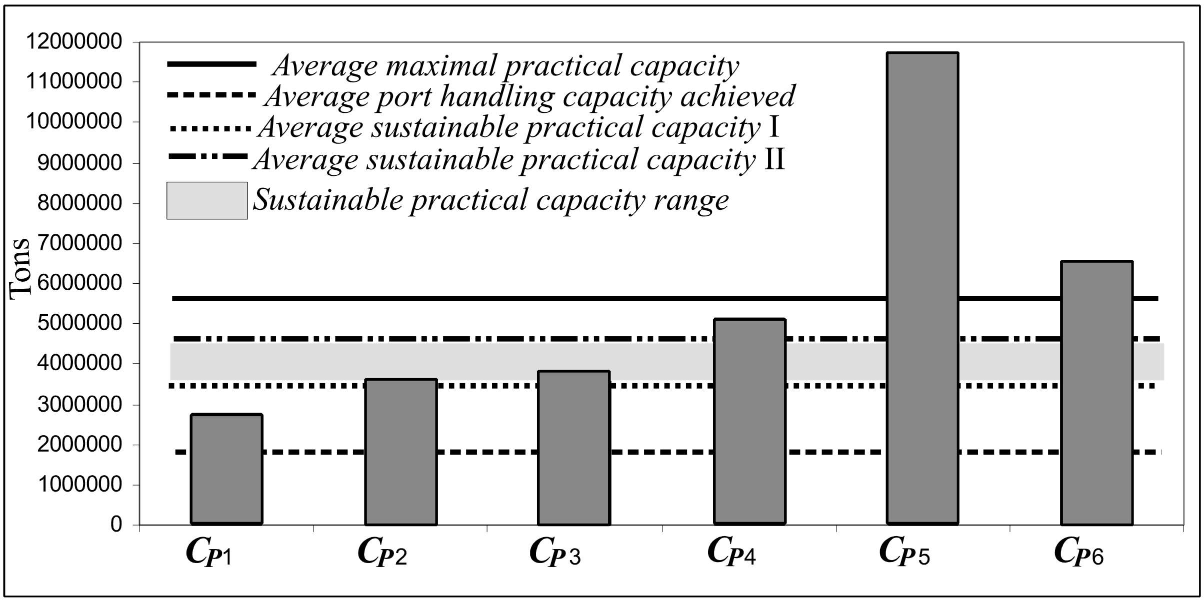


Figure 8: Average Sustainable Practical Capacity per year and Average Sustainable Practical Capacity for considered period of time (1983 – 2008)

Historically, the Port of Bar, has handled various cargo. The Port of Bar should have the ability to adopt future capacity demand from existing of future customers. The following Figure 9 outlines the projected growth of containerized cargo based on the above discussions and analysis. The annual growth rates for both standard containerized and Ro-Ro cargo take into consideration the cyclical nature of any commodity and use a percentage that averages out the highs and lows over the duration of the cycle. The following graph (Figure 9) compare the average sustainable practical capacity for containerized cargo, general cargo and bulk cargo as obtain from the model (Section 3) to the projected cargo throughput volumes to determine in what year the port of Bar could anticipate the capacity increase, assuming the specific components of each scenario from Figure 8.

From the viewpoint of planners, we should determine the system in which investment are to be made. The significant differences between these three, from the viewpoint of a planner, are that general and bulk cargo facilities have low capital and high operating costs where as container terminals have just the reverse. According to analysis of maximal practical capacity and sustainable practical capacity as well as project cargo throughput volumes we give model presented in Figure 9. This model based on Eqs. 1 and 2 (Section 3), and data collected from annual report of Port Authorities of Bar, 2009. Also, this model includes all specific results represented in Figure 8. However, investment for the containerized cargo subsystem should be considered in relation to results from Figure 9. The analysis, thus, implies that planners should consider development of container terminals as a priority.

Stated methodology can be used for fast and effective computations of the maximal practical capacity, sustainable practical capacity and average sustainable practical capacity in port systems. The application of numerical results in Figs. 8 and 9 is partially restricted because of the beforehand determined variables from projected cargo throughput volumes and others variables affecting these results.

However, presented methodology and results are convenient for different analyses, planning and development of port system, for example, increasing the number of berths or berths utilization depending on the optimum port capacity. The optimum size seaport and the optimum port capacity and associated costs could be extensively used in different analyses of investment model in real world.

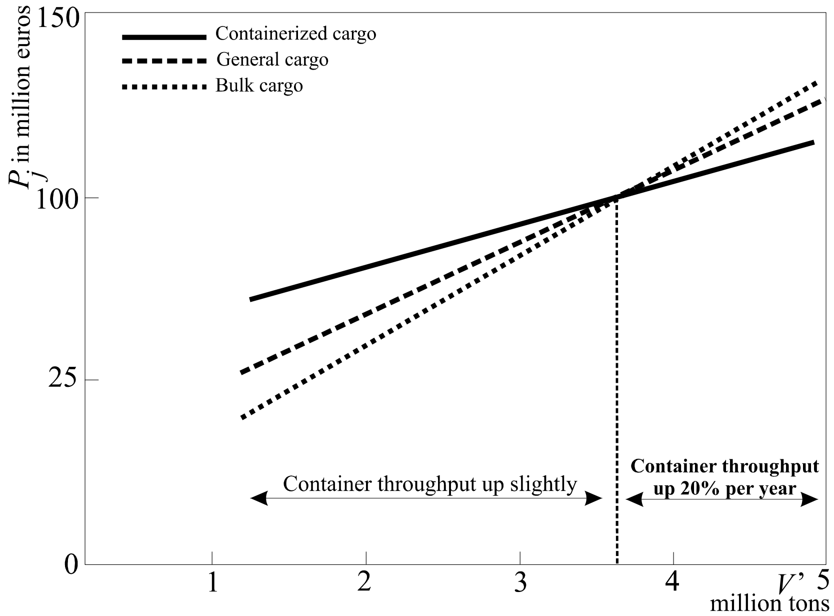


Figure 9: Minimum Cost Port

5. Conclusions

For the purposes of this evolution, it is assume, that in this case, container transport may be predominant. The current forecast projection for Port of Bar is based on the actual cargo anticipated for 2007 and 2008. The predominant change in the cargo volume forecast is from the installation of the container handling equipment at the container terminal property. The addition of the new equipment will increase the estimated containerized cargo for container terminal by 0.4 to 1.0 million ton per year. From previous year based off average volume of 2 million tons, we apply the same annual growth mid-point between the medium to high forecast projections scenarios present in Figure 8. Montenegro economy and infrastructure, it may be beneficial for the Port of Bar, to reassess the market forecast projection scenarios.

Economic impact study of the Port of Bar will be expressed as follows. Looking at all port operations, including the private sector in coming years, the study determines an overall direct, indirect and induced port industry impact. Furthermore, it is difficult to enumerate the impact facility like the Port of Bar provides when recruiting new industry to the land area. Finally, the Port of Bar could be considered a gateway to enhance model shift of cargo to water and should have the facilities available to assist local, regional and interregional planning and economic development efforts.

There are several ways in which the Port of Bar with the assistance of the Montenegro government and regional association and specialized planning and development agencies, may be able to increase capacity. Since most of this study requires a long duration to achieve final results, the Port of Bar and its regional partners (Serbia, Bosnia and Herzegovina, and Macedonia) should immediately begin to consider how different alternatives could be develop and implemented. Our study may be initial step to achieve a short term solution.

References

Agerschou, H., Dand, I., Ernst, T., Ghoos, H., Jensen O.J., Korsgaard, J., Land, J., McKay,S.T., Oumeraci, H., Peterson, J.B., Schmidt, L.R. and Svendsen, H.L. (2004), Planning and design of ports and marine terminals, Thomas Telford Ltd, 2nd Edition, London.
Annual Report of Port Authorities of Bar – ARPAB (2009), January 2009, Bar, Montenegro.

Dekker, S. (2005), Port Investment Towards an Integrated Planning of Port Capacity, The Netherlands TRAIL Research School, 2600 GA Delft, ISBN 90-5584-065-3.

DeNeufville, R. (1973), Role of Air Transportation in Sparsely Developed Areas, prepared for U.S., Dept. of Transportation; Office of the Assistant Secretary for Policy & International Affairs.

Dragović, B. (2000), Port and terminal modeling, Proceedings YU Science meeting – Water transport in 21st century, Faculty of transport, University of Belgrade, Belgrade, 421-426.

Dragović, B., Zrnić, Dj. and Radmilović, Z. (2006), Ports & container terminals modeling, Research monograph, Faculty of Transport and Traffic Engineering, University of Belgrade.

Gos, R.O. (1967), Towards an economic appraisal of port investments, Journal of Transport Economics and Policy, September: 249-272.

Frankel, G.E. (1979), Port Planning and Development, Center for Transportation Studies, Massachusetts Institute of Technology, Cambridge.

Frankel, G.E. (1987), Port Planning and Development, New York, John Wiley and Sons.

Frankel, G.E. and Chang, B.T. (1977), Port Capacity, Measures and Port Effectiveness Modeling, Center for Transportation Studies, Massachusetts Institute of Technology, Cambridge.

Fonteijin, H., Bishop, D. and Arvanitidis, M. (2006), Port capacity, Theoretical framework, Espo Statistics Committee, Goteborg 19th September 2006, Presentation based on meetings at the Port of Thessaloniki 24th & 25th August 2006.

Meyer, J.R. (1970), Techniques of Transportation Planning, Vol. 2, System Analysis and Simulation Models, BI, TRP, Washington.

Milosevic, S. (1995), Investigation of port capacity – Case study: Port of Bar, MSc dissertation, University of Montenegro, Maritime Faculty, Kotor, , Yugoslavia.

Noritake, M. and Kimura, S. (1990), Optimum allocation and size of seaports, Journal of Waterway, Port, Coastal, and Ocean Engineering, 116(2): 287-299.

Noritake, M. and Kimura, S. (1983), Optimum number and capacity of seaport berths, Journal of Waterway, Port, Coastal, and Ocean Engineering, 109(3): 323-339.

Talley, W.K. (2006), An economic theory of the port, Port Economics, Research in Transportation Economics, Volume 16: 43–65.

TranSystems Corporation (TSC) (2003), Port of Cleveland Capacity Assessment, Final report.

Wanhill, S.R.C. (1974) Further analysis of optimum size seaport, Journal of the Waterways, Harbors and Coastal Engineering Division, ASCE 100(4): 377-383.

Weille, J. and Ray, A. (1974), The optimum port capacity, Journal of Transport Economics and Policy, September: 244-259.

Škurić, M., Jovović, D, Čikmirović, E., Karanikić, Ž. and Popović, P. (2008), Optimizing capacity for port of Bar, 8th International Conference Research and Development in Mechanical Industry, RaDMI, Užice, Serbia: 1010-1019.

Obstacles to supply chain integration: empirical analysis of maritime firms

Jasmine Siu Lee Lam

Department of Transport and Regional Economics, University of Antwerp, Belgium.
School of Civil and Environmental Engineering, Nanyang Technological University, Singapore 639798.

Abstract

This paper aims to examine the possible barriers and problems to establish and maintain a partnership or collaboration among the members in a supply chain. Recommendations will be given on how to overcome these obstacles. The paper also aims to analyse supply chain integration of maritime firms. On top of theoretical analysis, we conduct empirical investigation by a survey in the form of interviews which targeted the professionals holding managerial positions in the top thirty container shipping lines as at 2007. Particularly, we examine the shipping lines' level of integration with the major shippers, and their level of integration with the major terminal operators. Findings show that integration exists more at the operational level, rather than the tactical and strategic levels. Inventory management is the area which is the least engaged. Individualism is a major obstacle to supply chain integration.

Keywords: Supply chain integration, obstacles, maritime, shipping lines

1. Introduction

Containerisation has been responsible for integration within the transport chain since its advent in the mid-1960s. In recent years, the maritime industry is progressing towards lower degree of fragmentation. Different forms of integration started to take place. It is not new to hear that various players in the supply chain work together to smoothen the cargo flow and information flow. While substantial benefits can be brought, there are obstacles to supply chain integration in practice.

After a thorough literature review (Lam, 2008), we find that publications in managing maritime transport as an integrated chain are very limited, and there is no paper which empirically investigates the integrated supply chain approach in shipping. This study attempts to fill this gap. Hence, this paper aims to examine the possible barriers and problems to establish and maintain a partnership or collaboration among the members in a supply chain. Recommendations will be given on how to overcome these obstacles. The paper also aims to analyse supply chain integration of maritime firms. Analyses are conducted based on both theoretical and empirical approaches.

2. Research Methodology

Theoretically, economics, management and supply chain concepts and theories are examined and applied. Empirically, a survey was conducted in 2007 with the professionals from the top thirty shipping lines in the world, based on Alphaliner (2007). The liner business is highly concentrated (Haralambides et. al., 2002; Lam et. al., 2007). Based on the slot capacity deployed in terms of TEU, the top thirty carriers accounted for a market share of 85.6%. With this high market share, the views from this group are considered representative of the major players in the liner industry. A pilot survey has been run before the actual field work. Two executives from top or middle management from each of the top thirty shipping lines are approached for the formal survey. Totally 60 executives from the lines form the sampling frame. The response rate is 90% with 54 successful responses. Several measures have been taken to mitigate the concerns on non-response bias and response bias.
In choosing a suitable data collection method, one has to consider issues such as research objective, problem definition, research settings and constraints. Compared to one-way communication when

completing the survey by the respondents, construct validity is enhanced in a two-way communication environment, where the definitions and concepts can be clarified for each respondent. This is important for the current study as the research topic is very new. Hence, door-to-door personal interview and telephone interview are considered appropriate. Face-to-face interviews are conducted whenever possible in order to achieve better results. However, as the shipping companies' offices are located in different parts of the world, it is not possible to travel to all the places for the interviews, given the budget and time constraints. Therefore, the second-best alternative is sought when long distance travel is not viable. Telephone interviews are conducted instead. Another advantage of conducting interviews is being able to secure a high response rate. A combination of close-ended questions and open-ended dialogue allows us to collect data and information for both quantitative and qualitative analyses.

3. Obstacles to Supply Chain Integration

3.1. Distribution of Benefits

3.1.1. Situations where benefits are distributed equitably

Successful management of a supply chain is a challenge to any firm or group of firms working together for a common business objective. Normally, each firm makes its decisions to pursue its individual objective. Assume the firms are commercial entities and each firm aims to maximise its profit, the total profit in the non-integrated supply chain is given by equation 1.

$$z = \max\ z_1 + \max\ z_2 + \max\ z_3 + ... + \max\ z_i \tag{1}$$

where
z is the total profit in a non-integrated supply chain
z_1 … z_i is the individual firm's profit in a non-integrated supply chain
i is the total number of supply chain members

To have a sustainable partnership, the chain members should share a common interest. Supply chain collaboration improves if the members of the chain take actions that together increase the total supply chain profit, more generally, the supply chain value. The chain must give all participants the possibility of benefiting. An ideal situation happens when the partnership produces an equitable distribution, i.e. a fair and reasonable distribution, of benefits among the chain members. The distribution of benefits can be estimated if the firms are willing to share the data. However, without sharing the supply chain data in many cases, it is each firm's perception of its own welfare that counts (Stiglitz, 2000). In the ideal case, all members are better off and they would be satisfied by the outcome. This is known as Pareto improvement, that is, a change which makes some firms better off without making any firm worse off (Stiglitz, 2000). This can be illustrated by equation 2 and equation 3.

$$Z = \max\ z_1(+z_{e1}) + \max\ z_2(+z_{e2}) + \max\ z_3(+z_{e3}) + ... + \max\ z_i(+z_{ei}) \tag{2}$$

$$Z - z = z_{e1} + z_{e2} + z_{e3} + ... + z_{ei} \tag{3}$$

where on top of the symbols denoted above,
Z is the optimal total supply chain profit after integration
z_{e1} … z_{ei} is the extra profit earned by the individual firm after integration

On top of the maximised individual profit, each supply chain member enjoys extra profit resulting from supply chain integration. The optimal total supply chain profit after integration Z is higher than the total profit in a non-integrated supply chain z. The higher margin is equitably distributed among the supply chain members. This is very conducive to develop a closer working relationship to achieve the common goal. The members are self-motivated in this circumstance.

3.1.2. Situations where benefits are distributed non-equitably

However, there are situations where the total supply chain profit can be enhanced by collaboration, but some chain members cannot directly enjoy the benefit. There could be two circumstances making them unwilling to collaborate. First, the total supply chain profit is increased at the expense of some chain members. Due to cost trade-offs, the chain members major in different areas of activities would display characteristics that put them in conflict with each other. Certain decisions may benefit the supply chain as a whole, but some members are actually adversely affected. For example, assume that a shipper exercises just-in-time manufacturing and aims to minimise the level of inventory. Yet, the shipper wants to offer flexibility and responsiveness to its customers. This decision requires a lean and agile, or so called 'leagile' (Goldsby et al., 2006), supply chain characterised by postponement strategy and close collaboration among the chain members. Transport cost would increase but the inventory carrying cost saving can balance the effect so that the ultimate result is better. Carriers would welcome this decision from shippers as revenue is increased. Nevertheless, warehouse operators may not react positively as their revenue would be reduced.

Second, the presence of gains that could be achieved might not be sufficient enough as a reason for collaboration as the costs of pursuing these gains by any one party could be substantial and might not be offset by the benefit obtained. If a firm can enjoy the benefit but the benefit cannot cover the cost of integration, it is still worse off after the integration. The cost of integration is the extra cost expends in collaboration which is involved in the flows of information, products and funds between various stages of the chain. For instance, the supply chain members may be required to invest more in information technology. They may have to develop a common database for sharing information and data which are essential for making decisions affecting the supply chain. They may even have to re-configure their information systems to ensure that they are compatible for smooth information flow. Another example of integration cost would be the effort to establish and maintain the relationship and the communication channel among the supply chain members.

Under either one of these circumstances, some chain members are worse off if the supply chain is more integrated. This can be illustrated by equations 4 and 5.

$$Z = \max z_1(-z_{e1}) + \max z_2(+z_{e2}) + \max z_3(+z_{e3}) + ... + \max z_i(+z_{ei}) \quad (4)$$

$$Z - z = -z_{e1} + z_{e2} + z_{e3} + ... + z_{ei} \quad (5)$$

As depicted by equation 5, the total supply chain profit after integration Z is still higher than the total profit in a non-integrated supply chain z. However, the optimal profit is achieved only at the expense of a supply chain member, firm 1. This firm will be worse off if it participates in supply chain collaboration and works towards the goal to maximise the total supply chain profit. While the other chain members can earn additional profits, firm 1's profit after integration is lower than the one given by acting on its own to maximise its firm's profit, i.e. max z_1 (- z_{e1}). This is known as a Kaldor-Hicks improvement, where the winners' gain is greater than the losers' loss (Mishan and Quah, 2007).

If the total supply chain profit is increased when some members are better off but some other members are worse off, it is obvious that those who are worse off have no incentive to collaborate. They will make decisions individually to maximise the firm's own profit, not the total supply chain profit. A mechanism should be created to motivate the participation from these firms. Naturally, those other firms which can enjoy greater benefits will take the initiative. We should also note that in some cases even all the chain members can achieve higher individual profits after integration, some chain members may be discontent if they cannot enjoy an equitable share of benefit. The problem deals with the fairness of the division of the total profits among the parties involved. The collaboration will tend to be unstable.

Hence, for the above-mentioned circumstances, there should be an agreement on distributing the benefits to ensure that all members can gain equitably from the partnership. The problem of incentive could be solved by an agreement to compensate those chain members who are worse off, like firm 1 in equation 4 (Ballou et. al., 2000). Also, the benefit allocated should be reasonable and sufficient to cover the cost of integration. As a whole, since partnership cannot be established and maintained by self-motivation, the agreement is to enforce a formal mechanism making all the chain members better off and satisfied with the outcome.

3.2. Mutual Dependence

Any collaboration should be built on mutual dependence. In other words, parties involved in the collaborative effort must recognise that their success is dependent on the success of each other. However, there are several conditions to fulfil if this situation is to be realised.

3.2.1. Major or minor business partner

From each supply chain member's perspective, the business with other parties in the supply chain should be, or at least perceived as, important to the firm. Also, the level of importance is expected to maintain in the future as the collaborative effort is for longer term. It ensures the certainty and stability of the supply chain partnership.

Very often, whether the business is significant can be judged based on whether the account is relatively major or minor. For example, if firm A is a major customer to firm B, and firm B is also a major supplier to firm A, both firms will treat each other's business as important. There is incentive from both sides to work together for better results. Kumar (1996) refers supply chain relationships of which the parties come together and exchange information and inputs in both directions as reciprocal interdependence. It is more likely to result in decisions that maximise supply chain profitability since the decisions have to take the objectives of both parties into consideration. In another case, assume firm 1 is a major customer to firm 2, but firm 2 is only a small supplier to firm 1. Firm 2 will treat firm 1 as key account, but firm 1 may not see firm 2 as an important partner. There is a lack of mutual dependence thus it is hard to establish a partnership between the firms.

Under such circumstance, the matter is beyond pure economical aspect to encompass the relational aspect of supply chain integration. For the above example, firm 1 may still be willing to collaborate with firm 2 if they have good relationship. Firm 2 has to put in continuous effort to manage the relationship with firm 1.

3.2.2. Visibility of the whole supply chain

High visibility of the implications exerted by the whole supply chain would encourage firms to work together. However, not many firms have high visibility of the whole system. For example, one company would know its immediate supplier and customer but not the supplier of its supplier or the customer of its customer. Often, a more dominant player within the chain will have higher visibility against the small players. The lack of trust and highly competitive nature of the business makes it even more difficult for firms to work for mutually beneficial outcomes. The supply chain members tend to be opportunistic and take actions without considering their impacts on the other parties.

The supply chain members can start collaborating with their major immediate customers and suppliers. The level of mutual dependence would be higher and it is more effective for bringing the partners together. When the parties learn that coordinated, joint efforts can lead to outcomes that exceed what the firm can achieve acting solely in its own interest, they will be more committed to the collaborative relationship. The supply chain partnership can be extended gradually afterwards.

Communication and trust are key ingredients for sustainable supply chain integration (LaLonde, 1998). Timely communication fosters trust by enhancing understanding and assisting in resolving conflicts. This is especially important when a high number of supply chain members are involved. In an empirical study, Panayides and So (2005) show that relationship orientation has a positive effect on key organisational capabilities, like organisational learning and innovation, promoting an improvement in supply chain effectiveness and performance. The study points out that a strong relationship between the logistics service provider and the client can increase integration, and consequently enhance the supply chain performance. Relationship management is an essential topic to be looked into for enhancing supply chain integration.

3.2.3. Shared values

It is quite often that the parties involved in the supply chain might have different corporate objectives. They could be pursuing aims other than pure commercial goals. In the study, we focus on commercial entities and assume that the supply chain members aim to maximise profit. Nonetheless, maritime supply chains usually involve governmental bodies such as customs and port authorities. These organisations are not commercial entities and normally do not aim at maximising profit. They often pursue goals aligning with national interest, for example, service excellence, facilitating trade, etc. Then they do not share common values with shippers, shipping lines and terminal operators. Therefore, difficulties may exist in building partnerships between commercial and non-commercial organisations.

Nevertheless, to overcome this obstacle, the commercial entities can work to align goals with those non-commercial organisations. Since enjoying higher profits would not be an incentive to encourage participation from non-commercial organisations, the approach should stress on other benefits that can be achieved by supply chain integration. For instance, synchronising the product flow can facilitate trade. The improvement of the overall supply chain performance can stimulate economic performance of the country. Again, communication cultivates collaboration by aligning perceptions and expectations on the joint effort.

3.3. Presence of Inertias

The presence of inertias in culture and organisation could also impede collaboration among the supply chain members. When the bias toward continuing past modes of behaviour is strong, the firms do not tend to seek new ways to manage shipping.

3.3.1. Complacency

If the corporate culture is risk adverse, conservative and resistant to make changes, the firm is not likely to promote supply chain collaboration, which is challenging and anticipates plenty of changes both internally and externally. The firm would value stability and comfort instead of innovation and improvement. It protects and reinforces traditional ideas and techniques. Hawkins and Gray (1999) empirically find out in the shipping context that senior managers' mindsets largely determine an organisation's strategic choices. Where senior managers are more conservative, an organisation is less likely to choose aggressive strategies. But in the fast changing business environment of container shipping, players have to innovate in order to counteract the market forces. Complacency makes people in a firm fail to advance. Complacent firms will be disadvantaged and overtaken by rivals.

The precursor of building supply chain partnership is the top management's determination to improve its supply chain performance. As stated by Porter (1998), the most important challenge to a firm is overcoming complacency and inertia to act on the new opportunities and circumstances. The challenge of action ultimately falls on the firm's leader. Therefore, the commitment of top management of the supply chain members is crucial to produce organisational change. The top management is to lead and motivate the whole firm in the process of achieving higher level of supply chain integration.

3.3.2. Realisation of benefits

Supply chain integration is a long term process. The partners have to be persistent in building and maintaining the partnership. As discussed previously, the partners involve the costs of integration which make the collaboration works. Normally, some of the costs are sunk costs at the beginning stage of the partnership. However, the benefits brought by the partnership can only be realised at a later stage. This means that even the firms can enjoy higher profits in the long run, they may have to bear loss in the short run. Because of this, some firms may wish to stay stagnant instead of taking the risk.

In order to alter the sluggish situation, the supply chain members have to recognise and focus on the long term supply chain performance and profitability. Similar to the above point, it is essential to have top management's support.

3.3.3. Reward system

The people in the firm are actors in the integration process. Supply chain collaboration among the firms requires the support of their staff. Hence, other than the incentive at the firm level, we have to consider the incentive at the individual level. The staff of a firm may be resistant to supply chain integration because their performance is evaluated by the traditional method based on the firm's individual key performance indicators and they are rewarded accordingly.

Therefore, establishing a reward system for identifying, measuring and passing the benefits arising from the collaboration to those who contribute to its success will encourage staff participation. The mechanism should create the impetus for self motivation.

3.3.4. Difficulty in start up

Another deterrent is the difficulty in starting up supply chain collaboration. It is particularly true for shipping as both research and industry practices in this regard are immature. There are too many considerations and the firms may not know where to start. They may hesitate and delay in taking actions.

To overcome this obstacle, a task force for building supply chain partnership should be set up. We recommend the model in the next section for planning and design purposes. The firms can draw up more detailed guidelines applicable to the particular situation in the maritime supply chain. They can take a gradual approach to implement the plan in realising integration.

4. Supply Chain Integration of Maritime Firms

4.1. Basis of the Survey Instrument

The survey is based on a normative model for achieving excellence in container shipping supply chains with the aim for better synchronisation and ultimately for value (i.e. profit for commercial setting) maximisation (Lam, 2008). It is a 3x4 matrix model having 12 cells. Examples of the cell activities are shown in figure 1. It is specified in mathematical form for quantitative analysis.

n / m	**Customer service** 1	**Inventory** 2	**Transportation** 3	**Order processing** 4
Strategic level 1	**11** Setting desirable service level	**12** Setting inventory management policies	**13** Selecting carriers; Selecting ports of call	**14** Determining system design; Order forecasting
Tactical level 2	**21** Collecting customers' feedback	**22** Determining safety stock level	**23** Seasonal capacity adjustments	**24** Assessing backorders
Operational level 3	**31** Handling customers' requests	**32** Replenishment quantities and timing	**33** Sea transportation; Loading and unloading	**34** Booking; Documentation

Figure 1: Model for achieving excellence in container shipping supply chains with numbers assigned

The mathematical model is specified as equations 6, 7 and 8:

Maximize $Z = f(x_{mn}, y_{mni}, a_{mn}, b_{mni})$ (6)

Subject to
$z_{ei} > 0$ (7)

$$Z = \sum z_i + \sum z_{ei} \quad (8)$$

where
Z is container shipping supply chain value
f is objective function
x is individual contribution of each cell
y is level of integration with supply chain member(s) of each cell
a is parameter of variable x
b is parameter of variable y
m is symbol for the level of activity, m = 1, 2, 3
n is symbol for the area of activity, n = 1, 2, 3, 4
i is symbol for the supply chain member(s), which can be downstream, denoted by d (shipper in the empirical test), or upstream, denoted by u (port/ terminal operator in the empirical test)
z_i is individual firm's profit in a non-integrated supply chain
z_{ei} is the extra profit earned by the individual firm after integration

4.2. Level of Integration with Major Shippers

In the survey, the respondents are asked to indicate to what extent the company engages their major shippers as supply chain partners in the cell activities by a continuum numerical scale. 1 denotes "not at all", 5 denotes "very large extent". Respondents can choose any value between the two extremes. The survey questions represent variables Y_{mnd} of the above model. After computing the mean scores based on the respondents' answers, it is found that the total mean score is 3.42. The most engaged cell activity with the major shippers is Y31D, representing customer service at operational level, with the mean score of 3.87. It is followed by Y33D, representing transportation at operational level, scoring 3.80 and Y34D, representing order processing at operational level, scoring 3.71. Cell Y12D, denoting inventory at strategic level, has the lowest score (2.89). The standard deviations of the cells Y_{mnd} are rather high, especially when compared to other cells. This means that the level of integration with the major shippers varies a lot across different liners. Table 1 depicts the mean score for each cell and the ranking, as well as the standard deviation.

As major shippers are important customers of the respondent companies (shipping lines), it is not surprising to see that they are most engaged in customer service at operational level. This cell activity is basic and essential for retaining the customers. As can be seen from table 1, the top three cell activities are all at the operational level. The results suggest that the respondent companies engage the major shippers more at this level. This is somewhat expected as it is easier to work with the supply chain partners in day-to-day operations. The level of commitment is lower than the tactical and strategic levels.

Table 1: Mean scores and ranking indicating the level of integration with major shippers

Rank	Cell		Mean score	Standard deviation
1	Y31D	Customer service, Operational	3.87	1.09
2	Y33D	Transportation, Operational	3.80	1.01
3	Y34D	Order processing, Operational	3.71	1.01
4	Y21D	Customer service, Tactical	3.67	1.18
5	Y23D	Transportation, Tactical	3.60	1.12
6	Y11D	Customer service, Strategic	3.46	1.24
7	Y32D	Inventory, Operational	3.41	1.03
8	Y13D	Transportation, Strategic	3.34	1.19
9	Y24D	Order processing, Tactical	3.23	1.11
10	Y22D	Inventory, Tactical	3.06	0.99
11	Y14D	Order processing, Strategic	3.01	1.12
12	Y12D	Inventory, Strategic	2.89	1.06
Total mean			3.42	

It seems that there is no distinct pattern on which area of cell activity is more engaged. But inventory may be the area which is the least engaged. As told by some interviewees, shippers may not be willing to collaborate in this area because they do not see the benefit of doing so. To be more specific, shippers often bear lower cost when they control the inventory in their own way without considering the effects on the whole container shipping supply chain. The reluctance from shippers could explain why cells Y12D, Y22D and Y32D indicate lower level of integration. It is especially true for the strategic level, whose score is lower than the medium level (2.89).

4.3. Level of Integration with Major Terminal Operators

The respondents are asked to indicate to what extent the company engages the major terminal operators as supply chain partners in the cell activities by a continuum numerical scale from 1 to 5, same as the last sub-section. The survey questions represent variables Y_{mnu} of the mathematical model. The total mean score is 3.54. The most engaged cell activity with the major terminal operators is Y33U, representing transportation at operational level, with the mean score of 4.05. The relatively consistent scores given by the respondents for this cell are reflected by the cell's lowest standard deviation (0.77) among Y_{mnu}. It is followed by Y31U, representing customer service at operational level, scoring 3.90 and Y23U, representing transportation at tactical level, scoring 3.80. Cell Y22U, denoting inventory at tactical level, has the lowest score (3.04). Table 2 depicts the mean score for each cell and the ranking, as well as the standard deviation.

Table 2: Mean scores and ranking indicating the level of integration with major terminal operators

Rank	Cell		Mean score	Standard deviation
1	Y33U	Transportation, Operational	4.05	0.77
2	Y31U	Customer service, Operational	3.90	0.87
3	Y23U	Transportation, Tactical	3.80	0.93
4	Y34U	Order processing, Operational	3.70	0.79
5	Y11U	Customer service, Strategic	3.65	0.89
6	Y21U	Customer service, Tactical	3.62	0.96
7	Y13U	Transportation, Strategic	3.48	1.17
7	Y24U	Order processing, Tactical	3.48	0.91
9	Y32U	Inventory, Operational	3.40	0.94
10	Y14U	Order processing, Strategic	3.27	1.00
11	Y12U	Inventory, Strategic	3.10	0.87
12	Y22U	Inventory, Tactical	3.04	1.10
Total mean			3.54	

We cannot see a very clear pattern of Y_{mnu} from the survey results. But since rank 1, 2 and 4 are cell activities at the operational level, similar to the last sub-section, it reveals that the respondent companies engage the major terminal operators more at this level, rather than the tactical and strategic levels. Also similarly, inventory may be the area which is the least engaged. Collaboration in the inventory area is to a large extent operational and reactive. In general, it is not the shipping lines' and terminal operators' priority in their agenda. This coincides with shippers' low priority put on in this area in terms of collaboration with the shipping lines.

As a whole, the level of integration with the major terminal operators, as indicated by the total mean score of 3.54, is moderately higher than the level of integration with the major shippers, as indicated by the total mean score of 3.42. To statistically test whether the two means are significantly different from each other, we perform the Analysis of Variance (ANOVA) test. The F-statistic is 4.1766 and its associated p-value is 0.0412. This means, the test statistic is significant at 5% level of significance. It proves that the two means are significantly different from each other.

5. Conclusions

The study has examined the obstacles to supply chain integration of maritime firms theoretically and empirically. The three levels of management based on the time frame involved in the normative model are: strategic, tactical and operational. The strategic and tactical levels involve longer time frame and wider scope of management. Empirical findings show that integration exists more at the operational level, rather than the tactical and strategic levels. Fostering collaboration at the tactical and strategic levels could be the step forward for bringing more benefits for the container shipping supply chain. In terms of the areas of activities, inventory management is the area which is the least engaged. Individualism is a major obstacle to supply chain integration. This study would be an interesting piece of work to various parties such as researchers, policy makers and market analysts.
For future research, the supply chain value generated by container shipping can be estimated by the survey data collected. Also, the proposed model is versatile and opens up new horizons in research. For example, the model can be applied, and modified to a certain extent if suitable, to different cases in practice.

References

Alphaliner (2007), www.alphaliner.com, last revision date: January 2007, accessed on 15 May 2007.

Ballou, R.H., Gilbert, S.M. and Mukherjee, A. (2000), New managerial challenges form supply chain opportunities, Industrial Marketing Management, Vol. 29, No. 1, pp.7-18.

Goldsby, T.J., Griffis, S.E. and Roath, A.S. (2006), Modeling lean, agile, and leagile supply chain strategies, Journal of Business Logistics, Vol. 27, No. 1, pp. 57-80.

Haralambides, H.E., Cariou, P. and Benacchio, M. (2002), Costs, benefits and pricing of dedicated container terminals, International Journal of Maritime Economics, Vol. 4, No. 1, pp. 21-34.

Hawkins, J.E. and Gray, R. (1999), Making strategic choices for Asia-Pacific shipping, International Journal of Maritime Economics, Vol. 1, No. 2, pp. 57-71.

Kumar, N. (1996), The power of trust in manufacturer-retailer relationships, Harvard Business Review, Nov-Dec, pp. 92-106.

LaLonde, B.J. (1998), Building a supply chain relationship, Supply Chain Management Review, Vol. 2, No. 2, Fall, pp. 7-8.

Lam, J.S.L. (2008), Managing Container Shipping Supply Chains: Modelling, Empirical Analysis and Applications. PhD dissertation, University of Antwerp.

Lam, J.S.L., Yap, W.Y. and Cullinane, K. (2007), Structure, conduct and performance on the major liner shipping routes, Maritime Policy and Management, Vol. 34, No. 4, pp. 359-381.

Mishan, E.J. and Quah, E. (2007), Cost-Benefit Analysis, 5th ed., Routledge, New York.

Panayides, P.M. (2006), Maritime logistics and global supply chains: towards a research agenda, Maritime Economics and Logistics, Vol. 8, No.1, pp. 3-18.

Panayides, P.M. and So, M. (2005), Logistics service provider–client relationships, Transportation Research Part E, Vol. 41, Issue 3, pp. 179-200.

Porter, M.E. (1998), The Competitive Advantage of Nations: with a New Introduction, Free Press, New York.

Stiglitz, J.E. (2000), Economics of the Public Sector, 3rd ed. WW Norton & Company, New York.

An application of the Porter's diamond framework: the case of Hong Kong airfreight industry

Yui Yip Lau

Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hong Kong.

Abstract

The Hong Kong airfreight industry faces keen domestic rivalry due to the emergence of Cathay Pacific's new air cargo terminal. The emergence of new air cargo terminal brings the airlines with low switching cost and strong bargaining power over air cargo terminals. They face cut-throat competition among each other. This paper investigates the internal and external environments and examines the performance of Hong Kong airfreight industry like the growth rate of air cargo throughput, operational efficiency and market share. Regarding the competitive business environment, we identify the key factors of strengthening the role of Hong Kong as an international air cargo hub. This paper applies the Porter's diamond framework to illustrate that how Hong Kong airfreight industry utilizes their inherent resources and enhances capabilities to compete with neighboring competitors like Guangzhou Baiyun International Airport and Singapore Changi Airport in dynamic and challengeable environment. This paper applies the concept of regional competitiveness to critically assess Hong Kong's potential evolution into an international air cargo hub. It provides an opportunity to look at the competition encountered by Hong Kong airfreight industry and grasps feasible opportunity in the external environment. The sustainable competitive advantage and first mover advantage will be achieved.

Keywords: Hong Kong Airfreight Industry, Domestic Rivalry, Porter's diamond framework, Regional Competitiveness, Sustainable Advantage, International Air Cargo Hub

1. Introduction

According to the Policy Address delivered by the Chief Executive of Hong Kong in 2001, the government of the HKSAR has recognized that the logistics industry was one of the four pillars in Hong Kong economy. The airfreight industry is an important element in Hong Kong logistics industry. In the review of airfreight industry in Hong Kong, there is a recent report conducted by GHK (Hong Kong) Ltd, an individual research organization authorized by The Airport Authority Hong Kong. Hong Kong is still the prime hub for forwarding logistics business among Southern China region. Hong Kong not only provides an extensive and excellent of transportation system, but also endows the leading international network connecting China to other parts of the world. Thus, it requires shorter flight time during the whole shipment. Compared with Taiwan, the required flight time is 6% shorter; Compared with Manila, the required flight time is 10% shorter; Compared with Singapore, the required flight time is 36 % shorter. It can help to save the total fuel cost by HK$40 millions annually as there are fewer flights time required (Schwieterman, 1993). In terms of invisible cost, including flight connections, custom clearance, cargoes handling efficiency, facilities, security, and Hong Kong International Airport (HKIA)'s advantages has outweighed over our neighborhood competitors. According to the report from GHK (Hong Kong) Ltd, it shows that using different airports can charge different charges and then affect the cost of shipment. From the Table 1, it can see that the shippers can gain the greatest cost advantage if they use HKIA. It revealed that there is the competitiveness of Hong Kong air cargo business.

Table 1: Comparison of Tangible Cost between Hong Kong International Airport, Guangzhou Baiyun International Airport and Shenzhen Baoan International Airport

Destination \ Airport	Hong Kong International Airport	Guangzhou Baiyun International Airport	Shenzhen Baoan International Airport
Los Angeles	HK$19,750	HK$24,900	HK$27,000
Frankfurt	HK$27,150	HK$28,400	HK$29,300
Tokyo	HK$18,800	HK$18,900	HK$19,300

Sources: GHK (Hong Kong) Ltd, 2006

In fact, Hong Kong has ranked as one of the world's leading international airports since 1996, handling about 3.7 million tons of cargo in 2007. The air cargo has recorded around 1.3% of Hong Kong's total cargo throughput, but it contributes to 35% of its total external trade value at HK$1,946 billions in 2007 (source: Hong Kong International Airport, 2008). Besides, there is a report commissioned by the Airport Authority Hong Kong, cargo volumes handled at HKIA will surge by 6% a year in the coming year. Last but not least, the Airports Council International mentions that Hong Kong is ranked the first in handling international cargo and the fifth in international passengers respectively in 2007. Table 2 illustrates the air cargo throughput at HKIA.

Table 2: The air cargo throughput at HKIA

Year	**Tonnage ('000 tons)**
1998	1,629
1999	1,974
2000	2,241
2001	2,074
2002	2,479
2003	2,642
2004	3,090
2005	3,402
2006	3,579
2007	3,742

Source: Hong Kong International Airport, 2008

However, there have been a few studies to review in airfreight industry, especially in Hong Kong. This study helps to illustrate that how Hong Kong airfreight industry utilize their inherent resources and enhance capabilities to compete with neighboring competitors in the dynamic and challengeable environment. The resources are heterogeneity, rare, imperfectly mobile, imitable and non-substitutable (Peteraf, 1993). The critical resources can either tangible like infrastructure, airport facilities and configuration, or intangible like individual expertise and skills, know-how, reputation and customs particle that the Hong Kong airfreight industry owns, controls and access to on a semi-permanent basis (Helfat and Peteraf, 2003; Lai, 2004; Valentin, 2001). On the other hand, the capacity is related to the competences and capabilities of the Hong Kong airfreight industry to perform a coordinated set of tasks to build, integrate and reconfigure the internal and external resources and capabilities so as to appropriately match the opportunities in the environment (Helfat and Peteraf, 2003; Teece et al., 1997). In the literature, it is suggested that the core competence can help the firms to differentiate themselves so as to build up their customer-focused capabilities (Lai, 2004). The customer-focused capability is the concept of market orientation and it comprises segmental focus, relevancy, responsiveness and flexibility (Zhao et al., 2001). Following this concept, it can help the Hong Kong airfreight industry to differentiate from their competitors on providing a wide variety of air cargoes service and attain better service performance in the final outcome (Lai, 2004).

Therefore, this study aims to (1) explore the internal and external environments of the Hong Kong airfreight industry, (2) apply the Porter's diamond framework to identify the key factors of strengthening the role of Hong Kong as an international air cargo hub and how Hong Kong airfreight industry utilizes their inherent resources and enhances capabilities to compete with neighboring competitors, (3) propose directions for the Hong Kong airfreight industry to sustain competitive advantage and enhance regional competitiveness in Asia region, (4) discuss under-explored topic for future research in Hong Kong airfreight industry.

2. The Current State of the Hong Kong Airfreight Industry

Thanks to globalization, the Hong Kong airfreight industry facing the stiff competition and rising expectations from their customers. Accordingly, the airfreight industry not only provides port-to-port transport, but also offers a wide variety, customized and differentiated customer services with different customer requirements like priority lift, courier lift, secure lift, live animal lift and fresh lift. The Hong Kong airfreight industry requires changing from manual operation to automatic operation so as to deliver the services and the product at the right time, at the right place, at the right condition and at the right people. Recently, Cathay Pacific (CX) initiates actively to become Asia's biggest carrier. CX uses HK$8.22 billions to take over the business of Dragon Airline by stock acquisition (source: Cathay Pacific Airways Limited, 2006). This acquisition make Cathay Pacific Asian‘s biggest airline which allows CX to develop its mainland flight route. After CX took over Dragon Airline‘s share, it is accessibility to mainland market will bring in more cargo capacity for CX. Because of profitable routes into the mainland's fast-growing airfreight market and the leading air cargo terminal operator in Hong Kong, the Super Terminal 1 is insufficient to provide effective cargo handling service to CX. In 2007, the Super Terminal 1 annual throughput of air cargoes around 2.63 million tones and the land area is 171,322 sq.m. The overall land use efficiency is approximately 15.4 tons per sq.m (source: Hactl, 2007; Hactl Press Release, 2008). Their efficiency will be further reduced when the cargo volume increase in the coming years. At the same time, a 10-year contract between CX and Hong Kong Air Cargo Terminals Limited (Hactl) will be expired in 2009. In a view of this, CX develops an independent air cargo terminal seems to be inevitable. CX invests an approximately HK$4.8 billions to set up an independent cargo terminal in the second half of 2011 and have gained the award of franchise for 20 years at HKIA (source: Cathay Pacific Cargo Press Release, 2008). It can deal with the airline's rapid growth in cargo volume and product enhancement, especially mainland China. With Cathay Pacific Airways plans to set up an independent new air cargo terminal at HKIA, it brings an additional cargo handling capacity and attracts more cargo flights at HKIA. It further consolidates Hong Kong as an international air cargo hub in the world.

In case of CX’s new terminal is recognized as a third cargo terminal operator (CTO), Asia Airfreight Terminal (AAT) will be found herself in a more stiff competition due to other airlines can also use CX’s cargo terminal. Nevertheless, CX establishes new air cargo terminal creates competitive pressure and rivals on Hactl and AAT. On the other hand, CX intends to establish its own airfreight terminal by backward vertical integration (Slack and Fremont, 2005). It can be viewed as the value-added chain in a total integrated logistics system and integrated conglomerates (Davis and Duhaime, 1992; Notteboom, 2002). Through backward vertical integration, company sets up their subsidiaries that the whole supply chain are managed itself. (i.e. they produce some of the inputs and use it in the production process. The final outcome is to produce own products). In this case, CX hopes to provide air cargoes handling services by its own subsidiary company in the return of benefits. In the commercial level, CX can enjoy the low cost competitive advantage due to efficiency increases and input utilization (Fronmueller and Reed, 1996). Compared with Hactl’s land use efficiency, the target of new air cargo terminal can bring over 25 tones per sq.m (source: Cathay Pacific Cargo Press Release, 2008). Also, CX can develop the diversification in this industry and one stop-shop services to customers. It can improve the customer service level accordingly (Notteboom, 2002; Slack and Fremont, 2005). In the strategic level, CX can enjoy the lower cost of operation and then can offer the lower price to the customers. In other words, it can attract other customers to use CX air cargo service from other air cargo terminals and build up their cargo network extensively. On the negative side, the backward vertical integration creates significant costs for CX like administrative cost, strategic cost (i.e. sunk costs and commitment escalation) because of the problem of complexity in coordination and communication within CX (Fronmueller and Reed, 1996).

The Hong Kong airfreight industry not only faces domestic rivalry, but also encounters neighboring pressure in Asia region. Thanks to open skies policies, airline market deregulation and alliance, it creates the pressure and strong competition in the airfreight industry (Francis et al., 2002; Park, 2003). The open skies policies stimulate the liberalization in the aviation industry. It will further remove the capacity control of each route and entry barriers (Forsyth et al., 2006). In addition, it permits airlines to set prices and quantities without any restriction while they are free to form the ownership arrangements and alliance without government intervention (Adler and Hashai, 2005; Forsyth et al., 2006). Consequently, the notion of regional competitiveness facilities the Hong Kong airfreight industry to encounter a competitive market in Asia region. The regional competitiveness (European Commission, 1996) is defined as "the ability to produce goods and services which meet the test of international markets, whilst at the same time maintaining high and sustainable levels of income, or more generally, the ability of regions to generate, while being exposed to external competitions, relatively high incomes and employment levels." The regional competitiveness leads to the Hong Kong airfreight industry to achieve a sustained competitive advantage. The sustainable competitive advantage is related to the Hong Kong airfreight industry implements value creating strategy which is not able to duplicate and difficult to implement or imitate by the neighboring rivals in a longer period (Barney, 1991; Dierickx and Cool, 1989; Oliver, 1997). The nature of airfreight industry is fundamentally heterogeneity that they have comparative advantages and disadvantages in the dynamic environmental circumstances (Barney, 1991; Peteraf, 1993). According to Ghemawat (1986), there are three effective strategies mention that it can create the sustainable competitive advantage in Hong Kong airfreight industry. They are (1) the size in the targeted market, (2) superior access to resources or customers and (3) restrictions on competitors' options. Hence, they will exploit their inherent strength and weakness as well as responding to opportunities and threaten others (Barney, 1991; Valentin, 2001).

3. Application of the Porter's Diamond Framework in Hong Kong Airfreight Industry

The concept of regional competitiveness has gained considerable attention and widespread support from the associated institutions and academicians. It brings significant effect on enhancing economic performance like wage, employment rate, exchange rate, trade volume and GPD per capital as well as gaining sustainable competitive advantage of the regional development (Porter, 2003). In the regional competitiveness, the location factors cannot be ignored because (1) the localization of industry foster the specialized local providers of inputs to production, (2) the speedy of the diffusion of information generates technological spillovers in the localized industry, (3) the pooling of specialized labor stimulates the local demand (Budd and Hirmis, 2004). It facilitates the agglomeration economies and upgrades the productivity in the clusters (Porter, 2000). According to Porter (2000), each industry can be grouped into the cluster of related and supporting industries plays a significant role on the economic development. The Hong Kong airfreight industry can be grouped as the clusters where it consists of freight forwarders, third party logistics service providers, airline companies, air cargo terminal operators, government, truckers, trade associations. The clusters are also linked by commonalities and complementarities that create the spillovers across airfreight industry and professional and academic institutions (Porter, 2000; Porter, 2003). Indeed, the Porter's diamond framework (1998) and the regional competitiveness can provide a theoretical foundation to assess how Hong Kong match its internal competences to its external environment and hence consolidate as an international air cargo hub (Porter, 1981).

3.1. Porter's Competitive Advantage of Nations

The Porter's national 'diamond' framework integrates the comparative advantage of different industries with the theory of competitive strategy (Grant, 1991). The nature of industry is fundamentally heterogeneous in resources and capabilities leads to have a comparative advantages and disadvantages in the dynamic environmental circumstances (Peteraf, 1993). With the analysis of the Porter's competitive advantage of nations, it is the firms rather than nations and it is the principal actors. In fact, the principal role of the nation is the 'home base' which can help to shape the identity of the firm, the character of the top management, the approach to strategy and organization and to determine the availability and qualities of the resources available to the firm. In the Porter's diamond framework (1998), it has identified four interacting determinants that constitute the 'national demand'. Hence, it affects industry's ability to establish and sustain competitive advantage in the competitive environment (Grant, 1991).

- *Factor conditions:* it refers to the factors of production creates the industry's comparative advantage in the international market. Basic factors can provide the initial advantages like the natural resources, climate and location (Grant, 1991; Porter, 1998). For instance, Hong Kong is located at an optimal location. People can move toward over half of the world population within 5 flying hours. There are over 85 airlines (including 19 cargo-only carriers) operated in HKIA. The HKIA has connected with more than 150 locations in 47 countries around the world. Aircrafts land off and on the airport for about 800 times everyday. The frequency of flight provides shippers and passengers' flexibility in cargo-shipping and traveling (source: Hong Kong International Airport, 2007). Due to Hong Kong's high accessibility, Hong Kong have been established as an aviation hub in Asia-Pacific region. However, the basic factors cannot provide sustainable competitive advantage leads to the individuals, governments and companies invest in the advanced factors. The specialized factors of production include individual expertise and skills, know-how, infrastructure, technology and communication (Grant, 1991; Helfat and Peteraf, 2003; Lai, 2004; Porter, 1998; Valentin, 2001). Such as, The Airport Authority Hong Kong invests HK$4.5 billions to expand its apron in order to serve new large aircraft A380 and the Low Cost Carrier like Jetstar Asia and Northwest Airlines. HKIA can maintain its competitive advantage as it increases its capacity in terms of passenger and cargo without providing extra timeslot. Most importantly, The Airport Authority Hong Kong invests HK$3 billons on the addition of Cargo Stands and taxiways. There will be 10 additional Cargo Stands in 2007. Thus, it can increase their cargo capacity as HKIA forecasts that the cargo growth will be expected to average 6% per year in the next 20 years (source: Hong Kong International Airport, 2007). The advanced and the specialized factors can help Hong Kong airfreight industry gain the strategic position with the factors of production.
- *Demand conditions:* the industry innovates faster and the sophisticated 'home demand' can create the competitiveness. The industry needs to respond the sophisticated home demand by rapid improvement of product and offering the superior product quality, features and service (Porter, 1998). Such as, HKIA and Hactl through his subsidiary Hong Kong Air Cargo Industry Services Limited (Hacis) have formed the collaboration on establishing the Air Cargo Consolidation Centre (ACCC) in the Futian Free Trade Zone in Shenzhen. It is just only 1 km away from the Lok Ma Chau border crossing in Hong Kong. Under ACCC, It can help to improve efficiency for exporters and cargo agents and ensure the reliability and enhance connectivity within the HKIA catchment area. Most importantly, it can facilitate the flow of goods to Hong Kong as well as enhancing HKIA's position as a primary cargo gateway for China (source: Hactl Press Release, 2003).
- *Related and supporting industries:* the related and supporting industries are regarded as the complementary products or services of the industries. The close working relations and the ongoing coordination of related supporting industries enhance the competitive advantage of the industries (Porter, 1998). There are many actors participate in the Hong Kong aviation industry like the airline companies, freight forwarders, air cargo terminals, third party logistics service providers and trucker. Through the information technology, it can closely coordinate with different parties in the same platform. Take the real case of AAT. AAT has participated in the Digital Trade and Transportation Network (DTTN) pilot where it transforms the house manifest data from the forwarders' Freight Management system automatically to the AAT preferred format through the DTTN platform. Hence, AAT can receive timely and accurate information as well as the forwarders can save time in capturing data of the AIMS. The AIMS is a web-based application system and platform that enrich the operational transparency and fully integrate with government, airlines and other air cargo community systems. Under this tool, it facilitates the shipment declaration process by submitting the house manifest faster and accurately and traces the shipment record in real time (source: http://www.aat.com.hk).

- *Firm strategy, structure and rivalry:* the rivalry is an essential ingredient of the competitive advantage of the industry. The domestic rivalry leads to the visible pressure on the firm to lower costs, improve quality and innovation. Hence, it can upgrade the competitive advantage of the industry (Grant, 1991). For instance, Hong Kong airfreight industry is a competitive business leads to the domestic rivalry between each other. Hactl and AAT are the same nature of business and they compete with each other. Because of the technology improvement and process re-engineering, Hactl has increased their handling capacity from 2.5 millions tons to 3.5 millions tons of cargo. At the same time, AAT Terminal 2, which cost HK$1.75 billions in construction, will be completed by the end of 2006 with an annual handling capacity of 910,000 tons. Together with its terminal 1, AAT can provide 1.5 millions handling capacity (source: Hong Kong International Airport, 2008). After CX established their air cargo terminal, it will further increase the cargo handling capacity and efficiency and the airlines have more choices of airfreight terminal and gain the benefit like rebate.

The 'diamond' framework is an interactive system which the four determinants are highly correlated and interdependently with each other and upgrade the competitive advantage of industry over time. The domestic rivalry accelerates the Hong Kong airfreight industry innovates and invests the advanced factors of production to offer specialized supporting service (Grant, 1991). The supporting industries are able to respond the home demand (Porter, 1998). Such as, Hacis has implemented the Superlink China Direct. It is the customs-bonded air-road connection between HKIA and 57 locations around the PRD region. Also, the Northbound and Southbound modes provide consolidated truck services and chartered truck service for Import and Export cargos. In addition, there are six inland cargo depots serve as a cargo collection and distribution centre in the PRD region. Apart from this, the inland cargo depots with applicable IATA city code which enable airlines to use a direct Master Airway bill to ship cargo from overseas to major cities in the PRD regions. In other words, it can provide one-stop service to the shippers. However, the domestic rivalry and the customers are highly sophisticated and knowledgeable leads to rapid improvement and upgrade their service by the Hong Kong airfreight industry. The advanced and specialized factors like physical infrastructure, skilled personnel and knowledge bases in airfreight industry can help the Hong Kong airfreight industry to upgrade their competitive advantage continuously (Porter, 1981; Porter, 1998). Hacis extend his Superlink China Direct Northbound service between HKIA and Humen, Dongguan. Hacis forms the joint venture with Dongguan Humen Great Trade Container Port Company Limited so as to extend their cargo network at the Inland Cargo Depots in the Pearl River Delta (PRD) as well as the vast manufacturing base in South China. The integration of South China region enhances the cross-boundary efficiency and hence reducing cargo transit time and transportation cost accordingly. It can provide time-definite delivery of cargo and stable service to shippers (source: Hacis Press Release, 2005).

The government plays a vital role on enhancing the role of Hong Kong airfreight industry as an international air cargo hub. The government measurements can influence the four determinants indirectly and partially through the variety of actions in the Porter's 'diamond' framework (Porter, 1998), shown as Figure 1.

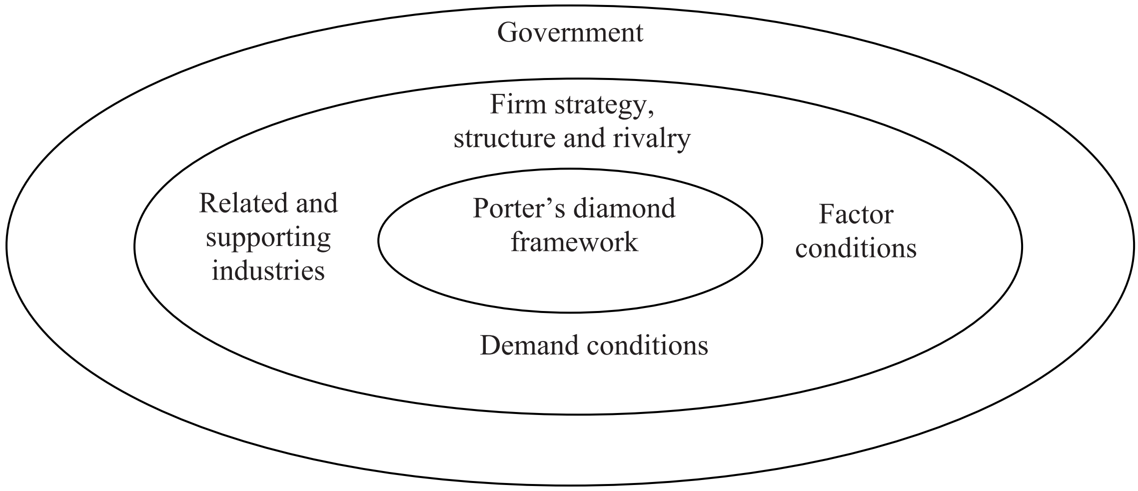


Figure 1: Porter's Diamond Framework
Source: Porter, 1998

- Providing the subsidies to the airfreight industry in Hong Kong. For the land or surface connectivity, it must continue an appropriate physical infrastructure development in order to streamline the border crossings and ease of infrastructural bottlenecks. For instance, the Hong Kong-Shenzhen Western Corridor provides direct access to the eastern PRD area while Hong Kong-Zhuhai-Macau Bridge enhances the connections between HKIA and western PRD area. It can integrate the hinterland within PRD area and bring more cargo sources from PRD area to Hong Kong (source: HKIA Master Plan 2025).
- Establishing a trade-freedom city. Hong Kong has recorded their weighted average tariff rate was zero percentage in 2005 and Hong Kong's economy is 90.3% free in 2008 (source: 2008 Index of Economic Freedom). As most of the commodity enter Hong Kong are tariff-free, it can attract the enterprises do the investment in Hong Kong and absorb cargo.
- Preparing the fifth freedom to Hong Kong by the CAAC. There will be the ASEAN Open Skies in 2015. There will be full liberalization of air services within Asia region (Ionides, 2008). It implies that Hong Kong will be easier to enter into the mainland and increase the catchment area in mainland regions due to without any restrictions for carriers to fly within Asia region.
- Developing creative and value added logistics services to shippers. The government collaborates with air cargo operators to develop streamline intermodal transportation network through customs clearance technologies like a common e-platform. It can simplify the custom clearance procedure and enhance the flexibility of cargo delivery (source: HKIA Master Plan 2025).

Recently, the academic institutions and literature reviewers are increasing awareness of regional and national economies by using the concept of "industry clustering" (Porter, 2003). A cluster emphasis the importance of coordination and mutual improvement while the reducing level of rivalry intensity and competition in the community (Porter, 2000). According to Porter (2003), he defines "clusters act as a geographical proximate group of interconnected companies, suppliers, service providers and associated institutions in a particular field." The cluster analysis is a classification tool that can help the Hong Kong airfreight industry characterized as clusters of related and supporting industries (Punj and Stewart, 1983). It can help to define different cluster boundaries in the Hong Kong airfreight industry (Lun et al., 2009).

- 1st party users: parties that own the products to deliver like Small Medium Enterprises (SMEs), global traders.
- 2nd party users: parties that own the airport facilities or vehicles to offer delivery and logistics services like airlines, air cargo terminal operators, warehouse operators, truckers.
- 3rd party users: parties that provide customized services to the shippers like third party logistics providers, freight forwarders.

- 4th party users: parties that implement the trading regulations like trade associations and government to meet customer requirements and increase customer service level.
- 5th party users: parties that provide consultation services and research studies like research centre and academic institutions to enhance the regional competitiveness of Hong Kong airfreight industry.

Clusters are important because of the externalities connect the constituent industries like common technologies, skills and knowledge (Porter, 2000). It creates knowledge spillovers and interconnections between upstream firms like air cargo terminal operators and downstream firms like freight forwarders. Thus, it not only facilitates the vertical and horizontal linkages among the boundaries of industries, but also encourages the agglomeration activities and specialized service in a regional economics (Budd and Hirmis, 2004; Cooke, 1997; Grant, 1991; Porter, 2003). Recently, the development of Traxon is a breakthrough in the Hong Kong airfreight industry. It can gain global connectivity and worldwide access with different partners like airlines, freight forwarders, airports, GHAs, GSAs. Besides, it can increase efficiency and enhance flexibility due to real time communication in the electronic neutral platform (source: http://www.traxon.com).

3.2. Regional Competitiveness

The competitiveness is a universal term which is mostly applicable to the business and economic environment. It is a way of benchmarking which is used for the firm to evaluate themselves and compare the internal performance with competitors. Hence, the firm gets superior performance by ongoing improvement (Budd and Hirmis, 2004; Francis et al., 2002). According to the Department of Trade and Industry (UK), “the competitiveness is the ability to produce the right goods and services of the right quality, at the right price and at the right time” (source: http://www.dft.gov.uk). The competitiveness is also defined as the firms has the ability to compete in the international market through the policy inputs like business environment, physical infrastructure and knowledge infrastructure as well as the essential conditions like business performance, productivity, price and cost and labor supply so as to increase efficiency and effective (National Competitiveness Council, Ireland, 2006). In addition, the Organization for Economic Cooperation and Development (1996) identify the national competitiveness is “the degree to which it can, under free and fair market conditions, produce goods and services which meet the test of inter-national markets, while simultaneously maintaining and expanding the real incomes of its people over the long term.”

According to the study of Porter (2003), the regional competitiveness is related to the Hong Kong airfreight industry competes with neighboring rivals by attracting investment from foreign, private and public capital, creating innovation environments by skilled employees, entrepreneurs and creative workers and facilitating the technological development. To illustrate, HKIA can attract the foreign, private and public capital due to receive revenue earnings from air side support services franchises and there is profit attributable to equity shareholder of HK$2,273 millions. Besides, the HKIA creates innovation environments with about 60,000 skilled labor workforces. Last but not least, the airfreight terminal like Hactl and AAT has fully automation so as to increase cargo handling capacity and enhance security level (source: HKIA Master Plan 2025). The HKIA creates the agglomeration economies by a high employment level and a high income consists of aeronautical revenue like landing fees, parking fees, aerobridge fees and passenger services and security tax and non-aeronautical revenue like airport concession fees and franchise fees in the region. It can contribute to Hong Kong’s evolution into an international air cargo hub.

4. Conclusion

This paper discusses the importance and the role of airfreight industry in Hong Kong. In the coming years, CX will build own cargo terminal leads to the stiff competition between each other and reduce the level of monopolize in the aviation industry. This brings the significant effect on Hong Kong’s airfreight industry. The airlines will have more choice of air cargo terminal and gain the benefit like reducing the cargo handling charges and rebate due to the cut-throat policy. Besides, it can increase the cargo handling capacity so as to meet the increasing demand in the future. In this study, we apply the concept of regional competitiveness to critically assess Hong Kong's potential evolution into an international air cargo hub. To deal with this situation, we address some practical measures for Hong Kong airfreight industry to encounter the competition and grasp the feasible opportunity in the external environment. Thus, it can gain sustainable competitive advantage and first mover advantage.

Actually, there are a few studies and literature to discuss and analysis on the airfreight industry, particularly in Hong Kong. In the past study, few of the literature use the qualitative method to analysis in the aviation industry. This study not only explores the internal and external situation of airfreight industry in Hong Kong, but also applies the Porter's diamond framework to identify the key factors of strengthening the role of Hong Kong as an international air cargo hub and how Hong Kong airfreight industry utilizes their inherent resources and enhances capabilities to compete with neighboring competitors. To supplement that idea, we also apply the concept of regional competitiveness to highlight some important points about how to sustain Hong Kong as an international air cargo hub in the literature. To begin with, the government should develop the hardware in the air side and land side within the transportation network so as to increase HKIA's capacity and connective. Hence, it can build up the multi-modal transportation. With the help of closer regional linkage and new facilities with advance technology, Hong Kong airfreight industry's crowns will not being dethroned by the neighboring competitors. Secondly, Hong Kong needs to establish a trade-freedom city where it can attract the enterprises do the investment in Hong Kong and absorb cargo. Thirdly, HKIA needs to cooperate with mainland like the customs clearance and the bilateral agreement and freedom of air so as to increase the connection to mainland area. In other words, it can attract more airlines to use HKIA and then increase revenue in the final outcome. Last but not least, Hong Kong's airfreight industry is not only to maintain a cost-effective expansion strategy, but also need to provide value-added service so as to exceed customer expectation. We believe that Hong Kong's airfreight industry will have prosperity in the future.

On the other side, we should take account of some limitations in this study. This study mainly focuses on the business view to analysis the airfreight industry in Hong Kong. However, there is a lack of study on the airfreight operation and focus on the cargo side in the aviation industry only. On the other hand, this study mainly focuses on the qualitative analysis to explain the framework for managing Hong Kong airfreight industry. In the future research, we carry out an extensive survey so as to get a comprehensive study and thoroughly analysis of the Hong Kong airfreight industry. This can inspire our new insights about the framework for managing Hong Kong airfreight industry.

5. Acknowledgments

The earlier research of this paper was received an award from the Chartered Institute of Logistics Transportation in Hong Kong's Student Scholarship 2007.

References

Adler, N. and Hashai, N. (2005), Effect of open skies in the Middle East region, Transportation Research Part E 39: 878-894.

Air Cargo Forum (2006), Look into the competitiveness of HKIA as an international and regional air cargo hub, GHK (Hong Kong) Ltd, 2006.

Asia Airfreight Terminal (2005), DTTN Pilot Run with AAT, Asia Airfreight Terminal.

Barney, J. (1991), Firm resources and sustained competitive advantage, Journal of Management 17(1): 99-120.

Budd, L. and Hirmis, A.K. (2004), Conceptual Framework for Regional Competitiveness, Regional Studies 38(9): 1015-1028.

Cathay Pacific Airways Limited (2006), Annual Report 2006, Cathay Pacific Airways Limited

Cathay Pacific Services Ltd awarded new cargo terminal franchise in Hong Kong, Cathay Pacific Cargo Press Release, 18 March, 2008.

Cooke, P. (1997), Regions in a Global Market: the experience of Wales and Baden-Wurttemberg, Review of International Political Economy 4(2): 349-381.

Davis, R. and Duhaime, I.M. (1992), Diversification, vertical integration, and industry analysis: new perspectives and measurement, Strategic Management Journal 13: 511-524.

Dierickx, I. and Cool, K. (1989), Asset stock accumulation and sustainability of competitive advantage, Management Science 35(12): 1504-1511.

European Commission (1996), Cohesion and Competitiveness: Trends in the Regions, 6th Periodic Report on Social and Economic Situation and Development of the Regions in the Community, European Commission, Luxembourg.

Francis, G., Humphreys, I. and Fry, J. (2002), The benchmarking of airport performance, Journal of Air Transport Management 8: 239-247.

Forsyth, P., King, J. and Rodolfo, C.L. (2006), Open skies in ASEAN, Journal of Air Transport Management 12: 143-152.

Fronmueller, M.P. and Reed, R. (1996), The competitive advantage potential of vertical integration, Omega 24(6): 715-726.

Ghemawat, P. (1986), Sustainable Advantage, Harvard Business Review 64(5): 53-58.

Grant, R.M. (1991), Porter's competitive advantage of nations: an assessment, Strategic Management Journal 12: 535-548.

Hacis launches SuperLink China Direct service extension to Humen, Dongguan, Hacis Press Release, April 12, 2005.

Hactl and Airport Authority in Shenzhen Cargo Consolidation Centre, Hactl Press Release, October 3, 2003.

Hactl sets new annual tonnage record, Hactl Press Release, January 10, 2008.

Hong Kong Air Cargo Terminals (HACTL) (2007), SuperTerminal 1 – The world's largest air cargo terminal at your service.

Helfat, C.E. and Peteraf, M.A. (2003), The dynamic resource-based view: capability lifecycles, Strategic Management Journal 24: 997-1010.

HKIA Master Plan 2025 (2007), Hong Kong International Airport, Airport Authority Hong Kong

Hong Kong International Airport (2007), HKIA: Where Opportunities Converge, Hong Kong International Airport Annual Report 2006/2007, Hong Kong International Airport, Hong Kong.

Hong Kong International Airport (2008), Air Cargo Development: Aviation Logistics Business, Hong Kong International Airport, Hong Kong.

Ionides, N. (2008), Asian open skies on track, Airline Business 24(1): 22.

Lai, K.H. (2004), Service capability and performance of logistics service providers, Transportation Research Part E 40: 385-399.

Lun, Y.H.V., Lai, K.H. and Cheng, T.C.E. (2009), Intermodal transport capability, Shipping and Transport Logistics Book Series 1:17-33.

National Competitiveness Council, Ireland (2006) Annual Competitiveness Report 2006, Volume 1: Benchmarking Ireland's Performance, National Competitiveness Council, Ireland.

Notteboom, T.E. (2002), Consolidation and contestability in the European container handling industry, Maritime Policy and Management 29(3): 257-269.

Oliver, C. (1997), Sustainable competitive advantage: combining institutional and resource based view, Strategic Management Journal 18(9): 697-713.

Park, Y. (2003), An analysis for the competitive strength of Asian major airports, Journal of Air Transport Management 9: 353-360.

Peteraf, M.A. (1993), The cornerstones of competitive advantage: a resource-based view, Strategic Management Journal 14: 179-191.

Porter, M.E. (1981), The contributions of industrial organization to strategic management, Academy of Management Review 6(4): 609-620.

Porter, M.E. (1998), The Competitive Advantage of Nations, New Edition, Macmillan, London.

Porter, M.E. (2000), Location, Competition, and Economic Development: Local Clusters in a Global Economy, Economic Development Quarterly 14(1): 15-34.

Porter, M.E. (2003), The Economic Performance of Regions, Regional Studies 37 (6&7): 549-578.

Punj, G. and Stewart, D.W. (1983), Cluster analysis in marketing research: review and suggestions for application, Journal of Marketing Research 20(2): 134-148.

Schwieterman, J.P. (1993), Air cargo and the opening of China, The Chinese University Press, Hong Kong.

Slack, B. and Fremont, A. (2005), Transformation of Port Terminal Operations: From the Local to the Global, Transport Reviews 25(1): 117-130.

Teece, D.J., Pisano, G. and Shuen, A. (1997), Dynamic capabilities and strategic management, Strategic Management Journal 18(7): 509-533.

Valentin, E.K. (2001), SWOT analysis from a resource-based view, Journal of Marketing Theory and Practice 9(2): 54-69.

Zhao, M., Droge, C. and Stank, T.P. (2001), The effects of logistics capabilities on firm performance: customer-focused versus information-focused capabilities, Journal of Business Logistics 22(2): 91-107.

Solving train blocking and scheduling problem by using a bi-level programming model and genetic algorithm

Haiying Li, Zhaoxia Yang, Lingyun Meng, Xi Jiang and Jianrui Miao

State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, 100044, China.

Abstract

This contribution deals with the train blocking and scheduling problem with flexible time window to better use railway resources and meet freight shippers' requirements from railway logistics perspective. This implies not only the need for organizing car blocks into trains, but in particular the burden of scheduling trains. To solve such problems, the application of bi-level programming and genetic algorithm techniques is proposed. Regarding the bi-level model, mathematical formulations are presented. In the bi-level model, a solution of the upper level planning stage determines the train operating plan. The aim of the upper stage is to maximize the total weighted sum of the profit of railway company and the convenience of shippers. More specifically, the profit of railway company mainly depends on the revenue by transporting goods and the cost caused by operating trains on railway lines and marshalling cars in yards. Based on the solution in the upper stage, the lower level planning stage is designed to determine how cars are grouped into trains in order to minimize the cost of shippers, in result, the solution in the lower stage reflects the profit of the upper stage. After mathematical formulations, a parallel genetic algorithm is designed and implemented, including solution procedure, sorting, selection, crossing and mutation rules. Finally, comprehensive numerical experiments demonstrate the applicability of the method.

Keywords: Railways transportation; Train Blocking; Train scheduling; Bi-level programming; genetic algorithms

1. Introduction

Railway has been faced with fierce competition and decreasing market shares in the area of freight transportation in China. In this context, more sophisticated and better planning tools are necessary in order to obtain a more rational and efficient utilization of available resources.

Railway typically develops operating plans to govern the movement of cars. Car movements are performed by assembling cars in trains through successive classification in order to achieve efficiency through economies of scale, and cars having the same travel direction are classified (i.e. grouped) into blocks according to a grouping or blocking plan. Blocks with the same general travel direction are then organized into trains according to a train makeup or train formation plan. After sufficient cars have been accumulated, trains will be dispatched from terminals according to train schedules (timetables),which may be represented on a distance-time graph. Line-haul movements begin from a departure yard and terminate at another terminal. Upon the arrival of train at a destination, cars having different destinations are reclassified into blocks and again grouped into other trains according to the same network-wide train makeup plan. The subsequent blocks are again grouped into other trains according to the train makeup plan.

It is well known that, operating plans are developed for a medium-term (tactical) planning horizon, and must be updated as traffic conditions and customers' need changes. As noted above, operating plans include not only train blocking and makeup plan, but also train schedules (timetables). Therefore, timetables should be jointly developed along with train connections and frequencies to better utilize the railway resources and meet the requirement of shippers. In practice, however, combination of train blocking plan/train makeup plan and train schedules is very difficult due to the large scale of real-world systems. Consequently, railway companies have followed a sequential process: at first, blocking and train formation are determined, and then timetables are designed around these connections. The operating plans should provide the most efficient way of moving all required traffic, subject to customer service requirements, and technological constraints on train and yard capacity.

A wide range of studies have been devoted to solve the train blocking problem during the last decades. Existing train blocking models can be categorized to 0-1 programming, to name a few, Cao and Zhu (1992), Cao and Zhu (1993a,b), Yang et al. (2002), non-linear programming, namely, Lin and Zhu (1994a,b), Lin et al. (1995) and Multi Agent System(MAS) based method (Bocker et al., 2001). At the same time, lots of studies have been focused on the train scheduling problem in the last three decades, particularly regarding the development of mathematical formulations and solution algorithms. Szpigel (1973) first modeled the train scheduling problem as a mixed integer program. Jovanovic and Harker (1991) presented a nonlinear integer programming model that minimizes the deviation between planned schedules and actual schedules, and used a branch-and-bound procedure to generate feasible meet-pass train plans. Carey (1994a,b) and Carey and Lockwood (1995) developed an iterative decomposition approach for solving the train scheduling and pathing problem in a rail network with one-way and two-way tracks. Higgins et al. (1996) developed a nonlinear integer programming model for minimizing the total train delay and fuel cost with variable train velocities. Brannlund et al. (1998) proposed an integer programming model to find a profit-maximizing timetable and designed a Lagrangian relaxation method to dualize track capacity constraints. Based on a graph-theoretic formulation for the periodic-scheduling problem in a rail line, Caprara et al. (2002) presented a Lagrangian relaxation solution method. Kroon and Peeters (2003) developed a periodic event scheduling model to consider variable trip times for the cyclic railway scheduling problem. Zhou and Zhong (2005) proposed a multi-mode resource-constrained project scheduling formulation to consider acceleration and deceleration time losses in double-track train scheduling applications, and designed several dominance rules to find Pareto optimum schedules in an intercity passenger corridor. The train scheduling problem is known to be NP-hard (Cai et al., 1998; Caprara et al., 2002), and optimal solutions are typically unattainable in large-scale and complex instances. To meet the computational requirements in real-world applications, a heuristic approach becomes a necessity for generating feasible and suboptimal solutions.

Although there are a number of existing approaches to solve train blocking and sharing problem, train scheduling problem respectively, literature on solving the combination of train blocking and scheduling problem is scarce. Dyke(2006) presented a multi model applied systems to solve such problems in the context of North America railroad. However there are many differences on the freight transportation mode between North America railroad and Chinese railway, thus those system are not quite suitable for solving the train blocking and scheduling problem in China. Li and Yang (2007) presented a MAS-based approach to combine the train blocking and scheduling problem and corresponding algorithms, aiming at the characteristics of Chinese railway. In their study, they discussed the definitions of various types of agents, interacting mechanism between agents and so on.

Train blocking and makeup models produce a transportation plan that completely describes the routing of freight, the set of trains to be and their respective frequency. As these models do not take train scheduling into consideration, it may be difficult to later find a timetable accommodating all planned trains and satisfying line and yard capacity. Thus, the requirements of shippers could not be met very well. Hence, compound models, which address both the train blocking and the scheduling aspects of freight transportation, can significantly help to improve service reliability and also reduce costs.

This paper is motivated by the need for combining train blocking and scheduling with flexible time window to respond to the requirement of freight shippers for high-level reliability of freight transportation. The challenge is to effectively account for the seamless combination of blocking train units and scheduling trains. The paper on hand provides a bi-level programming model which covers both train blocking problem and train scheduling problem. For this purpose, the reminder of the paper is structured as follows. Section 2 formalizes the problem as a mathematical program and states the complexity, then section 3 presents a heuristic(parallel genetic algorithm) solution procedure, which is evaluated in two comparative computational studies(section 4). Finally, section 5 summarizes the paper and specifies the future research challenges.

2. Mathematical Formulations

2.1. Rail Network Representation

We use a graph $G=(V,A)$ to represent a railway network. V is the set of vertices, $V=\{v_1,v_2,\ldots v_p\}$, p is number of vertices. Each vertex represents one yard on the network. A is the set of arcs, $A=\{a_1,a_2,\ldots a_q\}$, q is the number of arcs. a_{ij} represents the arc from v_i to v_j, $a_{ij}=(v_i,v_j)$, $a_{ij}\in A$, $v_i\in V$, $v_j\in V$, and $(v_i,v_j)\neq(v_j,v_i)$. w_{ij} is the weight of a_{ij}.

2.2. Notation

Table 1: Subscripts and parameters used in mathematical formulations

Symbol	Description
N_{mn}^k	= car block, including Q, $t_S^{'}$, $t_S^{''}$, $t_E^{'}$, $t_E^{''}$, P_{mn}.
Q	= number of cars in N_{mn}^k
$t_S^{'}$	= the expected earliest departure time for N_{mn}^k
$t_S^{''}$	= the expected latest departure time for N_{mn}^k
$t_E^{'}$	= the expected earliest arrival time for N_{mn}^k
$t_E^{''}$	= the expected latest arrival time for N_{mn}^k
P_{mn}	= serial of arcs in the path from v_m to v_n
C_{ij}	= capacity of the section a_{ij}, maximum number of trains that can be served in the section over some period time(i.e. one day)
M_{v_i}	= capacity of the yard v_i, maximum number of cars that can be marshaled in the yard over some period time (i.e. one day)
m_f	= maximum number of cars that can be grouped into one single train
T	= set of trains
d_i^{ll}	= the earliest departure time for train i at section l
d_i^{lu}	= the latest departure time for train i at section l
a_i^{ll}	= the earliest arrival time for train i at section l
a_i^{lu}	= the latest arrival time for train i at section l
d_i^l	= the earliest departure time for train i at the origin yard
d_i^u	= the latest departure time for train i at the origin yard
a_i^l	= the earliest arrival time for train i at the destination yard
a_i^u	= the latest arrival time for train i at the destination yard
t_{ij}^S	= start time of capacity loss time window for maintenance reason
t_{ij}^E	= end time of capacity loss time window for maintenance reason
v_i^B	= origin yard of train i
v_i^E	= destination yard of train i
Y_{mn}^k	= serial of trains that car blocks N_{mn}^k belongs to

B_{v_i}	= realized marshalling count of yard v_i
$M_{mn}^{k,i}$	= binary variable, 1 if N_{mn}^{k} is marshaled in yard v_i, 0 otherwise
$P_e^{a_{ij}}$	= binary variable, 1 if train e passes through arc a_{ij}, 0 otherwise
l_{mn}	= length of all sections that trains in Y_{mn}^{k} pass through
S_{mn}^{k}	= revenue to railway company from N_{mn}^{k}
$Z_{mn}^{t,i}$	= cost to railway company for operating train i from v_m to v_n at time section t
L_{mn}^{k}	= convenience to shippers from N_{mn}^{k}
C_{mn}^{k}	= cost to shippers for N_{mn}^{k}
$Z_{mn}^{m_c,i}$	= punishment cost for the less realized grouped cars in train i with count m_c
$Z_{v_i}^{t}$	= average required marshalling time for yard v_i
ω_1	= weight of railway company revenue
ω_2	= weight of shippers convenience

Table 2: Decision variables used in mathematical formulations

Symbol	Description
$x_{mn}^{t,i}$	= binary variable, 1 if allocate one train (i) from v_m to v_n at time section t, 0 otherwise
$y_{mn}^{k,i}$	= binary variable, 1 if train i contains N_{mn}^{k} ,0 otherwise

2.3. Mathematical Formulations

2.3.1. Upper-level Planning Formulations

The upper-level planning can be formulated as below.

Maximize $\omega_1 R + \omega_2 S$ (1)

$$R = (\sum_{m=1}^{P}\sum_{n=1}^{P}\sum_{k=1}^{K} S_{mn}^{k} - \sum_{i\in T} Z_{mn}^{t,i} - \sum_{i\in T} Z_{mn}^{m_c,i} - \sum_{v_i\in V} B_{v_i} Z_{v_i}^{c}) \quad (2)$$

$$S = \sum_{m=1}^{p}\sum_{n=1}^{p}\sum_{k=1}^{K} N_{mn}^{k}(Q) L_{mn}^{k} \quad (3)$$

Subject to,

(1) Section capacity constraints

$$\sum_{e\in T} P_e^{a_{ij}} \le \mathrm{C}_{ij}, a_{ij} \in A \quad (4)$$

(2) Capacity loss constraints for maintenance reason, see Figure 1.

$$\begin{cases} t_{ij}^{E} \le d_i^{ll} \le 24, t_{ij}^{E} \le a_i^{lu} \le 24 \\ t_{ij}^{E} \le d_i^{ll} \le 24, 0 \le a_i^{lu} \le t_{ij}^{S} \\ 0 \le d_i^{ll} \le t_{ij}^{S}, 0 \le a_i^{lu} \le t_{ij}^{S} \end{cases} \quad (5)$$

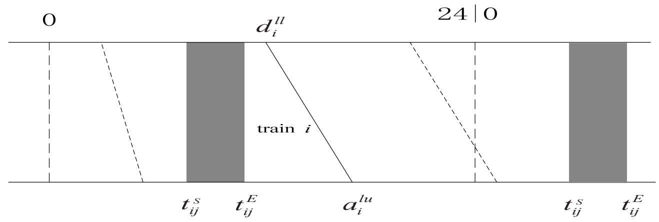


Figure 1: Illustration of capacity loss for maintenance reason

R in equation(1) represents the revenue of railway company and *S* in equation(1) represents the convenience of shippers. In the paper, we assume that a linear combination of *R* and *S* with respective weight ω_1 and ω_2 is maximized. $N_{mn}^{k}(Q)$ is the volume of N_{mn}^{k}.

2.3.2. Lower-level planning formulations

The lower-level planning can be formulated as:

$$\text{Minimize } C = \sum_{m=1}^{P}\sum_{n=1}^{P}\sum_{k=1}^{K} C_{mn}^{k} \tag{6}$$

C represents the cost of shippers.

Subject to,

(1) Yard capacity constraints

$$\sum_{m=1}^{p}\sum_{n=1}^{p}\sum_{k=1}^{K} M_{mn}^{k,i} N_{mn}^{k}(Q) \le \mathrm{M}_{v_i}, v_i \in V \tag{7}$$

(2) Physical limits to the number of cars that can be grouped into a single train

$$\sum_{m=1}^{p}\sum_{n=1}^{p}\sum_{k=1}^{K} y_{mn}^{k,i} N_{mn}^{k,i}(Q) \le \mathrm{m}_{f}, i \in T \tag{8}$$

$N_{mn}^{k,i}(Q)$ represents the count of cars marshaled into train *i of* N_{mn}^{k}.

(3) OD transfer constraints, For each Y_{mn}^{k},

$$\begin{cases} p \ge 1 \\ v_i^{B} = v_m, i = 1 \\ v_{i-1}^{E} = v_i^{B}, p \ge i \ge 2 \\ v_i^{E} = v_n, i = p \end{cases} \tag{9}$$

These constraints are considered to ensure each car block is marshaled into trains finally.

(4) Train transfer headway constraints, for any Y_{mn}^{k}

$$a_i^{u} + Z_{v_i^{E}}^{t} \le d_{i+1}^{l}, i=1...p-1 \tag{10}$$

(5) Flow balance constraints

$$O_I + \sum_{m=1}^{p}\sum_{n=1}^{p}\sum_{k=1}^{K}\sum_{\substack{i\in T,且 \\ v_i^{E}=I}} y_{mn}^{k,i} N_{mn}^{k}(Q) = \sum_{m=1}^{p}\sum_{n=1}^{p}\sum_{k=1}^{K}\sum_{\substack{i\in T,且 \\ v_i^{B}=I}} y_{mn}^{k,i} N_{mn}^{k}(Q) + D_I \tag{11}$$

O_I represents the departure train units on yard *I* and D_I the arrival train units.

3. Parallel Genetic Algorithm

By considering the complexity of the above model, normal algorithms are too slow or not capable enough to come up with good enough solutions. Thus, heuristic algorithms are needed to solve instance problems of real-world size. Although there are quite a few meta-heuristic methods that could be used to solve NP problems, genetic algorithm has its own advantages. Genetic algorithm is a search technique used in computing to find exact or approximate solutions to optimization and search problems. The major advantage of genetic algorithms is their flexibility and robustness as a global search method. In addition, genetic algorithms are parallel because they have multiple offspring thus making it ideal for large problems where evaluation of all possible solutions in serial would be too time taking. Furthermore, they perform well in problems where the fitness landscape is complex, where the fitness function is discontinuous, noisy, changes over time or has many local optima. In this study, we try to design and implement a specific genetic algorithm based on analyzing the characteristics of the problem to solve the above model. The table below shows the parameters used in the genetic algorithm.

Table 3: Parameters used in the genetic algorithm solution procedure

Symbol	Description
POP_SIZE_U	= Population size for the upper-level model
POP_SIZE_L	= Population size for the lower-level model
Gen_Num_U	= Iteration number for the upper-level model
Gen_Num_L	= Iteration number for the lower-level model
P_c_U	= Probability of crossing for the upper-level model
P_c_L	= Probability of crossing for the lower-level model
P_m_U	= Probability of mutation for the upper-level model
P_m_L	= Probability of mutation for the lower-level model

Solution procedure can be described as below.

Step 1: Representation and initialization of chromosomes for the upper-level model
Select a feasible solution N_0 for the upper-level model, let $X^r = \{x_{mn}^{t,i}\}$ represent a chromosome for the upper-level model. Each code in the chromosome is the value of $x_{mn}^{t,i}$, generate POP_SIZE_U chromosomes, that is $\{X^r / X^r = (x_1^r, x_2^r \ldots\ldots x_n^r), r = 1, 2, \ldots\ldots POP\_SIZE\_U\}$. The steps used to generate chromosomes could be described as below.
(1) Select a random direction d from R_n;
(2) If N_0+d is feasible, let it be a new chromosome, otherwise, generate a new d , until N_0+d is feasible;
(3) Repeat steps (1) and (2) to get POP_SIZE_U chromosomes.

Step 2: Evaluation and Selection
Execute each chromosome to get each chromosome's fit value based on step6, sort the chromosomes on the basis of fit values. Select POP_SIZE_U chromosomes to execute the following steps.

Step 3: Crossover
Create a new population of POP_SIZE_U number by applying the following operations. The operations are applied to choose from the population with a probability based on fitness.
(1) Darwinian Reproduction: Reproduce an existing chromosome by copying it into the new chromosome.
(2) Create two new chromosomes from two existing chromosomes by genetically recombining randomly chosen parts of two existing chromosome s using the crossover operation applied at a randomly(according to P_c_U) chosen crossover point within each chromosome.

Step 4: Mutation
Create a new population of POP_SIZE_U number by applying the following operations. The operations

are applied to choose from the population with a probability based on fitness.

(1) Darwinian Reproduction: Reproduce an existing chromosome by copying it into the new chromosome.
(2) Create one new chromosome from one existing chromosome by mutating a randomly (according to P_m_U) chosen part of the chromosome.

Step 5: Iterations

Iteratively perform the above steps (2)~(4) until the termination criterion Gen_Num_U has been satisfied.

Step 6: Solving the lower-level model based on the input from the upper-lower model

Step 6.1 Representation and initialization of chromosomes for the lower-level model

Select a feasible solution N_0 for the upper-level model, let $X_L{}^r = \{y_{mn}^{k,i}\}$ represent a chromosome for the lower-level model. Each code in the chromosome is the value of $y_{mn}^{k,i}$, generate POP_SIZE_L chromosomes, that is $\{X^r / X^r = (x_1^r, x_2^r......x_n^r), r = 1, 2, ...POP\_SIZE\_L\}$. The steps used to generate chromosomes could be described as below.

(1) Select a random direction d from R_n;
(2) If N_0+d is feasible, let it be a new chromosome, otherwise, generate a new d, until N_0+d is feasible;
(3) Repeat steps (1) and (2) to get POP_SIZE_L chromosomes.

Step 6.2: Evaluation and Selection The same with step2.

Step 6.3: Crossover The same with step3.

Step 6.4: Mutation The same with step4.

Step 6.5: Iterations The same with step5.

4. Numerical Experiments

4.1. Experimental data

4.1.1. Rail network and weights of arcs

Figure 2 depicts a rail network for experiments which consists of 11 yards and 16 weighted sections.

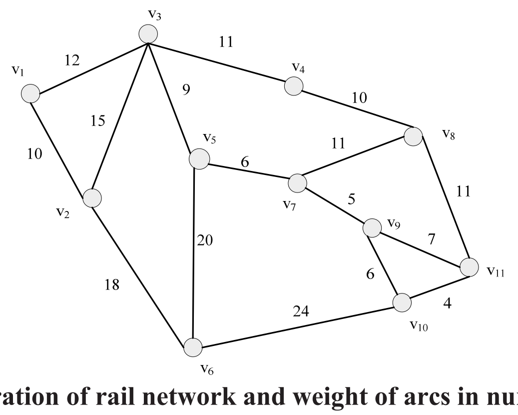


Figure 2: Illustration of rail network and weight of arcs in numerical experiments

4.1.2. Path Matrix

We assume that we already know the path from any node to another node before train blocking and scheduling. Without loss of generality, we use the shortest path according to the weight (see Table 4).

Table 4 Path matrix

v_i	1	2	3	4	5	6	7	8	9	10	11
1		(v_1,v_2)	(v_1,v_3)	(v_1,v_3,v_4)	(v_1,v_3,v_5)	(v_1,v_2,v_6)	(v_1,v_3,v_5,v_7)	(v_1,v_3,v_4,v_8)	(v_1,v_3,v_5,v_7,v_9)	(v_1,v_2,v_6,v_{10})	(v_1,v_3,v_4,v_8,v_{11})
2	(v_2,v_1)		(v_2,v_3)	(v_2,v_3,v_4)	(v_2,v_3,v_5)	(v_2,v_6)	(v_2,v_3,v_5,v_7)	(v_2,v_3,v_4,v_8)	(v_2,v_3,v_5,v_7,v_9)	(v_2,v_6,v_{10})	(v_2,v_6,v_{10},v_{11})
3	(v_3,v_1)	(v_3,v_2)		(v_3,v_4)	(v_3,v_5)	(v_3,v_5,v_6)	(v_3,v_5,v_7)	(v_3,v_4,v_8)	(v_3,v_5,v_7,v_9)	(v_3,v_5,v_7,v_9,v_{10})	(v_3,v_4,v_8,v_{11})
4	(v_4,v_3,v_1)	(v_4,v_3,v_2)	(v_4,v_3)		(v_4,v_3,v_5)	(v_4,v_3,v_5,v_6)	(v_4,v_8,v_7)	(v_4,v_8)	(v_4,v_8,v_7,v_9)	(v_4,v_8,v_{11},v_{10})	(v_4,v_8,v_{11})
5	(v_5,v_3,v_1)	(v_5,v_3,v_2)	(v_5,v_3)	(v_5,v_3,v_4)		(v_5,v_6)	(v_5,v_7)	(v_5,v_7,v_8)	(v_5,v_7,v_9)	(v_5,v_7,v_9,v_{10})	(v_5,v_7,v_9,v_{11})
6	(v_6,v_2,v_1)	(v_6,v_2)	(v_6,v_5,v_3)	(v_6,v_5,v_3,v_4)	(v_6,v_5)		(v_6,v_5,v_7)	(v_6,v_5,v_7,v_8)	(v_6,v_{10},v_9)	(v_6,v_{10})	(v_6,v_{10},v_{11})
7	(v_7,v_5,v_3,v_1)	(v_7,v_5,v_3,v_2)	(v_7,v_5,v_3)	(v_7,v_8,v_4)	(v_7,v_5)	(v_7,v_5,v_6)		(v_7,v_8)	(v_7,v_9)	(v_7,v_9,v_{10})	(v_7,v_9,v_{11})
8	(v_8,v_4,v_3,v_1)	(v_8,v_4,v_3,v_2)	(v_8,v_4,v_3)	(v_8,v_4)	(v_8,v_7,v_5)	(v_8,v_7,v_5,v_6)	(v_8,v_7)		(v_8,v_7,v_9)	(v_8,v_{11},v_{10})	(v_8,v_{11})
9	(v_9,v_7,v_5,v_3,v_1)	(v_9,v_7,v_5,v_3,v_2)	(v_9,v_7,v_5,v_3)	(v_9,v_7,v_8,v_4)	(v_9,v_7,v_5)	(v_9,v_{10},v_6)	(v_9,v_7)	(v_9,v_7,v_8)		(v_9,v_{10})	(v_9,v_{11})
10	(v_{10},v_6,v_2,v_1)	(v_{10},v_6,v_2)	(v_{10},v_9,v_7,v_5,v_3)	(v_{10},v_{11},v_8,v_4)	(v_{10},v_9,v_7,v_5)	(v_{10},v_6)	(v_{10},v_9,v_7)	(v_{10},v_{11},v_8)	(v_{10},v_9)		(v_{10},v_{11})
11	(v_{11},v_8,v_4,v_3,v_1)	(v_{11},v_{10},v_6,v_2)	(v_{11},v_8,v_4,v_3)	(v_{11},v_8,v_4)	(v_{11},v_9,v_7,v_5)	(v_{11},v_{10},v_6)	(v_{11},v_9,v_7)	(v_{11},v_8)	(v_{11},v_9)	(v_{11},v_{10})	

4.1.3. *Cars origin-destination matrix*

Table 5 shows the need for freight transportation of shippers in the way of cars origin-destination matrix. Each O-D unit represents one N_{mn}^{k} .

Table 5 Cars origin-destination matrix

v_i	1	2	3	4	5	6	7	8	9	10	11
1		(110,8,15,8,13) (200,0,24,0,24)	(80,0,24,0,24)	(100,6,10,18,22)	(90,14,16,8,10)	(100,8,10,5,6)	0	(100,0,24,6,7)	(130,0,24,16,24)	0	(138,23,24,16,23)
2	(240,6,8,0,24)		(100,0,24,0,24)	0	(290,13,15,20,22)	(78,0,24,0,24)	(170,0,24,0,24)	(95,0,23,0,24)	(156,15,17,0,24)	(210,15,19,13,20)	(310,1,3,6,8)
3	(99,8,12,13,17)	(165,8,9,0,24)		0	(230,0,12,7,8)	(142,23,24,0,6)	(156,12,18,14,18)	(100,2,10,15,17)	(160,9,15,8,14)	(97,5,9,10,18)	(130,0,24,0,24)
4	(145,0,24,0,24)	(100,8,24,0,24)	(300,0,10,0,24)		(110,0,24,0,24)	(110,8,24,7,16)	(163,0,24,0,24)	(130,0,6,0,6)	(140,12,14,8,18)	(160,0,24,0,24)	(99,0,24,0,24)
5	(200,0,24,0,24)	(165,0,24,0,24)	(200,14,19,2,5)	(93,15,19,17,20)		(120,0,24,0,24)	(260,8,10,17,20)	(160,8,14,16,20)	(91,0,24,0,24)	(197,8,10,17,20)	(165,0,24,0,24)
6	(140,9,12,18,24)	(100,0,24,0,24)	(100,6,10,18,20)	(150,0,24,0,24)	(236,9,19,17,23)		(130,8,10,19,20)	(150,0,24,0,24) (150,8,10,17,20)	(187,8,10,17,20)	(95,0,24,0,24)	(129,0,24,0,24)
7	0	(92,0,24,0,24)	(110,0,24,0,24)	(245,7,24,15,24)	(75,17,24,12,24)	0		(139,0,24,0,24)	(98,0,24,0,24)	(68,5,9,11,24)	(210,13,24,15,24)
8	(240,0,24,0,24)	0	(187,0,17,14,24)	(100,7,15,6,24)	0	(190,0,24,0,24) (152,0,24,0,23)	(100,13,22,0,24)		(164,0,24,0,24)	(100,0,24,3,24)	(110,0,24,0,24)
9	(100,15,19,13,17)	(159,0,24,0,24)	0	(100,11,24,0,24)	(143,0,24,0,24)	(100,0,24,0,24)	(230,4,8,9,22)	(145,0,24,0,24)		(100,8,11,7,23)	(97,7,12,13,19)
10	(100,0,24,0,24)	(180,11,24,16,24)	(89,0,24,0,24)	0	(320,0,13,11,23)	(234,0,24,0,24)	(200,4,16,8,10)	(110,0,24,0,24)	(100,0,24,0,24)		(135,0,24,0,24)
11	0	(87,0,24,0,24)	(100,0,24,0,24)	(280,15,24,17,22)	(96,8,24,11,24)	(100,0,24,0,24)	(100,12,24,12,22)	(90,0,15,7,12)	(95,0,24,0,24)	(93,0,15,0,20)	

4.1.4. Parameters Settings for Experiments

We set m_f = 55, M_{v_i} = 1000 for each yard, C_{ij} = 80 for each section, ω_1=0.5 and ω_2=0.5 for both experiment A and B. Table 6 proposes the value of $Z_{v_i}^t$ for each yard in experiments.

Table 6 Marshalling time of each yard

	1	2	3	4	5	6	7	8	9	10	11
$Z_{v_i}^t$	4	4	5	3	2	4	3	6	4	3	4

The method to realize S_{mn}^k and $Z_{mn}^{t,i}$ for experiments can be described as (12),(13) and (14).

$$S_{mn}^k = N_{mn}^k(Q) \bullet l_{mn} \quad (12)$$

$$Z_{mn}^{t,i} = \mu(l_{mn}) = l_{mn} \quad (13)$$

$$l_{mn} = \sum_{a_{ij} \in p_{mn}} w_{ij} \quad (14)$$

4.2. Experiment A

In experiment A , we assume that the rail company itself afford the punishment cost $Z_{mn}^{m_c,i}$.

$$C_{mn}^k = S_{mn}^k + \upsilon(N_{mn}^k(Q)) \bullet \Delta_t \quad (15)$$

$$Z_{mn}^{m_c,i} = l_{mn} \bullet m_c \quad (16)$$

4.2.1. Parameter settings for genetic algorithm in experiment A

The value of parameters for the upper-level model are POP_SIZE_U = 20, P_m_U =0.5, P_c_U =0.2, Gen_Num_U =100. Regarding the lower-level model , the parameters value are POP_SIZE_L=15, P_m_L =0.3, P_c_L =0.6, Gen_Num_L =100.

4.2.2. Model Analysis

The number of variables is

$$C_{11}^2 \times 2 + 102 \times C_{11}^2 \times 2 = 110 + 110 * 102 = 11330. \quad (17)$$

The number of constraints is

$$11 + 16 + 110 + 102 + 16 \times 110 \times 2 + 11 = 3770. \quad (18)$$

4.2.3. Results

The genetic algorithm has been implemented by Microsoft Visual C++ 6.0(more than 7000 code lines) and runs on a Pentium D, 3.4GHz PC, with 512MB RAM memory. The time cost is 187 seconds. The results are given below.

(1) The total weighted revenue is 39640.14, the revenue of railway company is 41470.00, the convenience of shippers is 37810.27, and the cost of shippers is 2609.00.
(2) Railway company totally operates 369 trains, the average utilization rate of section capacity is 40% and the average utilization rate of yard capacity is 63%.

Partial train departure time window and the number of cars that grouped into the train are shown in table A1, table A2 shows the train marshalling scheme.

4.3. Experiment B

In this experiment, we use the same parameters with experiment A, and we assume that the rail company and shippers jointly afford the punishment cost $Z_{mn}^{m_c,i}$.

$$C_{mn}^{k} = S_{mn}^{k} + \upsilon(N_{mn}^{k}(Q)) \bullet \Delta_t + \sum_{i=1}^{p} c_i \tag{19}$$

$$Z_{mn}^{m_c,i} = l_{mn} \bullet m_c - \sum_{m=1}^{p}\sum_{n=1}^{p}\sum_{k=1}^{K} y_{mn}^{k,i} c_i \tag{20}$$

$$c_i = (m_f - \sum_{m=1}^{p}\sum_{n=1}^{p}\sum_{k=1}^{K} y_{mn}^{k,i} N_{mn}^{k,i}(Q)) / N_i^c \tag{21}$$

(1) The total weighted revenue is 39209.05, the revenue of railway company is 46455.00, the convenience of shippers is 31963.10, and the cost of shippers is 4088.00.
(2) 363 trains are operated by railway company, the average utilization rate of section capacity is 39% and the average utilization rate of yard capacity is 66%.

Partial train departure time window and the number of cars that grouped into the train are shown in table A3,Table A4 shows the train marshalling scheme.

4.4. Comparative Result Analysis

We presented the below comparative analysis according to the results from experiment A and B. Figure 3 shows that in experiment B, the revenue of railway company and the cost of shippers rise significantly with the reduction of the convenience of shippers. There is no obvious difference between experiment A and experiment B on the yard capacity utilization and section capacity utilization which can be made on the basis of Figure 4 and Figure 5. Figure 6 shows the number of cars that trains contain in experiment A and B.

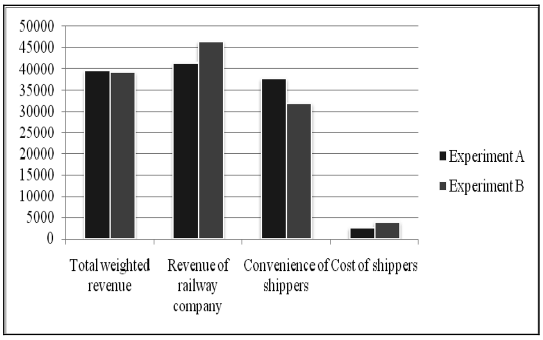


Figure 3: Comparison analysis of revenue and cost

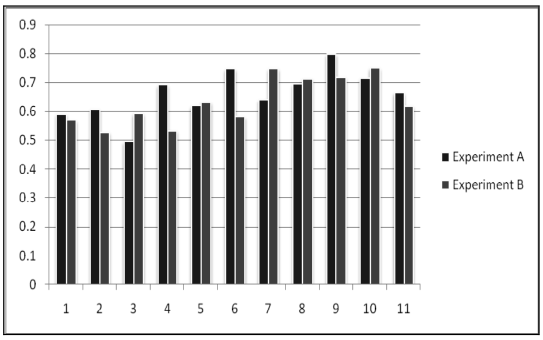


Figure 4: Comparison analysis of yard capacity utilization

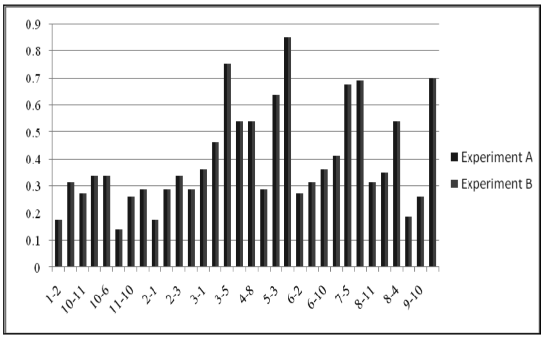


Figure 5: Comparison anaylisis of section capacity utilization

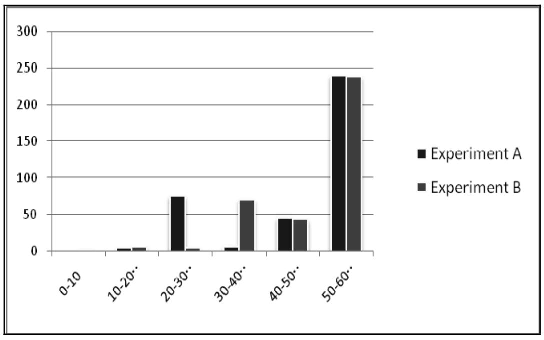


Figure 6: Comparison anaylisis of the number of cars that trains contain

5. Conclusions and Future Research

This paper introduced a bi-level mathematical programming model to solve the train blocking and scheduling combination problem . The upper level planning stage aims to maximize the weighted revenue which depends on the number of trains operated by railway company and the number of cars that marshaled in yards.

The lower level is designed to minimize the cost of shippers which is mainly related to the solution that cars are grouped into trains. The main benefit is the provision of a heuristic solution that is ideal for real-world size. Although the proposed methodology is sufficient for real-world problem, it does not guarantee an optimal solution, because of the heuristic property. As a result, further research is necessary to develop an solution methodology that can get the optimal solution and can also meet the requirements of solving real-world problem.

6. Acknowledgements

The work described in this paper was jointly supported by The Foundation of Beijing Jiaotong University (No. 2004SZ004) and The National High Technology Research and Development Program of China(863 Program)(No. 2007AA11Z136). We would like to give our sincere thanks to them! The authors are responsible for all the results and opinions expressed in this paper.

References

Bocker, J., Lind, J. and Zirkler, B., (2001), Using a multi-agent approach to optimize the train blocking and sharing system. European Journal of Operational Research, 143(2001), 242-252

Brannlund, U., Lindberg, P.O., Nou, A., and Nilsson, J.E. (1998), Railway timetabling using Lagrangian relaxation. Transportation Science 32(4), 358–369.

Cai, X., Goh, C.J., and Mees, A.I. (1998), Greedy heuristics for rapid scheduling of trains on a single track. IIE Transactions 30 (5), 481–493.

Cao, J. and Zhu, S., (1992), Optimizing train blocking plan, 0-1 programming[J]. Journal of the China Railway Society. 14(4).49-57

Cao, J. and Zhu, S., (1993a), The 0-1 quadratic programming method for the single group train formation plan 0n railway network [J]. Journal of the China Railway Society.15(2). 62-70

Cao, J. and Zhu, S., (1993b), Synthetic optimization of train formation plan on railway network. Journal of Southwest Jiaotong University.5

Caprara, A., Fischetti, M. and Toth, P., (2002), Modeling and solving the train timetabling problem. Operations Research 50 (5), 851–861.

Carey, M., (1994a), A model and strategy for train pathing with choice of lines, platforms and routes. Transportation Research Part B 28 (5), 333–353.

Carey, M., (1994b), Extending a train pathing model from one-way to two-way track. Transportation Research Part B 28 (5), 395–400.

Carey, M. and Lockwood, D., (1995), A model, algorithms and strategy for train pathing. Journal of the Operational Research Society 46 (8),988–1005.

Carl van Dyke, (2006), Strategic Policies of Multi Model Applied Systems. MULTIRAIL[C].

Li, H. and Yang, H., (2007), Study on The Theory and Methodologies of Synchronizing Train Blocking and Scheduling. Beijing Jiaotong University.

Higgins, A ., Kozan, E. and Ferreira, L., (1996), Optimal Scheduling of Trains on a Single Line Track. Transportation Research Part B. 30.147-161

Jovanovic, D. and Harker, P.T., (1991), Tactical Scheduling of rail operations: the SCAN I system. Transportation Science. 25(1), 46–64

Kroon, L.G. and Peeters, L.W.P., (2003), A variable trip time model for cyclic railway timetabling. Transportation Science 37 (2), 198–212.

Lin, B. and Zhu, S., (1994a). The nonlinear 0-1 programming model and simulated annealing algorithm for optimizing the train formation plan. Journal of the China Railway Society. 16(2). 61-66

Lin, B. and Zhu, S., (1994b).The optimization method for the train formation plan with yard capacity limit[J].Journal of the China Railway Society.29(5).488-493

Lin, B. and Zhu, S. et al., (1995), Study on Direct Train Formation Plan for Loading Area. China Railway Science.16(2).108-114

Szpigel, B., (1973), Optimal Train Scheduling on a Single Track Railway, in: M .Ross(Ed.). Operational Research 72.North-Holland, Amsterdam, 1973, 343-352

Yang, S., Shi, F. and Li, Z., (2002), An artificial neural network method for marshalling plan. Journal of Changsha Railway University. 20(3).79-84

Zhou, X. and Zhong, M., (2005), Bi-criteria train scheduling for high-speed passenger railway planning applications. European Journal of Operational Research 167 (3), 752–771.

Appendices

Table A1: Partial departure time window and marshalling volume in experiment A

train index	v_m	v_n	departure time window	marshalling volume	cost
13312	2	3	15:00—15:30	14	15.00
14126	6	7	18:30—19:00	15	26.00
11627	9	7	7:00—7:30	16	5.00
10859	9	10	3:30—4:00	17	6.00
10935	6	5	4:00—4:30	20	20.00
10624	8	4	2:30—3:00	20	10.00
11555	2	6	7:00—7:30	20	18.00
11341	3	1	6:00—6:30	20	12.00
…	…	…	…	…	…

Table A2: Partial train marshalling scheme in experiment A

train units index	volume	v_m	v_n	train serial	transport cost
38	10	5	3	10043	9.00
39	13	5	4	10043 -> 12113	18.00
64	18	7	9	10068	5.00
66	53	7	11	10070	12.00
69	20	8	4	10074	10.00
72	17	8	6	10076	37.00
71	25	8	6	10076	37.00
88	50	10	2	10092	42.00
…	…	…	…	…	…

Table A3: Partial train departure time window and marshalling volume in experiment B

Train index	v_m	v_n	departure time window	marshalling volume	cost
12223	3	4	10:00—10:30	10	11.00
12567	4	8	11:30—12:00	10	10.00
13471	7	1	15:30—16:00	15	27.00
11320	11	10	5:30—6:00	15	4.00
10989	11	9	4:00—4:30	15	7.00
15157	10	7	23:00—23:30	15	11.00
10442	1	3	2:00—2:30	15	12.00
13472	7	2	15:30—16:00	19	30.00
…	…	…	…	…	…

Table A4: Partial train marshalling scheme in experiment B

train units index	volume	v_m	v_n	train serial	transport cost
17	10	3	5	10244	11.00
10	10	2	8	11552 -> 12223 -> 12567	114.00
53	10	6	8	13025 -> 10156 -> 11497	48.00
50	10	6	4	13025 -> 10923 -> 10683	70.00
40	10	5	6	13675	26.00
58	11	7	2	11165 -> 11032	33.00
64	13	7	9	11168	7.00
37	15	1	4	10442 -> 10683	56.00
…	…	…	…	…	…

Sequencing minimum product set on mixed model U-lines to minimize work overload

Jinlin Li, Jie Gao and Linyan Sun

School of Management, The State Key Laboratory for Manufacturing Systems Engineering,
Xi'an Jiao Tong University, Xi'an 710049, China.
E-mail: gaoj@mail.xjtu.edu.cn (J. Gao)
E-mail: lysun@mail.xjtu.edu.cn (L. Sun)

Abstract

Mixed model U-shaped line (MMUL) is widely used in JIT production systems. In MMUL, products need to be processed in both side of a station. As a result, the job sequences in different stations might be different and be differ from the sequence in which products are launched. This paper discusses the model sequencing problem on MMUL and explicitly considers the objective of minimizing the total overload in steady state. Some characteristics of the problem are identified and a mathematical model is formulated. At last, a branch-and-trim algorithm is proposed to solve the problem optimally.

Keywords: mixed model U-line, minimum product set, overload, sequencing

1. Introduction

In mixed model production system, different products (models) with similar production characteristics can be produced in the same line with neglectable setup cost. Managers usually need to decide the order in which the products should be processed in an existing balanced assembly line. To help make this decision, lots of researches have been done in the model sequencing topic in mixed model assembly line (MMAL). Generally speaking, two main sequencing objectives are considered (Boysen et.al, 2007) in the literature:

- Leveling part usage. This objective is related to JIT production philosophy, which tries to get constant material requirements for all the parts used in the line. This goal have been studied by Miltenburg (1989), Inman and Bulfin (1991).
- Minimizing work overload. This objective is usually related to paced line, closed stations and a material handling system running at a constant speed (such as conveyor). Since different model may require different processing time in each station, a sequence with consecutive work intensive models may induce overload. The unfavorable situation is usually handled by hiring utility worker, allowing line stoppage or accelerating the operations. As a result, extra cost and inefficiency occurs. Minimizing overload in paced line with closed station has been proved NP-hard in strong sense (Tsai, 1995). In single-station cases, some efficient optimal methods are developed (Yano and Rachamadugu, 1991; Tsai, 1995). Exact algorithms, such as branch-and-bound procedure (Zhao and Ohno, 2000) can be applied to solve small sized multi-station problems, but heuristics (Bolat and Yano, 1992; Bautista and Cano, 2008) and meta-heuristics (Kim et.al, 1996) are in more common use.

In this paper, we consider the objective of minimizing overload, particularly in U-shaped layout. Compared to straight line, U-line has several advantages, such as improved visibility and communication, possible reduction of stations, convenience for rebalance and adjustment. Lots of introduction and description about U-line can be found in books on the topic of JIT production. The balancing problem in single-model U-line was firstly discussed in Miltenburg and Wijingaard (1994). However, up to now, few researches on the topic of U-line balancing or sequencing problem can be found and a major of these investigated the single-model U-line. Mixed model U-line (MMUL) balancing problem was firstly studied by Sparling and Miltenburg (1998). After that, some researchers tried to use meta-heuristics to solve the MMUL balancing and sequencing problem simultaneously (Miltenburg, 2002; Kim et.al, 2006; Kara et.al, 2007). In all these papers, little attention is paid to the overload occurring during the process. In this paper, the minimization of the overload on MMUL is explicitly considered.

Besides, the sequencing problem studied in this paper is based on minimum product set (MPS) principle, which means the products are launched into the line in a cyclic sequence. (The MPS is the smallest product set having the same proportion as the total demand.) This production manner is also called cyclic production, which is widely used in industry and proved effective in smoothing the workload and getting constant production rate.

This paper focuses on minimizing the total overload in long run in the context of a paced MMUL. The contributions of this study include: 1) some characteristics of the MMUL sequencing problem are identified; 2) a mathematical formulation is built to model the problem; 3) the branch-and-trim method in Bolat (1997) is modified to solve the MMUL sequencing problem optimally.

The paper is organized as follows: in section 2, a detailed problem description is given and some characteristics of the problem is discussed, based on which a mathematical model is formulated; in section 3, an optimal branch-and-trim algorithm is proposed and applied to an example to prove its validity; finally, a summary and conclusion is given in section 4.

2. Mixed Model U-line Sequencing Problem

2.1. Problem Description

Consider a paced U-shaped production line, connected by a conveyor belt moving at a constant speed, as showed in Figure 1. The balanced line is made up of several stations with closed boundaries, which is represented by dashed line. These boundaries divide the U-line into many sections. In U-shaped line, some stations are composed of two disconnected sections (such as station 1, 4 in Figure 1), spanning both sides of the line, which called "crossover stations" and the others are the same as stations in straight line named "straight stations".

Several kinds of products are assembled simultaneously in the line in a cyclic production manner, i.e., products are launched into the belt in a basic sequence Π cyclically, for example, CACBCACBCACB..., Π=CACB. Products visit each section in the same order and are not removed from the belt until they reach the end of the line. Define the tasks required to perform to a product in a section as a job, e.g. job B_4^f represents the required tasks performed in front section of station 4 to a product from type B. The standard processing time of each job in each station is known and deterministic. Workers walk downstream inside the U-shaped line to do jobs according a non-preemptive first-come, first-served (FCFS) policy. When a worker finishes a job or reaches the downstream boundary of a section, he stops the job on hand and moves to the next job. If he reach the upstream boundary but the next job has not arrived, he waits there. Workers move much faster than the belt so the traveling time of each worker can be ignored. Sometime, overload occurs if a worker can't finish a job inside the section boundaries with standard operation time.

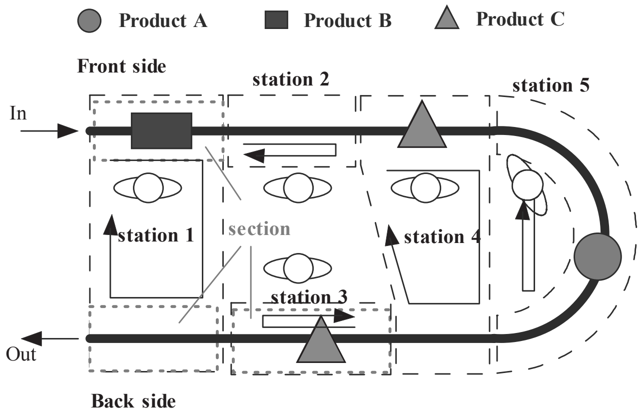


Figure 1: An illustration of mixed model U-line (MMUL)[1]

No loss of generality, assume the downstream moving speed of the conveyor belt is 1, and there is only one worker in each station. For the convenience of description, define the following notation:

Indices:

m	index for stations
n	index for product types
k	index for position in the product sequence
T	a parameter indicating which side a section belongs to, T=f/b represents front/back side respectively.

Parameters:

M	number of stations
N	number of product types
c	cycle time
d_n	number of product from type n in MPS
D	total number of products in MPS,
$t_{m,n}^{\mathrm{T}}$	required processing time for a product from type n in side T of station m
l_m^{T}	length of side T at station m
L_m	the distance from the upstream boundary of front section to the upstream boundary of back section in cross station m

Decision variables:

$x_{n,k}$	equal to 1 if a product from type n is assigned to kth position in the sequence, else equal to 0

Calculated decision variables:

Π	the basic product sequence according to which products are launched into the line
$\Pi(k)$	index of the product type in the kth position of Π, $\Pi(k) = \sum_{n=1}^{N} x_{n,k} n$.
$sp_{m,k}^{\mathrm{T}}$	position from the upstream boundary when the worker at station m start processing the jth product at side T.
$ep_{m,k}^{\mathrm{T}}$	position from the upstream boundary when the worker at station m stop processing the kth product at side T.
$u_{m,k}^{\mathrm{T}}$	overload (unfinished work) when kth product is processed at side T of station m.
$w_{m,k}^{\mathrm{T}}$	idle (waiting) time after kth product is processed at side T of station m
$U_m(\Pi)$	total overload at station m during processing a MPS for a given Π
$W_m(\Pi)$	total idle time at station m during processing a MPS for a given Π
$U(\Pi)$	total overload during processing a MPS for a given Π
$W(\Pi)$	total idle time during processing a MPS for a given Π

[1] Stations are numbered according the balancing procedure introduced in Miltenburg (1994), which guarantee the station m+1, …, M are placed between the front and back sections of crossover station m.

In crossover station m, a product firstly enters the front side to be processed and L_m time units later, it enter the back side, i.e., processing a product needs two separate jobs. Let $p_m=\lfloor L_m / c \rfloor$, $a_m= L_m-p_m c$ ($\lfloor x \rfloor$ is the largest integer not larger than x). Then the job $\Pi(k)_m^{\mathrm{b}}$ is released after the job $\Pi(k+p_m)_m^{\mathrm{f}}$ and before the job $\Pi(k+1+p_m)_m^{\mathrm{f}}$. Hence, the job sequence in crossover station m is $\Pi(p_m+1)_m^{\mathrm{f}}, \Pi(1)_m^{\mathrm{b}}, \Pi(p_m+2)_m^{\mathrm{f}}, \Pi(2)_m^{\mathrm{b}}, \ldots$. Jobs in front side and in back side alternate in the job sequence. Table 1 is an example, the basic sequence is CACB and the job sequence is $C_m^{\mathrm{f}} B_m^{\mathrm{b}} A_m^{\mathrm{f}} C_m^{\mathrm{b}} C_m^{\mathrm{f}} A_m^{\mathrm{b}} B_m^{\mathrm{f}} C_m^{\mathrm{b}}$. The jobs sequences in different stations might be different. This is quite different from the situation in straight line, where the job sequence in each station is the same with the sequence in which products are launched.

Table 1: the job sequence in a crossover station with p_m=3

product launching sequence		CACBCACB…
Generated job sequence (p_m=1)	Front side	C_m^{f} | A_m^{f} | C_m^{f} | B_m^{f} ...
	Back side	B_m^{b} | C_m^{b} | A_m^{b} | C_m^{b} ...

If all the jobs are started within the station boundaries, the worker has to walk clockwise or counter-clockwise to do jobs in two sides, as showed in Figure 1. This phenomenon is also described in Spariling and Milterburg (1998). However, if the sequence is not proper and the length of one side is much longer than the other side at the station, some jobs might be missed, i.e., when the worker finishes a job and turns around, the next job has left the section in other side. In fact, the worker ignores the missed job and move to the next job after the missed job. This fact can be treated as the worker firstly move to a dummy position out of the station, stay there 0 time and then move the next job after the missed job, since the worker's traveling time can be ignored. The end position for the missed job can be set to be same as the dummy start position. This treatment won't affect the amount of overload or idle time and the start positions for the next job. Hence, following relationships hold:

$$sp_{m,k}^{\mathrm{f}} = \max\{0, ep_{m,k-p_m-1}^{\mathrm{b}} + a_m - c\} \quad (1)$$

$$ep_{m,k}^{\mathrm{f}} = \min\{\max\{sp_{m,k}^{\mathrm{f}}, l_m^{\mathrm{f}}\}, sp_{m,k}^{\mathrm{f}} + t_{m,\Pi(k)}^{\mathrm{f}}\} \quad (2)$$

$$sp_{m,k}^{\mathrm{b}} = \max\{0, ep_{m,k+p_m}^{\mathrm{f}} - a_m\} \quad (3)$$

$$ep_{m,k}^{\mathrm{b}} = \min\{\max\{sp_{m,k}^{\mathrm{b}}, l_m^{\mathrm{b}}\}, sp_{m,k}^{\mathrm{b}} + t_{m,\Pi(k)}^{\mathrm{b}}\} \quad (4)$$

The overload and waiting time can be calculated using:

$$u_{m,k}^{\mathrm{f}} = \max\{0, \min\{t_{m,\Pi(k)}^{\mathrm{f}}, t_{m,\Pi(k)}^{\mathrm{f}} + sp_{m,k}^{\mathrm{f}} - l_m^{\mathrm{f}}\}\} = sp_{m,k}^{\mathrm{f}} - ep_{m,k}^{\mathrm{f}} + t_{m,\Pi(k)}^{\mathrm{f}} \quad (5)$$

$$u_{m,k}^{\mathrm{b}} = \max\{0, \min\{t_{m,\Pi(k)}^{\mathrm{b}}, t_{m,\Pi(k)}^{\mathrm{b}} + sp_{m,k}^{\mathrm{b}} - l_m^{\mathrm{b}}\}\} = sp_{m,k}^{\mathrm{b}} - ep_{m,k}^{\mathrm{b}} + t_{m,\Pi(k)}^{\mathrm{b}} \quad (6)$$

$$w_{m,k}^{\mathrm{f}} = \max\{0, a_m - ep_{m,k}^{\mathrm{f}}\} = sp_{m,k-p_m}^{\mathrm{b}} - ep_{m,k}^{\mathrm{f}} + a_m \quad (7)$$

$$w_{m,k}^{\mathrm{b}} = \max\{0, c - a_m - ep_{m,k}^{\mathrm{b}}\} = sp_{m,k+p_m+1}^{\mathrm{s}} - ep_{m,k}^{\mathrm{b}} + c - a_m \quad (8)$$

A straight station can be seen as a special crossover station with empty workload in the other side. For example, if straight station m is in front side of the U-line, it can be treated as a crossover station with $t_{m,n}^{\mathrm{b}} = 0$ for all n, $l_m^{\mathrm{b}} = l_m^{\mathrm{f}}$ and $a_m = p_m = 0$. Then equations (1)-(8) turn to:

$$sp_{m,k}^{\mathrm{b}} = ep_{m,k}^{\mathrm{b}} = ep_{m,k}^{\mathrm{f}}, \quad sp_{m,k}^{\mathrm{f}} = \max\{0, ep_{m,k-1}^{\mathrm{f}} - c\}, \quad ep_{m,k}^{\mathrm{f}} = \min\{l_m^{\mathrm{f}}, sp_{m,k}^{\mathrm{f}} + t_{m,\Pi(k)}^{\mathrm{f}}\};$$

$$u_{m,k}^{\mathrm{f}} = \max\{0, t_{m,\Pi(k)}^{\mathrm{f}} + sp_{m,k}^{\mathrm{f}} - l_m^{\mathrm{f}}\}, \quad w_{m,k}^{\mathrm{b}} = \max\{0, c - ep_{m,k}^{\mathrm{f}}\}, \quad u_{m,k}^{\mathrm{b}} = w_{m,k}^{\mathrm{f}} = 0$$

It's easy to verify these results are same with the relationships reported in former studies (Bolat, 1997; Zhao et.al., 2004; etc.) except the idle time after the kth job is split to two parts: $w_{m,k}^{\mathrm{f}}$ and $w_{m,k}^{\mathrm{b}}$. If straight station m is in back side, set $t_{m,n}^{\mathrm{f}} = 0$ for all n, $l_m^{\mathrm{f}} = l_m^{\mathrm{b}}$. In the following context, all stations in U-line are treated as crossover stations.

It's widely accepted high work overload leads to additional costs and inefficiency. The objective of this study is to arrange the products entering the line in a sequence Π so as to minimize the total overload
$U(\Pi)=\sum_{m=1}^{M}U_m(\Pi)=\sum_{m=1}^{M}\sum_{k=1}^{D}(u^{\text{f}}_{m,k}+u^{\text{b}}_{m,k})$.

2.2. Characteristics of MMUL sequencing problem

Recall the MMUL sequencing problem is based on MPS principle. Products are launched according to a basic sequence cyclically. The generating job sequence in each station is also cyclic. Zhao et. al. (2004) has briefly proved in straight line the start position for each product replicates cyclically, which is called "steady state". Following conclusion proves similar conclusion can be made in the context of MMUL.

Lemma 1: For an arbitrary Π, an integer K exists satisfying:
$\forall k>K,\ sp^{\text{f}}_{mk}=sp^{\text{f}}_{m(k+D)}$

Proof: The thought is similar to the idea in Zhao et.al (2004). From Eqs. (1)-(4):
$sp^{\text{f}}_{m,k+1}=\max\{0,\min\{l^{\text{b}}_m,t^{\text{b}}_{m,\Pi(k-p_m)}+\max\{0,\min\{l^{\text{f}}_m,t^{\text{f}}_{m,\Pi(k)}+sp^{\text{f}}_{m,k}\}-a_m\}\}-c+a_m\}=f(sp^{\text{f}}_{m,k})$.
f(x) is a non-decreasing piecewise linear function of *x*, and the linear coefficient in each piece is 1 or 0. $\forall k_0>p_m$, comparing sp^{f}_{m,k_0} and sp^{f}_{m,k_0+D}, the result must be one of following three cases:
1) $sp^{\text{f}}_{m,k_0}=sp^{\text{f}}_{m,k_0+D}$. If this happen, $\forall k>k_0$, $sp^{\text{f}}_{m,k}=sp^{\text{f}}_{m,k+D}$ holds according to Eqs. (1)-(4), then station *m* reaches steady state.
2) $sp^{\text{f}}_{m,i}<sp^{\text{f}}_{m,i+D}$. Then consider the infinite sequence $\{sp^{\text{f}}_{m,i+jD}\}$($j$=0,1,…). Notice $sp^{\text{f}}_{m,i+jD}=f^{(D)}(sp^{\text{f}}_{m,i})$ is a non-decreasing piecewise linear function of $sp^{\text{f}}_{m,i}$. Meanwhile, the maximum value of $sp^{\text{f}}_{m,i}$ is restricted by $\max\{l^{\text{b}}_m-c+a_m,l^{\text{f}}_m-c\}$ according to Eqs. (1)-(4). Thus, the infinite sequence $\{sp^{\text{f}}_{m,i+jD}\}$ (j=0,1,…) must converge to a limit value . After reaching the limit value, the situation transfer to case (1).
3) $sp^{\text{f}}_{m,i}>sp^{\text{f}}_{m,i+D}$. $\{sp^{\text{f}}_{m,i+jD}\}$ (j=0,1,…) is an infinite non-increasing sequence with lower bound 0, must converging a limit value. Then after reaching the limit value, the situation transfer to case (1).

This means in steady state, $U_m(\Pi)$, $W_m(\Pi)$ are equal to the total overload, idle time during processing any consecutive *2D* jobs (including the jobs with empty content). Redefine $sp^{\text{T}}_{mk}, ep^{\text{T}}_{mk}, u^{\text{T}}_{mk}, w^{\text{T}}_{mk}$, $U_m(\Pi), W_m(\Pi), U(\Pi), W(\Pi)$ as the definitions in steady state, following relationship holds:
$$cD=\sum_{n=1}^{N}(t^{\text{f}}_{m,n}+t^{\text{b}}_{m,n})d_n-U_m(\Pi)+W_m(\Pi). \qquad (9)$$

The left side is the total available time for processing a MPS, and the right side is the sum of actual processing time (required processing time minus overload) and idle time. The equation is intuitive and can be verified by Eqs. (1)-(8). Following from Eqs. (9), some other results can be obtained.

Corollary 1: For a given Π, in any station *m*, $U_m(\Pi)$, $W_m(\Pi)$ is independent with the initial position where the work starts.

Proof: According to Eqs. (1)-(8), in steady state, $U_m(\Pi)$ is a non-increasing function and $W_m(\Pi)$ is a non-decreasing function of the initial position, while Eq. (9) implies $[U_m(\Pi)-W_m(\Pi)]$ is invariant for different initial position. Hence, for a given Π, no matter where the worker starts his work, the status in steady state are the same. □

Corollary 2: $\min \sum_{m=1}^{M} U_m(\Pi) \Leftrightarrow \min \sum_{m=1}^{M} W_m(\Pi)$.

Proof: Following from Eq. (9), $U_m(\Pi) - W_m(\Pi) = \sum_{n=1}^{N} (t_{m,n}^{\mathrm{f}} + t_{m,n}^{\mathrm{b}}) d_n - cD$=constant. □

Corollary 3: For any Π, the lower bound for $U_m(\Pi)$ is given by:
$LBU_m = \sum_{n=1}^{N} (t_{m,n}^{\mathrm{f}} + t_{m,n}^{\mathrm{b}}) d_n - cD$.

Proof: Since $W_m(\Pi) \geq 0$, the result follows directly from Eq. (9). □

Corollary 1 means in steady state, any product in the sequence can be seen as the first position in the basic sequence Π. For example, the overload of basic sequence CACB in steady state is the same with the sequences ACBC, CBCA, BCAC. That is to say, each solution of has *D* equivalent solutions. Thus, the number of different solutions for the problem is $(D-1)!/\prod_{n=1}^{N} d_n!$.

In this paper, a product sequence is evaluated by its steady-state performance. For a given Π, $U_m(\Pi)$ can be derived from the performance at station *m* during processing the first 2*D* jobs. In detail, the procedures are as follows, noting that *ew* is the total evitable idle time, and *md* is the moving distance between the start position for the *j*th job and the start position for the (*j*+2*D*)th job.
Step 1: Let k=1, $sp_{m,1}^{\mathrm{f}}$=0, md=+∞, ew=0;
Step 2: Calculate $ep_{m,k}^{\mathrm{f}}$, $u_{m,k}^{\mathrm{f}}$, $w_{m,k}^{\mathrm{f}}$, $sp_{m,k-p_m}^{\mathrm{b}}$ using Eqs. (2),(5),(7), (3);
Step 3: If $\min\{w_{m,k}^{\mathrm{f}}, md\}>0$, $ew=ew+\min\{md, w_{m,k}^{\mathrm{f}}\}$; else, $md=\min\{md, l_m^{\mathrm{f}} - ep_{m,k}^{\mathrm{f}}\}$;
Step 4: Calculate, $ep_{m,k-p_m}^{\mathrm{b}}$, $u_{m,k-p_m}^{\mathrm{b}}$, $w_{m,k-p_m}^{\mathrm{b}}$, $sp_{m,k+1}^{\mathrm{f}}$ using Eqs. (4),(6),(8),(1);
Step 5: If $\min\{w_{m,k-p_m}^{\mathrm{b}}, md\}>0$, $ew=ew+\min\{md, w_{m,k-p_m}^{\mathrm{b}}\}$; else, $md=\min\{md, l_m^{\mathrm{b}} - ep_{m,k-p_m}^{\mathrm{b}}\}$;
Step 6: If $k=D$, go to step 7; else, $k=k+1$, go to step 2;
Step 7: $U_m(\Pi) = \sum_{k=1}^{D} (u_{m,k}^{\mathrm{f}} + u_{m,k-p_m}^{\mathrm{b}}) + \max\{0, sp_{m,D+1}^{\mathrm{f}} - ew\}$, Stop.

The principle of these procedures is as follows. Since the next 2*D* job will be processed starting from position $sp_{m,D+1}^{\mathrm{f}}$, denote $\Delta sp_{m,j}$ as the increment of the start position for the *j*th job, $\Delta sp_{m,j}$ will transmit to succeeding jobs. For the *j*th job, if the sum of original end position and the increment of start position is larger than the section length, overload increases $\Delta u_{m,j}$ and the increment of start position for the (*j*+1)th job turns to $\Delta sp_{m,j} - \Delta u_{m,j}$. If idle time exists after the *j*th job, then $\Delta sp_{m,j}$ will lead to decrement of idle time, the lower value between $\Delta sp_{m,j}$ and original idle time.

2.3. Mathematical Formulation

According to Lemma 1, the objective function is Min $\sum_{k=1}^{D} (u_{m,k}^{\mathrm{f}} + u_{m,k}^{\mathrm{b}})$. From Eqs. (5) and (6),

$$\begin{aligned}\sum_{k=1}^{D} (u_{m,k}^{\mathrm{f}} + u_{m,k}^{\mathrm{b}}) &= \sum_{k=1}^{D} (sp_{m,k}^{\mathrm{f}} + sp_{m,k}^{\mathrm{b}} - ep_{m,k}^{\mathrm{f}} - ep_{m,k}^{\mathrm{b}}) + \sum_{k=1}^{D} (t_{m,\Pi(k)}^{\mathrm{f}} + t_{m,\Pi(k)}^{\mathrm{b}}) \\ &= \sum_{k=1}^{D} (sp_{m,k}^{\mathrm{f}} + sp_{m,k}^{\mathrm{b}} - ep_{m,k}^{\mathrm{f}} - ep_{m,k}^{\mathrm{b}}) + \sum_{n=1}^{N} (t_{m,n}^{\mathrm{f}} + t_{m,n}^{\mathrm{b}}) d_n\end{aligned}.$$

$\sum_{n=1}^{N} (t_{m,n}^{\mathrm{f}} + t_{m,n}^{\mathrm{b}}) d_n$ is constant. Hence, the problem can be modeled as follows:

Model 1: Min $\sum_{m=1}^{M} \sum_{k=1}^{D} (sp_{m,k}^{\mathrm{f}} + sp_{m,k}^{\mathrm{b}} - ep_{m,k}^{\mathrm{f}} - ep_{m,k}^{\mathrm{b}})$
s.t. Eqs. (1)-(4) $\forall m, k$;
$ep_{m,k}^{\mathrm{T}} = ep_{m,wrap(k,D)}^{\mathrm{T}}$ $\forall m, k$, T=f or b (10)

$sp^{\mathrm{T}}_{m,k} \le l^{\mathrm{T}}_m$	$\forall m,k$, T=f or b	(11)
$\sum_{k=1}^{D} x_{n,k} = d_n$	n=1,…,N;	(12)
$\sum_{n=1}^{N} x_{n,k} = 1$	k=1,…,D;	(13)
$x_{n,k}$=0 or 1	n=1,…,N; k=1,…,D;	(14)

Eqs. (1)-(4) are the relationships between start positions and end positions. Noting that in constraint (10) , $wrap(k,D)=k+iD$, i is an integer so that the value is in the interval [1,D] . This constraint is to make ensure the line reach steady state. Constraint (11) means no job can be missed since the precedence relationships between jobs. This constraint can be omitted when $0 \le l^{\mathrm{f}}_m - a_m \le l^{\mathrm{b}}_m$, $0 \le l^{\mathrm{b}}_m + a_m - c \le l^{\mathrm{f}}_m$ because Eqs. (1)-(4) have imply this constraint. Constraint (12) ensures all products are processed and constraints (13), (14) imply products are launched one by one.

The complex functions in Eqs. (2), (4) make it difficult to solve Model 1. Actually, Model 1 can be relaxed to a mixed integer linear programming problem by applying condition (10), (11) to Eqs. (1)-(4) and relaxing the equalities to inequalities. The relaxation model is as follows.

Model 2: $\sum_{m=1}^{M}\sum_{k=1}^{D}(sp^{\mathrm{f}}_{m,k} + sp^{\mathrm{b}}_{m,k} - ep^{\mathrm{f}}_{m,k} - ep^{\mathrm{b}}_{m,k})$

s.t. $sp^{\mathrm{T}}_{m,k} \ge 0$	T= f or b; m=1,…,M; k=1,…,D;
$ep^{\mathrm{T}}_{m,k} \le l^{\mathrm{T}}_m$	T= f or b; m=1,…,M; k=1,…,D;
$sp^{\mathrm{f}}_{m,k} \ge ep^{\mathrm{b}}_{m,wrap(k-p_m-1,D)} - (c - a_m)$	m=1,…,M; k=1,…,D;
$ep^{\mathrm{f}}_{m,k} \le sp^{\mathrm{f}}_{m,k} + \sum_{n=1}^{N} t^{\mathrm{f}}_{m,n} x_{n,k}$	m=1,…,M; k=1,…,D;
$sp^{\mathrm{b}}_{m,k} \ge ep^{\mathrm{f}}_{m,wrap(k+p_m,D)} - a_m$	m=1,…,M; k=1,…,D;
$ep^{\mathrm{b}}_{m,k} \le sp^{\mathrm{b}}_{m,k} + \sum_{n=1}^{N} t^{\mathrm{b}}_{m,n} x_{n,k}$	m=1,…,M; k=1,…,D;

Eqs. (12), (13), (14)

Lemma 2: The optimal solution of Model 2 is optimal to Model 1.

Proof: Since Model 2 is relaxation of Model 1, it's only necessary to prove the optimal solution of Model 2 is feasible to Model 1. Assuming ($SP^{\mathrm{T}}_{m,k}$, $EP^{\mathrm{T}}_{m,k}$, $X_{n,k}$) is an optimal solution of Model 2. ($SP^{\mathrm{T}}_{m,k}$, $EP^{\mathrm{T}}_{m,k}$) must be the optimal solution for the following model.

$$\sum_{m=1}^{M}\sum_{k=1}^{D}(sp^{\mathrm{f}}_{m,k} + sp^{\mathrm{b}}_{m,k} - ep^{\mathrm{f}}_{m,k} - ep^{\mathrm{b}}_{m,k})$$

s.t. $sp^{\mathrm{f}}_{m,k} \ge \max\{0, ep^{\mathrm{b}}_{m,wrap(k-p_m-1,D)} - c + a_m\}$	m=1,…,M; k=1,…,D;
$ep^{\mathrm{f}}_{m,k} \le \min\{l^{\mathrm{f}}_m, sp^{\mathrm{f}}_{m,k} + \sum_{n=1}^{N} t^{\mathrm{f}}_{m,n} X_{n,k}\}$	m=1,…,M; k=1,…,D;
$sp^{\mathrm{b}}_{m,k} \ge \max\{0, ep^{\mathrm{f}}_{m,wrap(k+p_m,D)} - a_m\}$	m=1,…,M; k=1,…,D;
$ep^{\mathrm{b}}_{m,k} \le \min\{l^{\mathrm{b}}_m, sp^{\mathrm{b}}_{m,k} + \sum_{n=1}^{N} t^{\mathrm{b}}_{m,n} X_{n,k}\}$	m=1,…,M; k=1,…,D;

This is a linear programming with 4MD variables and 4MD independent constraints. The optimal solution ($SP^{\mathrm{T}}_{m,k}$, $EP^{\mathrm{T}}_{m,k}$) must be one of the vertexes of the feasible region, that is to say, all the constraints are equations at that point. Hence, the solution ($SP^{\mathrm{T}}_{m,k}$, $EP^{\mathrm{T}}_{m,k}$, $X_{n,k}$) satisfies the Eqs. (1)-(4), and feasible to Model 1. □

Lemma 2 implies the optimal sequence can be found by solving Model 2 optimally. Model 2 is a mixed integer linear programming problem with ND integer variables and 4MD real variables, which can be solved by traditional method such as branch-and-bound, or by commercial optimization software.

3. Branch-and-trim Algorithm

We propose a branch-and-trim algorithm to find the optimal solution. Bolat (1997) has applied this method to solve the sequencing problem in straight line. To describe the procedure in detail, some additional notations are introduced. For a node, let $R=(r_1, \ldots, r_N)$ be the vector of remaining product types, and $J_m(R)$ be the partial job sequence at station m. Denote u_m (w_m) as the total overload (idle time) for the sequenced jobs at station m, ew_m as the maximum evitable idle time for sequenced jobs and LB_m as the lower bound of overload for remaining products. The procedures are as follows.

1) Initialization;
2) Branch. Fill the next position with an available product, update R;
3) Evaluate the nodes with the criteria $\sum_{m=1}^{M}(u_m + LB_m)$;
4) Trim. If the performance of a node is worse than the best known complete solution, eliminate the node.
5) Repeating (2)-(4) until all product are sequenced.

According to the problem description in 2.1, the job sequence in each station is determined by the basic sequence Π and p_m. Before a complete Π is given, the generating job sequence is discontinuous. For example, if D=4, p_m=2, the partial product sequence is CA** (* represents the position is not filled yet), then $J_m(R)$ is $C_m^{\mathrm{f}} * A_m^{\mathrm{f}} ** C_m^{\mathrm{b}} * A_m^{\mathrm{b}}$, where job A_m^{b} and C_m^{f} are adjacent while A_m^{f} and C_m^{b} are isolated. Define "subsequence" as a serial of consecutive sequenced jobs in the generating job sequence. In the above example, there are 3 subsequences in the partial job sequence: $A_m^{\mathrm{b}}C_m^{\mathrm{f}}$, A_m^{f}, C_m^{b}. The number and content of subsequences depends on the number of sequenced products and p_m. Next, each step of the algorithm is introduced in detail.

3.1. Procedures

Initialization

Since any product can be seen as the first product in a basic sequence, we set the product with least demand to the first position to reduce the number of feasible solutions, i.e., $R=(d_1, \ldots, d_i\text{-}1, \ldots, d_N)$, where $d_i = \min_{1 \le n \le N}\{d_n\}$. The number of all feasible solution is $\dfrac{d_i(D-1)!}{\prod_{n=1}^{N} d_n!}$. Meanwhile, set the start positions for the first job in all subsequences in all stations to 0.

Branch

Product n is available if $r_n \geq 1$. Allocating any available product to the next position will generate a son node. If a product from type n is selected, update $J_m(R)$ by filling the $[2(D-\sum_{i=1}^{N} r_i)\text{-}1]$th position with job N_m^{f} and filling wrap$[2(D\text{-}\sum_{i=1}^{N} r_i + p_m), D]$th position with the job N_m^{b}. Then update R with $r_n = r_n\text{-}1$. If no product is available, stop.

Evaluating

This is the key procedure of the algorithm. According to the method calculating $U_m(\Pi)$, similar procedure can be used to update u_m, w_m, ew_m simultaneously. For a given subsequence π consisted of h jobs, denote $u_m(\pi)$, $w_m(\pi)$, $ew_m(\pi)$ as the overload, idle time and evitable idle time for the jobs in π. Let sp_j (ep_j) be the start (end) position of the jth job in π, u_j (w_j) be overload (idle time) of the jth job. $u_m(\pi)$, $w_m(\pi)$, $ew_m(\pi)$ can be updated in following steps.

Initialize: j=1; sp_j =0; $u_m(\pi)=w_m(\pi)=ew_m(\pi)$=0; $md=+\infty$;

Until $j=h$

$ep_j = \min\{\max\{sp_j, l_j\}, sp_j + t_j\}$;

$sp_{j+1} = \max\{0, ep_j - a_j\}$;

$u_j = sp_j + t_j - ep_j$;
$w_j = sp_{j+1} + a_j - ep_j$;
$u_m(\pi) = u_m(\pi) + u_j$;
$w_m(\pi) = w_m(\pi) + w_j$;
If $\min\{w_j, md\} > 0$
$ew_m(\pi) = ew_m(\pi) + \min\{md, w_j\}$
Else $md = \min\{md, l_j - ep_j\}$
$j = j+1$
End

Noting t_j is the required processing time for the jth job, l_j is the length of section where the jth job is processed. If the jth job is processed in front side, $l_j = l_m^{\mathrm{f}}$ and $a_j = a_m$; If the jth job is processed in back side, $l_j = l_m^{\mathrm{b}}$ and $a_j = c - a_m$. u_m (w_m, ew) is the sum of all $u_m(\pi)$ ($w_m(\pi)$, $ew_m(\pi)$) for any $\pi \subset J_m(R)$. The following result is used to calculate LB_m.

Lemma 3: The lower bound for the overload during processing remaining products in station m can be estimated by:

$$LB_m = \max\{0, \textstyle\sum_{n=1}^{N}(t_{m,n}^{\mathrm{f}} + t_{m,n}^{\mathrm{b}})d_n - cD - u_m + w_m - ew_m\}.$$

Proof: The required processing time for remaining products is $\sum_{n=1}^{N}(t_{m,n}^{\mathrm{f}} + t_{m,n}^{\mathrm{b}})r_n$. According to Lemma 1, the total available processing time for all products in a MPS is cD. The amount of time spent on these sequenced jobs is $\sum_{n=1}^{N}(t_{m,n}^{\mathrm{f}} + t_{m,n}^{\mathrm{b}})(d_n - r_n) - u_m + w_m$. The maximum evitable idle time during processing the sequenced jobs is ew_m. Hence, the total available time for remaining products is

$$cD - [\textstyle\sum_{n=1}^{N}(t_{m,n}^{\mathrm{f}} + t_{m,n}^{\mathrm{b}})(d_n - r_n) - u_m + w_m] + ew_m.$$

The lower bound of overload for remaining products is:

$$\begin{aligned} LB_m &= \max\{0, \textstyle\sum_{n=1}^{N}(t_{m,n}^{\mathrm{f}} + t_{m,n}^{\mathrm{b}})r_n - [cD - (\sum_{n=1}^{N}(t_{m,n}^{\mathrm{f}} + t_{m,n}^{\mathrm{b}})r_n - u_m + w_m) + ew_m]\} \\ &= \max\{0, \textstyle\sum_{n=1}^{N}(t_{m,n}^{\mathrm{f}} + t_{m,n}^{\mathrm{b}})d_n - cD - u_m + w_m - ew_m\} \end{aligned}$$ □

The performance of a node is evaluated with $\sum_{m=1}^{M}(u_m + LB_m)$.

Trim

If the performance of a node is worse than the best known complete solution, the node is eliminated. If a new complete solution is obtained, update the best known complete solution.

3.2. An Example

As an illustration, the proposed algorithm is applied to a modified MMUL instance based on the one in Kara et.al (2007). There are five stations in the U-line, three of which are crossover station. Five types of products are produced in the line. The cycle time is set to 29. Detailed parameters of this example are given in Table 2.

Table 2: parameters of the example

products / stations	A	B	C	D	E	l_m^T
1-f	13	6	21	21	22	49
1-b	5	7	9	13	0	42
2-f	18	18	23	34	10	49
2-b	10	17	10	2	13	34
3(back side)	20	17	21	32	32	39
4-f	11	10	2	26	17	37
4-b	7	11	12	15	2	34
5(front side)	55	27	37	29	33	66
d_n	2	1	1	3	2	

AADEBCDED is an optimal solution with 83 time units of overload in steady state. The optimal value is equal to the theoretical lower bound.

4. Conclusions

This paper deals with the model sequencing problem in a paced mixed model U-line. The problem is studied in the background of cyclic production and the objective is to minimize the overload in steady state. Some characteristics of the problem are identified. We proved the total overload in steady state is independent with the initial position, and the minimization of the total overload in steady state is equivalent to the minimization of the total idle time. Beside, a method to calculate the overload for a given sequence in steady state is presented. Based on these characteristics, a mathematic model is built and its optimal solution can be obtained by optimally solving a mixed integer linear programming. At last, a branch-and-trim algorithm is proposed to solve the problem optimally, and is applied to a modified instance to verify its validity.

References

Bautista and Cano (2008), Minimizing work overload in mixed-model assembly lines. International Journal of Production Economics, 112(2): 177–191.

Bolat A. and Yano C.A. (1992). Scheduling Algorithms to Minimize Utility Work at a Single Station on a Paced Assembly Line. Production Planning and Control, 3(): 393-405.

Bolat A. (1997), Efficient Methods for Sequencing Minimum Job Sets on Mixed Model Assembly Lines. Naval Research Logistics, 44():419-437

Boysen N., Fliedner M. and Scholl A. (2007). Sequencing mixed-model assembly lines: Survey, classification and model critique. European Journal of Operational Research, 192(): 349-373.

Inman and Bulfin (1991). Notes on Sequencing JIT Mixed-Model Assembly Lines. Management Science, 37(): 901-905.

Kara Y., Ozcan U. and Peker A. (2007). Balancing and sequencing mixed-model just-in-time U-lines with multiple objectives. Applied Mathematics and Computation, 184(): 566-588.

Kim Y. K., Hyun C.J. and Kim Y. (1996). Sequencing in Mixed Model Assembly Lines: a Genetic Algorithm Approach. Computers and Operations Research, 23(): 1131-1145.

Kim Y. K., Kim J.Y. and Kim Y. (2006). An endosymbiotic evolutionary algorithm for the integration of balancing and sequencing in mixed-model U-lines. European Journal of Operational Research, 168(): 838-852.

Miltenburg J. (1989). Level Schedules for Mixed-Model Assembly Lines in Just-in-Time Production Systems. Management Science, 35(): 192-207.

Miltenburg J. (2002). Balancing and Scheduling Mixed-Model U-Shaped Production Lines. International Journal of Flexible Manufacturing Systems, 14():119-151

Miltenburg J. and Wijingaard J. (1994). The U-line balancing problem. Management Science, 40():1378-1388

Sparling D. and Miltenburg J. (1998). The mixed-model U-line balancing problem. International Journal of Production Research, 36():485-501.

Tsai L.H. (1995). Mixed-model Sequencing to Minimize Utility Work and the Risk of Conveyor Stoppage. Management Science, 41(): 485-495.

Zhao X.B., Ohno K. and Lau H.S. (2004). A Balancing Problem for Mixed Model Assembly Lines with a Paced Moving Conveyor. Naval Research Logistics, 51():446-464

Zhao X.B. and Ohno K. (2000). Properties of a sequencing problem for a mixed model assembly line with conveyor stoppages. European Journal of Operational Research, 124():560-570

Yano C.A. and Rachamadugu R. (1991). Sequencing to Minimize Work Overload in Assembly Lines with Product Options. Management Science, 37(): 572-586.

Value of information sharing in marine mutual insurance

Kevin Li[1], *John Liu*[2], *Jia Yan*[3] *and Jie Min*[2]

[1]Department of Logistics & Maritime Studies, The Hong Kong Polytechnic University, Hong Kong.
Email address: k.x.li@polyu.edu.hk.
[2]Department of Logistics & Maritime Studies, The Hong Kong Polytechnic University, Hong Kong.
[3]School of Economic Sciences, Washington State University, US.

Abstract

With empirical evidence from marine mutual insurance (MMI), an impulse feedback model is constructed to address how information sharing can help increase both the social welfare and the efficiency of operation of the MMI system. Focusing on information sharing, this paper considers the premium policy optimization of a pure (i.e., non-stock consideration) mutual insurance system with a homogeneous market of identical members. Our findings confirm that the principle of information sharing can be attained under equal-risk pooling, but not necessarily under unequal-risk pooling, and reveal that quantifiable differences exist in the valuation of information sharing under the two schemes of risk pooling. It indicates that the key to successful MMI is equal-risk pooling. Algorithms are developed to compute the value of information sharing by solving the HJB equations and quasi-variational inequalities. It determines that information sharing can achieve best social welfare as well as efficient operation of a P&I Club. The conclusion provides a scientific basis for both managerial strategy and competition regulation. The findings are also applicable to a wide range of reserve and inventory management problems.

Key words: Information sharing; Mutual Insurance; P&I Club

1. Introduction

In the marine insurance market, mutual Clubs are not the pure commercial insurance enterprises as conventionally defined. Insurance is based on membership of the Club. An individual ship-owner, who wants to pool his risk in a mutual Club, must first obtain membership through payment of a membership fee, which is deemed to be the premium as defined in commercial insurance (Cass, Chichilnisky and Wu, 1996; Lamm-Tennant and Starks, 1993).

Numerous studies into the formation of mutual insurers concern adverse selection (Smith and Stutzer, 1990; Ligon and Thistle, 2005), moral hazards (Smith and Stutzer, 1995) and information asymmetry (Cabrales et al, 2003). At the outset of the insurance industry, there were only mutual insurers, and because of the existence of adverse selection and information asymmetry, the original risk pool degenerated into two types of sub-risk-pools formed by the homogeneous assureds.

There are a total 28 Marine Mutual Clubs, 13 of which are members of the International Group of P&I Clubs (see table 1). The 13 P&I Clubs that comprise the International Group of P&I Clubs (the IG) are mutual not-for-profit insurance organizations that between them cover third party liabilities, which include pollution, loss of life and personal injury, cargo loss and damage and collision risks. The Clubs are mutual organizations, that is, the shipowner members are both insured and insurers, as the members both own and control their individual clubs (Gold, 2002; Hazelwood, 2000). The day-to-day activities and operations of the Clubs are delegated to managers. The 13 P&I Clubs cover about 95% of the world's ocean-going vessels in terms of tonnage (JLT, 2003; Golish, 2005). This monopoly position, and the way that P&I Clubs operate, have triggered two European Commission (EC) investigations and rulings, in 1985 and 1999 respectively, in the field of competition law (Gyselen, 1999; EC, 1985, 1999).

	Name of Club	Year Established	Headquarters	IG member	Annual Growth	Size (gt) (million)
1	UK	1869	London	Yes	10.06%	120
2	Gard	1907	Arendal	Yes	4.76%	98
3	Britannia	1854	London	Yes	5.68%	80
4	Steamship	1974	London	Yes	1.38%	65
5	Standard	1885	London	Yes	13.03%	58
6	Japan	1950	Tokyo	Yes	2.24%	54
7	Skuld	1897	Oslo	Yes	-5.23%	52
8	W. England	1856	London	Yes	3.73%	46
9	N. England	1860	Newcastle	Yes	19.72%	43
10	London	1866	London	Yes	2.46%	28
11	American	1917	New York	Yes	19.33%	18
12	Swedish	1872	Goteborg	Yes	2.36%	15
13	China	1984	Beijing	No	30.47%	9
14	Shipowners	1855	London	Yes	9.86%	9
	Average				**9.11%**	**46**

Table 1: Major Marine Mutual Insurance

Notes: 1. This has been combined from various sources by the authors.
2. IG – International Group of P&I Clubs; g.t. – gross tonnage of entered ships.

A key message from the EC is that the P&I Club system should be more transparent both to its members and to the outside world, in order to ensure full implementation of the principle of information sharing (Garner, 1999; Macey, 2004). The International Maritime Organization (IMO) has also asked P&I Clubs to open their books, for the general interest of "safer ships and cleaner oceans." Such suggestion has always been rejected by P&I Clubs under the defense of 'protection of privacy', so the lack of data from P&I Clubs contributes to the need for academic research to be conducted in this area (Johnstad, 2000; Bennett, 2000). Much information about MMI in general has been "veiled in antiquity and lost in obscurity" (Dover, 1975). All of these discussions have pointed to one critical legal argument — whether MMI should open up or continue to keep closed information on claims records and other information. However, this paper determines that information sharing can achieve best social welfare as well as efficient operation of a P&I Club.

Assuming the standard Brownian motion characterization of a claim process associated with each individual vessel of a club (Tapiero and Jacque, 1987; Asmussen and Taksar, 1997; Siegl and Tichy, 1999), we develop for the MMI cost minimization problem an impulse feedback model of a pure MMI (Bensoussan and Lions, 1984; and Aubin, 2000). We focus the analysis on a comparison of optimal insurance positions (i.e., funds on hand) under two information structures — unequal-risk versus equal-risk. In practice, P&I Clubs collect premiums at the beginning of each policy year on 28 February. Therefore, we model a P&I Club as an impulse control system in the sense that the total reserve of the Club is "reset" by an impulse premium control, so that insurance claims incurred during the policy year can be sufficiently covered at a desirable level of risk pooling.

Through an impulse feedback analysis, we first calculate the optimal premium policies (in terms of total-cost minimization) under different risk pooling structures (unequal- versus equal-risk pooling). We then determined a quantifiable extra value in an unequal-risk pooling MMI system, as compared with an equal-risk pooling one. In this paper, this extra value term is referred to as the "value" of information sharing. We develop algorithms to compute the value of information sharing, by solving the HJB equations and quasi-variational inequalities.

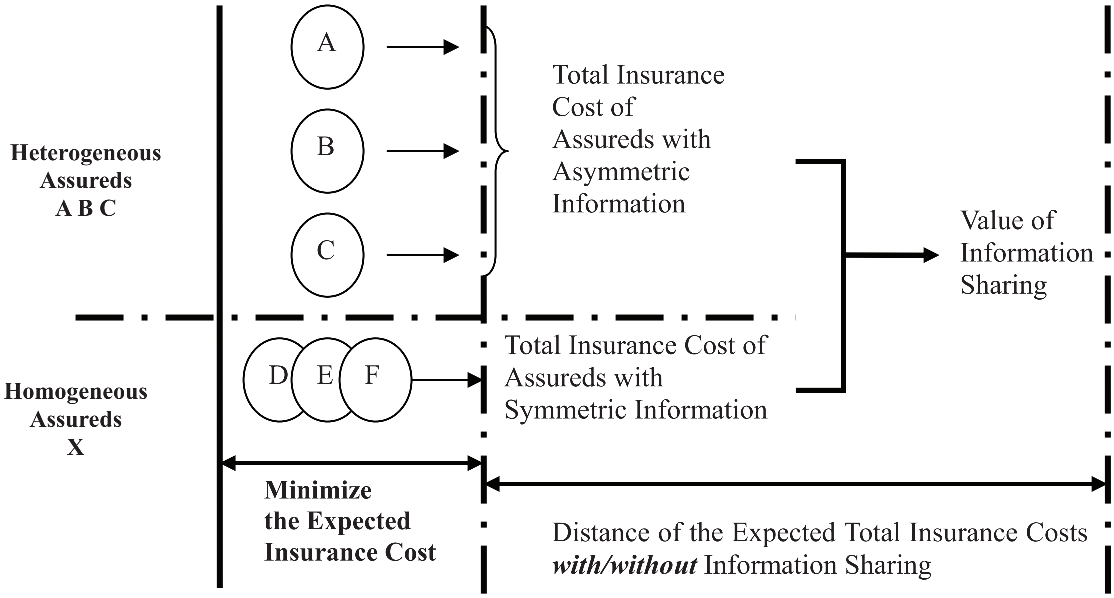


Figure 1: Value of Information Sharing

2. Impulse Feedback Model for Unequal-Risk Pooling

Consider a general mutual insurance Club of *n* members over a time horizon $[0,T]$ (when $T=\infty$, $[0,T]=[0,\infty)\cup\{\infty\}$), during which each member *i* establishes an account of insurance position y_t^i (i.e., individual account balance) at time *t*, and a record of the individual claims incurred by member *i* (denoted by x_t^i). The records of individual accounts and claims are kept confidential at the Club. Let each Gaussian claim process $x_t^i \sim N(\lambda_i, \sigma_i^2)$ be characterized as a drifted Brownian motion, as follows:

$$dx_t^i = \lambda_i dt + \sigma_i d\widetilde{w}_t^i \quad (i=1,\cdots,n), \tag{1}$$

where $d\widetilde{w}_t^i$ represents a Wiener differential. The Club will review the current insurance position and claims outlook at the beginning of each renewal period k (e.g., each year *k*, $k=1,2,\cdots$). Following the review, the Club then determines for each member a premium call (an impulse control), denoted as q_k^i, so as to "reset" the member's insurance position; i.e., $y_k^i = y_{k-1}^i + \int_{k-1}^{k} dy_t^i + q_k^i$, where dy_t^i is the differential of the insurance position during a review period $t\in[k-1,k)$. We characterize this differential using the following differential characteristics:

$$\begin{aligned} dy_t^i &= h_i(y_t^i)dt - dx_t^i \\ &= \mu_i(y_t^i)dt + \sigma_i dw_t^i \; (i=1,\cdots,n), \end{aligned} \tag{2}$$

where $h_i(y_t^i)$ represents the rate of individual position consumption (e.g., operational and management costs) excluding the claim coverage costs, $\mu_i(y_t^i) = h_i(y_t^i) - \lambda_i$, and $dw_t^i = -d\widetilde{w}_t^i$. For the rest of the paper, we adopt the following notation system: Denote by small letter the vector, e.g., the position vector as $y_t = (y_t^1, \cdots, y_t^n)$, and by a capital letter the sum $Y_t = \sum_{i=1}^{n} y_t^i$. The dynamics of an MMI system can be concisely presented in vector form as

$dy_t = \mu(y_t)dt + \Theta dw_t$,

where $\mu(y_t)$ is an $n \times 1$ vector drift, $\Theta = (\theta_{ij})_{n\times n}$ is a matrix disturbance with $\theta_{ii} = \sigma_i$ and $\theta_{ij} = 0$ for $i \neq j$, and dw_t is an *n*-dimension Wiener differential (Gollier and Wibaut, 1992; Taksar, 2000). As a function of the Club's state y_t, let $L_i(y_t)$ be a nonnegative non-decreasing management cost (Lagrangian) associated with member *i* of the Club, and $L(y_t) = \sum_{i=1}^{n} L_i(y_t)$ be the total MI operating cost of the Club. Similarly, there is an impulse cost (including the premium payment) associated with each premium call q_k^i for renewal year *k*, denoted by $K_i(q_k)$, and an aggregate impulse cost for the entire Club $K(q_k) = \sum_{i=1}^{n} K_i(q_k^i)$ for all $k \in T_q$. Thus, the Club is facing the problem of minimizing the total insurance costs under unequal-risk information (Kavadias and Loch, 2003; Kulkarni, Magazine and Raturi, 2004), which we formulate, in vector form, as an impulse control system (Bensoussan and Lions, 1984; Aubin, 2000), as follows:

$$\varphi\,(y_0) = \min_{q_k} E\left(\sum_{k\in T_q} K(q_k)e^{-rk} + \int_0^T L(y_t)e^{-rt}dt + \Omega(y_T)e^{-rT} \right), \text{ for any } y_0 \qquad (3)$$

Subject to:

$$\begin{cases} dy_t = \mu\,(y_t)dt + \Theta dw_t \\ y_k = y_{k-1} + \int_{k-1}^{k} dy_t + q_k \end{cases}$$

where $\Omega(y_T)$ is a terminal function.

2.1. Optimal Premium Policy under Unequal-Risk: Impulse Feedback Control

According to impulse control theory, the value function $\varphi(y)$ is a solution to the following quasi-variational inequalities (QVI) system (see Polyanin and Zaitsev, 1995; Aubin, 2000; Liu, 2004):

$$\begin{cases} \text{i)} \quad r\varphi(y) - A\varphi(y) - H(y, D\varphi) \le 0 \\ \text{ii)} \quad \varphi(y) - (\varphi * K)(y) \le 0 \\ \text{iii)} \quad \big(\varphi - (\varphi * K)\big)\ \big((r\varphi - A\varphi) - H\big) = 0 \end{cases} \qquad (4)$$

where:

1) $D\varphi = (D_y\varphi) = (\varphi_{y^1}, \cdots, \varphi_{y^n})$ represents a gradient of φ, where φ_{y^i} is the marginal insurance cost shared by member *i*.
2) $A = \sum_{i=1}^{n} \frac{1}{2}\sigma_i^2 \frac{\partial^2}{\partial (y^i)^2}$ is a second-order differential (in viscosity sense) operator.
3) $H(y, D\varphi) = \langle D\varphi, \mu \rangle + L$ is the Hamiltonian of the MI system (3).
4) $(\varphi * K)(y) := \inf_q \big(\varphi(y+q) + K(q)\big)$ is termed the *inf-convolution* of functions φ and *K*.

According to impulse control analysis, there exists for system (3) an optimal position $y^* = (y^{1*}, \cdots, y^{n*})$ termed an impulse feedback control, to which the actual position at the beginning of each renewal period must be "reset" by collecting premiums accordingly. For the sake of reference, we present below the optimal impulse feedback policy for the individual positions, but with the proofs omitted (Aubin, 2000; Liu, 2004).

PROPOSITION 1. *For the MMI system (3), let* y_{k^-} *be the position (vector) realized at review time k before the renewal premium is collected; i.e.,* $y_{k^-} = y_{k-1} + \int_{k-1}^{k} dy_t$ *. Then there exists an optimal position* y^**, such that the optimal renewal premium is determined by the following base-stock policy:*

$$q_k = \begin{cases} y^* - y_{k^-} & \text{if } y_{k^-} < y^* \\ 0 & \text{otherwise} \end{cases}$$

We shall note that the optimal impulse feedback policy of the base-stock type described in Proposition 1 is obtained without regard to the type of information structure (i.e., unequal-risk versus equal-risk). Therefore, we shall confine our analysis of MMI systems to feedback policy, under either the unequal-risk or equal-risk information structure. For convenience, we use the term *unequal-risk* (or *equal-risk*) impulse feedback to differentiate base-stock feedback types under an *unequal-risk* (or *equal-risk*) information structure.

2.2. Heterogeneous Membership: Unequal-Risk Pooling

While complete information about positions and claim records is kept at the Club, each member *i* is only informed of the individual optimal impulse feedback position y^{i*}, and her own claims process x^i. The Club now needs to determine a risk level for the MMI system (3). For this purpose, we define an optimal unequal-risk risk level ξ^{i*} as:

$$\xi^{i*} \equiv \Pr\left(y^{i*} < x^i\right) \quad \text{for each } i$$

Given $x^i \sim N(\lambda_i, \sigma_i)$, we can then immediately write

$$y^{i*} = \lambda_i + \Phi^{-1}(1 - \xi^{i*})\sigma_i \,. \qquad (5)$$

Under the unequal-risk impulse feedback, each member *i* is informed of y^{i*} and ξ^{i*}, which are kept confidential as its unequal-risk records. Note that an unequal-risk MMI system is in general unequal-risk pooling. At the Club level, an average risk level can be calculated using $\bar{\xi}^* = \frac{1}{n}\sum_{i=1}^{n} \xi^{i*}$ as an assessment of aggregate risk pooling. Let $\{y^*, \bar{\xi}^*\}$ denote the unequal-risk MI impulse feedback system, as described by Proposition 1.

3. Equal-Risk Pooling for Homogeneous Membership

Suppose that the managers of the same Club are compelled to operate with an equal-risk pooling scheme, under which an optimal base-stock policy, denoted by $\tilde{y}^{i*}$, is now open to every member of the Club. We shall note that the simplest implementation alternative of equal-risk pooling is to adopt homogeneous membership. Before we discuss the details of $\tilde{y}^* = (\tilde{y}^{1*}, \cdots, \tilde{y}^{n*})$, let us introduce the concept of equal-risk pooling regarding insurance claims. Let $\widetilde{Y}^* = \sum_{i=1}^{n} \widetilde{y}^{i*}$ be the total insurance position of the Club under an equal-risk information structure. In this case, the Club strives to maintain the aggregate position $\widetilde{Y}^*$ by collecting premiums from the members, which money is then used to collectively cover aggregate claims, $X = \sum_{i=1}^{n} x^i \sim N(\lambda,\ \sigma^2)$, where $\lambda = \sum_{i=1}^{n} \lambda_i$ and $\sigma = \sqrt{\sum_{i=1}^{n} \sigma_i^2}$. Because individual account records have been risk-equalized, the total minimized insurance position $\widetilde{Y}^*$ is expected to be reduced in comparison with the total unequal-risk position $Y^* = \sum_{i=1}^{n} y^{i*}$, where y^{i*} is the unequal-risk optimal position as given in Proposition 1. Thus, the allocation of total funds $\widetilde{Y}^*$ among heterogeneous members should no longer be determined based on an actual individual claims record x^i. Instead, the individual premium rates are determined according to the principle of information sharing; i.e., the principle of equal-risk sharing, as opposed to unequal-risk pooling as given in (5). The idea of the collective coverage of claims by the equal-risk allocation of premium contributions is termed *equal-risk pooling*.

3.1. Impulse Feedback Model under Equal-Risk Pooling

Thus, the result of equalizing risk pooling, (or equivalently homogenizing membership) is a potentially reduced effective share of claims coverage. The individual's effective share of claims coverage under equal-risk pooling, denoted as $\widetilde{x}^i \sim N(\widetilde{\lambda}_i, \widetilde{\sigma}_i^2)$, has the same mean as the actual individual claim x^i, but has a smaller variance (for $n > 1$) as a result of equal-risk pooling (or membership homogenizing); i.e.,

$$E(\widetilde{x}^i) = \widetilde{\lambda}_i = \lambda_i, \text{ and } \operatorname{var}(\widetilde{x}^i) = \widetilde{\sigma}_i^2 < \sigma_i^2. \qquad (6)$$

With the above characteristics, a member's effective share of the claims under equal-risk pooling is $\widetilde{x}^i \sim N(\lambda_i, \widetilde{\sigma}_i)$, as compared to its actual share of the claims $x^i \sim N(\lambda_i, \sigma_i)$. The effective and actual shares have the same mean λ_i, and only differ in standard deviation, specifically $\widetilde{\sigma}_i < \sigma_i$ for $n > 1$. Replacing actual σ_i with effective $\widetilde{\sigma}_i$, we can derive from equation (2) the dynamics for the effective position under equal-risk pooling, and express them in vector form as follows:

$$d\widetilde{y}_t = \mu(\widetilde{y}_t)dt + \widetilde{\Theta} dw_t,$$

where $\widetilde{\Theta} = \left(\widetilde{\theta}_{ij}\right)_{n\times n}$ with $\widetilde{\theta}_{ii} = \widetilde{\sigma}_i$ and $\widetilde{\theta}_{ij} = 0$ for $i \neq j$. Then, with the same cost structure, we can derive from system (3) an equal-risk impulse feedback MMI system as being

$$\widetilde{\varphi}(y_0) = \min_{\widetilde{q}_k} E\left(\sum_{k\in T_q} K(\widetilde{q}_k)e^{-rk} + \int_0^T L(\widetilde{y}_t)e^{-rt}dt + \Omega(\widetilde{y}_T)e^{-rT} \right), \text{ for any } y_0 \qquad (7)$$

subject to:

$$\begin{cases} d\widetilde{y}_t = \mu(\widetilde{y}_t)dt + \widetilde{\Theta} dw_t \\ \widetilde{y}_k = \widetilde{y}_{k-1} + \int_{k-1}^{k} d\widetilde{y}_t + \widetilde{q}_k \end{cases}.$$

The optimal equal-risk impulse feedback position $\widetilde{y}^* = (\widetilde{y}^{1*}, \cdots, \widetilde{y}^{n*})$ can then be obtained from the same QVI system (4), except for the operator $\tilde{\mathsf{A}}$, which is modified with $\tilde{\sigma}_i$ as:

$$\tilde{\mathsf{A}} = \sum_{i=1}^{n} \tfrac{1}{2} \tilde{\sigma}_i^2 \frac{\partial^2}{\partial(\tilde{y}^i)^2} .$$

3.2. Equal-Risk Pooling under Impulse Feedback

In this subsection, we show that the optimal position $\tilde{y}^* = (\tilde{y}^{1*}, \cdots, \tilde{y}^{n*})$ entails an equal-risk pooling scheme among all the members. For an equal-risk MMI system, an optimal aggregate risk level can be defined as

$$\tilde{\xi}^* \equiv \Pr\left(\tilde{Y}^* < X\right).$$

With the optimal aggregate risk level $\tilde{\xi}^*$, an optimal equal-risk MMI impulse system can be denoted as $\{\tilde{y}^*, \tilde{\xi}^*\}$. An optimal equal-risk impulse feedback MMI system $\{\tilde{y}^*, \tilde{\xi}^*\}$ can be determined by solving the corresponding QVI system. An optimal unequal-risk impulse feedback MMI system was previously obtained as $\{y^*, \xi^*\}$.

Unlike the unequal-risk pooling MMI impulse system, we show below that individual impulse feedback under equal-risk pooling $\tilde{y}^{i*}$, which is uniquely determined from the corresponding QVI system, can be implemented through an equal-risk level $\tilde{\xi}^*$ for all members.

PROPOSITION 2. *Denote by* $\{\tilde{y}^*, \tilde{\xi}^*\}$ *an optimal equal-risk impulse feedback MMI system. Then, under an equal-risk pooling scheme at level* $\tilde{\xi}^*$*, it holds that:*

$$\tilde{\sigma}_i = \frac{\sigma_i^2}{\sigma} \quad \text{and} \quad \tilde{y}^{i*} = \lambda_i + \Phi^{-1}(1 - \tilde{\xi}^*) \cdot \tilde{\sigma}_i , \qquad (8)$$

where $\tilde{\sigma}_i$ *is the effective standard deviation under equal-risk pooling as given in (6).*

Proof. By the constraint that $\tilde{\xi}^* = \Pr(\tilde{Y}^* < X)$, we can write that:

$$\tilde{Y}^* = \lambda + \Phi^{-1}(1 - \tilde{\xi}^*)\sigma ,$$

or, equivalently:

$$\sum_{i=1}^{n} \left(\lambda_i + \Phi^{-1}(1 - \tilde{\xi}^*) \frac{\sigma_i^2}{\sigma} - \tilde{y}^{i*} \right) = 0 .$$

Letting each item of the summation above be zero, we obtain:

$$\tilde{y}^{i*} = \lambda_i + \Phi^{-1}(1 - \tilde{\xi}^*) \frac{\sigma_i^2}{\sigma} .$$

Letting $\tilde{\sigma}_i = \dfrac{\sigma_i^2}{\sigma}$, we can write the above equivalently as follows:

$$\Phi(1 - \tilde{\xi}^*) = \Pr(\tilde{x}^i \le \tilde{y}^{i*}) ,$$

where $\tilde{x}^i \sim N(\lambda_i, \tilde{\sigma}_i^2)$. This concludes the proof of Proposition 2.

4. The Value of Information Sharing

4.1. Define the Value of Information sharing

We have thus far obtained two value functions associated respectively with two risk pooling structures of an MMI system, namely, the value function φ of an unequal-risk system $\{y^*, \bar{\xi}^*\}$ versus the value function $\tilde{\varphi}$ of an equal-risk system $\{\tilde{y}^*, \tilde{\xi}^*\}$. By definition, a value function represents the total value (a cost in this case) of a Club optimally operated under a specific risk pooling structure, either unequal-risk pooling or equal-risk pooling. In this sense, the difference between the two value functions can be used as a measure of the value (or the prize) for information sharing, especially the difference in the two value functions when evaluated at their respective optimal base-stock positions. For this purpose, we define the value of information sharing as:

$$P_{\text{mutual}} \equiv \varphi(y^*) - \tilde{\varphi}(\tilde{y}^*), \qquad (8^*)$$

where y^* and $\tilde{y}^*$ are the optimal base-stock positions respectively under unequal- and equal-risk.

First, we need to be assured that a nonnegative difference between the two value functions exists (i.e., $\varphi - \tilde{\varphi} > 0$, and therefore $P_{\text{mutual}} = \varphi^* - \tilde{\varphi}^* > 0$). Intuitively, the difference is nonnegative (i.e.,), since equal-risk pooling should reduce the total insurance cost. Now let us ascertain this intuition using rigorous analysis. Since $\tilde{\sigma}_i < \sigma_i$ for $n > 1$, we can write $\sigma_i = \tilde{\sigma}_i + \Delta\sigma_i$ for some $\Delta\sigma_i > 0$. Then, the HJB equation of an unequal-risk $\{y^*, \bar{\xi}^*\}$ can be derived from system (4) as follows:

$$r\varphi - \left(\sum_{i=1}^{n} \tfrac{1}{2}\left(\tilde{\sigma}_i^2 + (\Delta\sigma_i)^2\right)\frac{\partial^2}{\partial y_i^2}\right)\varphi - H(y, D\varphi) = 0\,. \qquad (9)$$

Or, equivalently, we can write the above as:

$$\tilde{F}(y, \varphi, D\varphi, \tilde{A}\varphi) - \left(\sum_{i=1}^{n} \tfrac{1}{2}(\Delta\sigma_i)^2 \frac{\partial^2}{\partial y_i^2}\right)\varphi = 0\,,$$

where $\tilde{F}(\cdot) = r\varphi - \left(\sum_{i=1}^{n} \tfrac{1}{2}\tilde{\sigma}_i^2 \frac{\partial^2}{\partial y_i^2}\right)\varphi - H(y, D\varphi)$ is an augmented Hamiltonian for the equal-risk MMI system $\{\tilde{y}^*, \tilde{\xi}^*\}$, of which the HJB equation can be written as:

$$\tilde{F}(y, \tilde{\varphi}, D\tilde{\varphi}, \tilde{A}\tilde{\varphi}) = 0\,.$$

With $\frac{\partial^2 \varphi}{\partial y_i^2} > 0$ and $\Delta\sigma_i > 0$ for some *i*, the HJB equation (9) of an unequal-risk MMI system can immediately verify that

$$\tilde{F}(y, \varphi, D\varphi, \tilde{A}\varphi) > \tilde{F}(y, \tilde{\varphi}, D\tilde{\varphi}, \tilde{A}\tilde{\varphi})\,.$$

The fact that the Hamiltonian *H* is identical for both unequal-risk and equal-risk systems implies that φ and $\tilde{\varphi}$ can only differ by an additive functional term; i.e., $\varphi = \tilde{\varphi} + \delta$ for some functional term $\delta \geq 0$. With this, the value of information sharing can be analytically measured by δ. In principle, the exact expression of this price term δ can be obtained by solving for φ and $\tilde{\varphi}$ from the respective HJB equations of the unequal-risk MMI system $\{y^*, \bar{\xi}^*\}$ and the equal-risk MMI system $\{\tilde{y}^*, \tilde{\xi}^*\}$, respectively. However, closed-form solutions are often unattainable, and even numerical solutions are still too complex to be tractable using numerical methods. In the Proposition below, we derive a more tractable lower-bound function $\underline{\delta}$, which gives the least cost difference caused by information privacy.

PROPOSITION 3. *Let φ and $\tilde{\varphi}$ be the respective value functions under the unequal-risk and equal-risk MMI systems with a strictly convex Lagrangian L. Then, for each non-impulse interval*

$[k, k+1)$ *for all* $k \in T_q$, *the following holds true:*

1) *The value functions* φ *and* $\tilde{\varphi}$ *are strictly convex for* $t \in [k, k+1)$ *; i.e.,* $\frac{\partial^2 \varphi}{\partial y_i^2} > 0$ *and* $\frac{\partial^2 \tilde{\varphi}}{\partial y_i^2} > 0$ *for all i.*
2) *There exists a functional* $\delta > 0$, *such that* $\varphi = \tilde{\varphi} + \delta$. *Therefore,* $P_{\text{privacy}} = \delta^* > 0$, *where* $\delta^* = \varphi(y^*) - \tilde{\varphi}(\tilde{y}^*)$.
3) *Given* φ *and* $\tilde{\varphi}$, *then* $\delta = \varphi - \tilde{\varphi}$ *has a nonnegative Hamiltonian, specifically,*

$$F(y, \delta, D\delta, \mathsf{A}\delta) = \left(\sum_{i=1}^{n} \tfrac{1}{2} (\Delta\sigma_i)^2 \frac{\partial^2}{\partial y_i^2} \right) \tilde{\varphi} > 0,$$

where $\Delta\sigma_i \equiv \sigma_i - \tilde{\sigma}_i$.

Proof. Item 1 above is a proven result in control theory (Fleming and Soner, 2006), and its proof is thus omitted. Given $\frac{\partial^2 \varphi}{\partial y_i^2} > 0$ for all *i*, item 2 of Proposition 3 can be verified from the HJB equation (9). To prove item 3 of the Proposition, we obtain from (9) the following variational inequality:

$$F(y, \varphi, D\varphi, \mathsf{A}\varphi) > \tilde{F}(y, \varphi, D\varphi, \tilde{\mathsf{A}}\varphi),$$

Where $F(\cdot)$ differs from $\tilde{F}(\cdot)$ only in the second-order operator A. Then, using $\varphi = \tilde{\varphi} + \delta$, we can verify from HJB (9) that

$$F(y, \tilde{\varphi}, D\tilde{\varphi}, \mathsf{A}\tilde{\varphi}) + F(y, \delta, D\delta, \mathsf{A}\delta) = 0.$$

Noting that $\tilde{F}(y, \tilde{\varphi}, D\tilde{\varphi}, \tilde{\mathsf{A}}\tilde{\varphi}) = 0$ and $\mathsf{A} = \tilde{\mathsf{A}} + \sum_{i=1}^{n} \tfrac{1}{2} (\Delta\sigma_i)^2 \frac{\partial^2}{\partial y_i^2}$, we can write the above HJB equation as follows:

$$F(y, \delta, D\delta, \mathsf{A}\delta) - \left(\sum_{i=1}^{n} \tfrac{1}{2} (\Delta\sigma_i)^2 \frac{\partial^2}{\partial y_i^2} \right) \tilde{\varphi} = 0.$$

Noting that $\frac{\partial^2 \tilde{\varphi}}{\partial y_i^2} > 0$ and $\Delta\sigma_i > 0$ for $n > 1$, the proof of item 3 of Proposition 3 is immediate from the above equality. With this, we conclude the proof.

4.2. Computing the Value of Information Sharing

In summary, the exact value of information sharing, denoted by $V_{\text{mutual}} \equiv \varphi(y^*) - \tilde{\varphi}(\tilde{y}^*)$, can be computed as follows:

1) Solve the unequal-risk and equal-risk HJB equations respectively for φ and $\tilde{\varphi}$.
2) Obtain the respective optimal base-stock positions, y^* (unequal-risk) and $\tilde{y}^*$ (equal-risk).
3) Then compute the value of information sharing, i.e., $V_{\text{mutual}} = \varphi(y^*) - \tilde{\varphi}(\tilde{y}^*)$.

The above solution for the exact price of privacy requires the solving of two HJB equations, one for the unequal-risk system and the other for the equal-risk system. These equations mostly require complex numerical methods to solve them. However, using Proposition 3, we can construct an approximate solution, which requires much less computation. The scheme for the approximate method of solution is described below:

1) Solve for the equal-risk $\widetilde{\varphi}$ for its HJB equation, and obtain $\widetilde{y}^*$.
2) Determine:

$$\Delta V^* = \left(\sum_{i=1}^{n} \tfrac{1}{2} (\Delta \sigma_i)^2 \frac{\partial^2}{\partial y_i^2} \right) \widetilde{\varphi}(\widetilde{y}^*).$$

3) Determine a linear functional difference $\hat{\delta} \in C^1$ with $D^2 \hat{\delta} = 0$ by solving the following first-order PDE system:

$$r\hat{\delta} - H(y, D\hat{\delta}) - \Delta V^* = 0 \text{, subject to: } \frac{\partial^2 \hat{\delta}}{\partial y_i^2} = 0 \text{ for all } i.$$

4) Then compute the approximate value of information sharing $\hat{V}_{\text{mutual}} = \hat{\delta}(\widetilde{y}^*)$.

The rationale of the above approximation is to seek the approximate difference in the form of a linear function (i.e., $D^2 \hat{\delta} = 0$), so as to avoid solving the HJB equations twice.

4.3. Value of Information Sharing and Volatility of Risk

Compared with the computation of the value of information sharing, a more important and interesting topic concerns the characteristics of the value of information sharing, i.e., $V_{\text{mutual}} = \varphi(y^*) - \widetilde{\varphi}(\widetilde{y}^*)$. By Item 3 of Proposition 3, the characteristics of the information sharing value are characterized in the corresponding Hamiltonian $F(y, \delta, D\delta, \mathsf{A}\delta)$, which has been obtained in Proposition 3 as:

$$F(y, \delta, D\delta, \mathsf{A}\delta) = \left(\sum_{i=1}^{n} \tfrac{1}{2} (\Delta \sigma_i)^2 \frac{\partial^2}{\partial y_i^2} \right) \widetilde{\varphi} > 0,$$

where $\Delta \sigma_i \equiv \sigma_i - \widetilde{\sigma}_i$ is the variability differential between non-risk pooling and equal-risk pooling. For the 'ideal' case of homogeneous membership with i.i.d. individual claim processes with identical $\sigma_i = \overline{\sigma}$ for all $i = 1, \cdots, n$, the variability differential can be determined as $\Delta \sigma_i = \sigma_i - \widetilde{\sigma}_i = (1 - \frac{1}{\sqrt{n}})\overline{\sigma}$. In terms of the Hamiltonian characterization of $V_{\text{mutual}} = \varphi(y^*) - \widetilde{\varphi}(\widetilde{y}^*)$, it can immediately be seen that the value of information sharing increases along with the average variability $\overline{\sigma}$ and the size of the Club *n*. Both parameters $\overline{\sigma}$ and *n* are measures or indicators of the volatility of underlying risks in terms of aggregate claims. With the finding in this paper that homogeneity facilitates optimal realization of the value of information sharing, it is without loss of generality to conclude that the more volatile the insurance risk is, the more competitive the mutual insurance becomes.

5. Findings and Implementation

First, let us summarize the useful observations and managerial implications that can be drawn from the results we have obtained so far in this paper, especially from Proposition 5.

5.1. Findings and Implications

1) The price of information privacy is mainly dependent on $\Delta \sigma_i = \sigma_i - \widetilde{\sigma}_i$, where $\widetilde{\sigma}_i = \frac{\sigma_i^2}{\sigma}$ and $\sigma = \sqrt{\sum_{i=1}^{n} \sigma_i^2}$. If the claims are deterministic ($\Delta \sigma_i = 0$), then the price of privacy would vanish, that is, there would be no difference between an unequal-risk and an equal-risk information structure without regard to the degree of heterogeneity of the members.
2) A unified tonnage-based premium policy can be justified only if the tonnage of a vessel is linearly

associated with the variability of the claims incurred by the vessel. This finding suggests that a rigorous and intensive statistical study of the correlation between tonnage and its claims needs to be conducted, so as to determine whether or not a tonnage-based premium policy is justifiable.

3) Under an unequal-risk information structure, the total cost minimizing premium policy entails an unequal-risk pooling scheme among the heterogeneous members of a P&I Club. Under the optimal unequal-risk premium policy, the individual's share of risk is solely determined by its actual claims record.
4) Given the fact that the tonnage-based premium policy has been practiced under unequal-risk information in P&I Clubs since their establishment 150 years ago, let us suppose that the tonnage-base premium policy is justified (i.e., that tonnage is linearly associated with the variance of claims). The above findings 1), 2), and 3) then explain the phenomenon in marine insurance that vessels of a similar tonnage tend to join the same P&I Club.

5.2. Implementation of Equal-Risk Pooling

The principle of information sharing implies the equal-sharing of risks. Noting that the equal-pooling of risk can be attained with an optimal premium policy under an equal-risk information structure, we only need to examine how to implement an unequal-risk equal-risk pooling scheme; i.e., equal-risk pooling in an unequal-risk MMI system $\{y^*, \bar{\xi}^*\}$. Two scenarios for implementing equal-risk pooling in an unequal-risk MMI system can be immediately considered: One with an equal-average level of risk $\bar{\xi}^* = \frac{1}{n}\sum_{i=1}^{n}\xi^{i*}$, where ξ^{i*} is as that determined in (5), and the other with an equal-level of risk $\tilde{\xi}^*$. In what follows, we examine the details of the implementation of unequal-risk equal-risk pooling.

The insurance threshold position for each member *i* under the equal-average risk $\bar{\xi}^*$ scheme, denoted by $\bar{y}^i(\bar{\xi}^*)$, can be determined to be:

$$\bar{y}^i(\bar{\xi}^*) = \lambda_i + \Phi^{-1}(1-\bar{\xi}^*)\sigma_i .$$

In vector form, we write the equal-average risk pooling position as $\bar{y}(\bar{\xi}^*) = (\bar{y}^1(\bar{\xi}^*),\cdots,\bar{y}^n(\bar{\xi}^*))$. From equation (5), we can determine that in general the equal-average risk pooling position $\bar{y}(\bar{\xi}^*)$ and the optimal unequal-risk pooling position y^* differ (i.e., $\bar{y}(\bar{\xi}^*) \neq y^*$). Individually, some would have a higher threshold position (i.e., $\bar{y}^i(\bar{\xi}^*) > y^{i*}$) if $\bar{\xi}^* < \xi^{i*}$, and some may have a lower threshold position (i.e., $\bar{y}^i(\bar{\xi}^*) < y^{i*}$) if $\bar{\xi}^* > \xi^{i*}$.

Since the total insurance cost is minimized under an optimal threshold y^*, we can conclude that the total cost under an equal-average risk scheme, denoted by $\varphi(\bar{y})$, can be no less than the unequal-risk value function $\varphi(y^*)$ (i.e., $\varphi(\bar{y}(\bar{\xi}^*)) \geq \varphi(y^*)$). This suggests that, in terms of total Club cost, under an unequal-risk MMI system it is worse to implement an equal-average risk pooling scheme than it is to implement an optimal unequal-risk pooling scheme.

Now let us examine what happens when an equal-equal-risk risk level $\tilde{\xi}^*$ is implemented for the unequal-risk MMI system. In this case, the individual threshold position, denoted by $\bar{y}^i(\tilde{\xi}^*)$, can be determined as:

$$\bar{y}^i(\tilde{\xi}^*) = \lambda_i + \Phi^{-1}(1-\tilde{\xi}^*)\sigma_i .$$

Using similar arguments, we can show that $\varphi(\bar{y}(\tilde{\xi}^*)) \geq \varphi(y^*)$. Thus, we can conclude that in terms of total insurance cost, an unequal-risk pooling system is better for the allocation of premiums in an of total insurance cost, an unequal-risk pooling system is better for the allocation of premiums in an unequal-risk MMI system.

6. Conclusion

It is concluded that an open policy or equal-risk information can lead to a more efficient MMI system overall for society, and to a greater degree of fairness and information sharing for the insured. The paper determines that information sharing can achieve best social welfare as well as efficient operation of a P&I Club. The study provides a scientific basis for future legislation on MMI competition law. The conclusion provides a scientific basis for both managerial strategy and competition regulation.

Acknowledgement

It is to acknowledge that the research is supported by The Hong Kong Polytechnic University under project No. B-Q909.

References

Asmussen, S. and Taksar, M. (1997), Controlled diffusion models for optimal dividend pay-out. Insurance: Mathematics and Economics 1997; 20; 1–15.

Aubin, J. (2000), Optimal impulse control problems and quasi-variational inequalities thirty years later: a viability approach. Optimal Control and Partial Differential Equations, 311-324. IOS Press.

Bennett, P. (2000), Mutuality at a Distance? Risk and Regulation in Marine Insurance Clubs. Environment and Planning A 2000; 32; 147-163.

Bensoussan, A., Lions, J. (1984), Impulse control and quasi-variational inequalities. Kent, England: Trans-Inter-Scientia, 1984.

Cabrales, A., Calvó-Armengol, A. and Jackson, M.O. (2003), La Crema: A Case Study of Mutual Fire Insurance. Journal of Political Economy 2003; 111; 425-458.

Cass, D., Chichilnisky, G. and Wu, H. (1996), Individual Risk and Mutual Insurance. Econometrica; 1996; 64; 333-341.

Dover, V. (1975), A handbook to marine insurance. 8th edition. UK: Witherby & Co Ltd; 1975.

European Commission (1985), O.J.L376/2 1985.

European Commission (1999), O.J.L125/12 1999.

Fleming, W. H., Soner, H. M. (2006) Controlled Markov processes and viscosity solutions. 2nd Edition, New York, N.Y.: Springer-Verlag, 2006.

Garner, B.A. (1999), Black’s Law Dictionary. 7th Edition. Minn.: West Group; 1999.

Gold, E. (2002), Gard Handbook on P&I Insurance. Arendal: Gard; 2002.

Golish, H. (2005), Marine Insurance – 2005/06. Drewry Shipping Consultants Ltd: London; 2005.

Gollier, C. and Wibaut, S. (1992), Portfolio Selection by Mutual Insurance Companies and Optimal Participating Insurance Policies. Insurance: Mathematics and Economics 1992; 11; 237-245.

Gyselen, L. (1999), P&I Insurance: the European Commission's Decision Concerning the Agreements of the International Group of P&I Clubs. Marine Insurance at the Turn of the Millennium, Antwerp: Intersentia; 1999.

Hazelwood, S.J. (2000), P&I Clubs: Law and Practice. 3rd edition. LLP Professional Publishing: London; 2000.

JLT (2003), A Fragile Stability – Insurance Market Overview. Jardine Lloyd Thompson: London; 2003.

Johnstad, T. (2000), Mutual Maritime Insurance Clubs: Co-operation and Competition. Annals of Equal-Risk and Cooperative Economics 2000; 71; 525-555.

Kavadias, S. and Loch, C.H. (2003), Optimal Project Sequencing with Recourse at a Scarce Resource. Production and Operations Management 2003; 12; 433-444.

Kulkarni, S.S., Magazine, M.J. and Raturi, A.S. (2004), Risk Pooling Advantages of Manufacturing Network Configuration. Production and Operations Management 2004; 13; 186-199.

Lamm-Tennant, J. and Starks, L.T. (1993), Stock Versus Mutual Ownership Structures: the Risk Implications. The Journal of Business 1993; 66; 29-46.

Ligon, J.A. and Thistle, P.D. (2005), The Formation of Mutual Insurers in Markets with Adverse Selection. Journal of Business 2005; 78; 529-555.

Liu, J.J. (2004), Coupled Impulse Feedback in Supply Order Fulfillment System. Work Paper Series 08-2004, LGT-HKPolyU, Hong Kong; 2004.

Macey, J.R. (2004), Efficient capital markets, corporate disclosure, and Enron. Cornell Law Review 2004; 89; 394-422.

Polyanin, A.D. and Zaitsev, V.F. (1995), Handbook of Exact Solutions for Ordinary Differential Equations. CRC Press; 1995.

Siegl, T.F. and Tichy, R.F. (1999), A process with stochastic claim frequency and a linear dividend barrier. Insurance: Mathematics and Economics 1999; 24; 51–65.

Smith, B.D. and Stutzer, M.J. (1990), Adverse Selection, Aggregate Uncertainty, and the Role for Mutual Insurance Contracts. The Journal of Business 1990; 63; 493-510.

Smith, B.D. and Stutzer, M.J. (1995), A Theory of Mutual Formation and Moral Hazard with Evidence from the History of the Insurance Industry. The Review of Financial Studies 1995; 8; 545-577.

Taksar, M.I. (2000), Optimal Risk and Dividend Distribution Control Models for an Insurance Company. Mathematical Methods of Operations Research 2000; 1; 1-42.

Tapiero, C.S., and Jacque, L. (1987), The Expected Cost of Ruin and Insurance Positions in Mutual Insurance. Journal of Risk & Insurance 1987; 54; 594-602.

Strict or relaxed: remedy for duty of disclosure

Kevin Li[1], *Yulan Wang*[2] *and Jie Min*[3]

Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hong Kong.

Abstract

One hundred years ago, when signing the marine insurance contract, the underwriter was not well-informed and in a disadvantageous position. As a result, Marine Insurance Law of UK 1906 regulates a strict ex post remedy with respect to the duty of disclosure. Today it has become the most significant regulation in the jurisdiction of the marine insurance disputes, widely referred by many countries during drafting the marine or non-marine insurance laws. But with the development of the shipping industry and the insurance market, the underwriter now has much better understanding of the IoM for the business. And Marine Insurance Law 1906 is believed outdated regarding the duty of disclosure. The amendment of the related clauses has been questioned by the academia, the legislators and the shipping industry.

In this paper, we classify the IoM into the three categories: (1) important for the underwriter only; (2) important for the policyholder only; (3) important for both parties. Based on the disclosed IoM, the underwriter uses a stochastic control model to price the premium rate. And the policyholder may provide the biased IoM information to obtain a low price. Through the analysis of the policyholder's utility before and after the ex post remedy, the weaknesses of the current marine insurance law are revealed. Some amendment suggestions are also provided in this study.

1. Introduction

Marine insurance is the oldest type of insurance born in UK. Marine insurance laws adopted in the world, currently, are mainly *Marine Insurance Law of UK 1906* or her derivative versions. With the development of shipping industry, the remedy of duty of disclosure regulated by the law seems so strict for policyholders. Assureds have mandatory obligation to disclose any information of materiality to marine insurers, otherwise, underwriters have rights to discharge insurance contract, refuse to cover the loss and refund the extra premium, if it is witnessed policyholders conceal some information. Since it is difficult for policyholders to identify whether the information is of materiality, the court will adopt the declaration done by underwriters. As a result, marine insurers have great motivation to reject their coverage obligations intentionally. Legal scholars advocate that the old version of marine insurance law is outdated, where the benefit of policyholders is not protected appropriately.

In marine insurance practice, the clauses about duty of disclose have become the pretext of underwriter to terminate contract maliciously, which makes people more dissatisfied with the present marine insurance law, especially the legal consequence of violation. In the case Mackay v. London

[1] Corresponding author: Associate Professor of Department of Logistics and Maritime Studies, M640, Faculty of Business, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, P. R. China. Tel: +852-2766-7919; Fax: +852-2330-2704; E-mail: K.X.Li@inet.polyu.edu.hk.
[2] Secondary author: Visiting Assistant Professor of Department of Logistics and Maritime Studies, M507n, Faculty of Business, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, P. R. China. Tel: +851-3400-3612; Fax: +852-2330-2704; E-mail: Yulan.Wang@inet.polyu.edu.hk.
[3] Third author: Ph.D Candidate of Department of Logistics and Maritime Studies, Faculty of Business, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, P. R. China. Tel: +852-2766-4070; Fax: +852-2330-2704; E-mail: 06900431r@polyu.edu.hk.

General Insurance Co, Lord Swift appealed to people for more attention on assured's disadvantageous situation under the present clauses of duty of disclosure. Assured is suffering rude attitude of underwriter. After collecting premium, underwriter rejects to cover the claim, even though the loss event is definitely not influenced by the information concealed by assured. Accordingly, from 1950s, the common law system countries, including UK, try to relax the strict request about duty of disclosure. After 100-year development, steal hull, advanced engineer, electronic equipment, high-level cargo packing technology, especially containerization, advanced navigation and security system are widely used in shipping industry. The risk level of vessel and cargo on board is significantly reduced. Underwriter is facing the reducing insurance risk caused by violating duty of disclosure, where insurer can employ developed acceptance technology, communication technology and special human resource to investigate the actual situation of insurance bids and estimate risk more precisely. A famous scholar in UK questioned if it is still applicable to conclude the rights and obligations of the two parties in the insurance contract in the 21st century basing on the dated case 250 years before.

(1) Duty of disclosure is a concept of marine insurance law. Law and the quality of its enforcement are potentially important determinants of what rights policyholders have and how well these rights are protected (see in la Porta et al, 1998). The purpose of this research is to find a better revision of duty of disclosure in the present marine insurance law. The whole procedure of marine insurance business can be divided into three steps: (1) the formation of insurance contract; (2) the occurrence of marine insurance disputation; (3) the remedy of law.

Intuitively, an underwriter and a policyholder would not accept a certain insurance contract, unless they predict that sufficient benefit can be satisfactorily obtained through such agreement. We can say that the contract is *expectedly fair*. Usually, policyholder is the more informed party rather than underwriter. When the marine insurance disputation happens, it must imply there is the actual unfairness contained in the contract. Such bias deviating from the expected fairness should be redressed by law. At the mean time, the efficiency and the enforcement of the law depend on the objective of legislation.

2. Literature Review

The issue of duty of disclosure is essentially connected to the information asymmetry existing in insurance business. The common view is asymmetric information are plaguing insurance market and reducing the social wealth. However, Garidel-Thoron (2005) obtains the opposite conclusion. By a two-period model, Garidel-Thoron (2005) finds that when the initial contract choice of a policyholder cannot be shared among underwriter in the second period, a strict increase in welfare is obtained. More exactly, in the second period, the ex post welfare of the insurance contract is improved, which, however, decrease the efficiency of the ex ante competition. Finally, the overall welfare decreases in these two periods.

Generally, information disclosure is much discussed in corporate finance, but few appear in the literatures of insurance economics. Analogous to de Garidel-Thoron (2005), Li and Madarasz (2008) describe a scenario that expert provides advices to the uninformed decision maker. Since there are the conflicting interests between the two parties, an expert can provide a bias report, right-biased or left-biased. Li and Madarasz (2008) discussed the uncertainty of the conflicting direction and the uncertainty of the conflicting extent, separately. The main results suggest that the expert of lower conflicting interest will obtain higher expected payoff under disclosure, whereas the expert of high conflicting interest will achieve higher expected payoff under nondisclosure. The decision maker always obtain higher expected payoff under nondisclosure. When there is only the uncertainty of conflicting direction, nondisclosure will increase welfare. If the uncertainty of conflicting extent exists, the decision maker and the expert of low conflicting interests can improve their situations more under disclosure.

The conclusions of de Garidel-Thoron (2005) and Li and Madarasz (2008) imply that the absolute information symmetry is not improving the social welfare. De Garidel-Thoron (2005) constructs a dynamic model, while Li and Madarasz (2008) adopted a comparative static model. The former study offers a method to analyze the information asymmetry among multiple underwriters and the insurance strategies of distinct types of policyholders. The latter research takes into consideration of how the uncertainty of conflicting interest direction and severity makes impact on the distribution of social wealth, which reveals the fundamental occasion of information asymmetry. Such consequence supports our intuitive understanding of Marine Insurance Law 1906 that the information symmetry obtained by executing the extremely strict remedies fails to promote the further development of the contemporary marine insurance business. Gu (2007) suggests to revise Marine Insurance Law of China from three perspectives, where the main purpose is to relax the remedy regulation about duty of disclosure and, indirectly, allow the existence of information asymmetry to some extents. Simpler than Li and Madarasz (2008), the scenario we want to emphasize is one-side bias. Underwriter takes the role of the decision maker in Li and Madarasz (2008), and policyholder is an expert of more information than the uninformed underwriter. Obviously, the two parties are of conflicting interest. Any policyholders have great incentive to underestimate their risk level and perform overconfident.

Differing from de Garidel-Thoron (2005), a more subdivisible temporal model is constructed by Boot and Thakor (2001), where the three types of investors seek for information through costly transmission, or firm disclosure policy after some investor becoming informed. When the true value of the firm becomes known, the investors obtain payoff. Compared with the social welfare analysis done by de Garidel-Thoron (2005) and Li and Madarasz (2008), Boot and Thakor (2001) emphasize the benefit or the loss with respect to each type of firm and investor. Under disclosure, the firm of good quality is always better off and the bad-quality firm always worse off. Intuitively, in the insurance market, the low-risk individual is also better off under duty of disclosure than the high-risk one, who is thus intent to conceal his risk state to obtain the identical contract for a low-risk policyholder.

Our issue in this study is not only an economic problem, but also a legislation topic about how to revise marine insurance law to prevent the marine insurance contract from the despiteful violation done by some party of the contract. In the law academia, it is questioned that the targets of legislation are difficult to be affirmed among the conflicting goals, efficiency, fairness, good faith and the protection of individual autonomy. Marine insurance contract is typically a contract between two firms, either long-term or short-term. Contract law presumes, in principle, the two parties of a contract want to maximize joint gains, with which the observation is conflicting in the real world (see in Schwartz and Scott, 2003). Each party of a contract wants to maximize its profit as much as possible rather than maximize the joint gains of the both sides. It implies that the maximal joint gain of the contract is not necessarily the sum of the maximal profits of the two parties. Marine insurance is of its own unique characteristics. In this market, there are two types of underwriters, stock and mutual, with the absolutely opposite objective functions. The former type wants to maximize the profit and the latter type performs non-profitable.

The general principle of contract law is not always available to guarantee the insurance contracts. Insurance contracts are risk shifting contracts subject to a distinct and heavily regulated legal field. Due to the interpretation style, the court can possibly deviate from the correct answer. Schwartz and Scott (2003) suggest that the expectation of judicial errors is zero with variance. It implies that we can describe the court's interpretation through an adapted Brownian process without the loss of generality. In the modern insurance industry, the contract is written in a relatively standardized style, before which a policyholder has obligations to submit his private information to the underwriter through filling some standardized forms. There is possibly a huge gap between the knowledge mastered by policyholder and underwriter, which induces that contracts are often incomplete in relevant respects. Law, correspondingly, has to contribute to completing such contract by providing the plain interpretation.

3. Basic Assumption

Let Φ denote the set of information mastered by policyholders. In Φ, there consist of N aspects of information totally. Each policyholder can rank all of the information by the materiality. Let X_i denote the actual value of the i^{th} information. For instance, if we suppose the first information is the tonnage of a vessel, the actual size of this vessel is recorded by X_1 in scale of ton. Analogously, if the second information is the age of a vessel, X_2 can be 10 years or some other numbers. X_i denotes the i^{th} aspect of information.

Limited by the knowledge pool and the investigation capability, each underwriter is impossible to focus on all sorts of information in Φ. Basing on the private understanding, an underwriter believes that there are n aspects of information are of importance, $n \leq N$. Since the underwriter is the information disadvantage side in the formation of insurance contract, it is impossible for him to rank the n aspects of information before obtaining the information set through investigation. Let $\Phi_u = \{X_1, X_2, ..., X_n\}$ denote the IoM set that the underwriter focus on, where each aspect of information has the equal evaluation with respect to the materiality.

On the other hand, each policyholder knows all N aspects of information in Φ but does not consider that all of the information is of materiality. Suppose the policyholder ranks m aspects of information which are of materiality he believes. Let subscript (1) indicate the least important information and subscript (m) denote the most important one. The policyholder has the information set of materiality, $\Phi_p = \left\{X_{(1)}, X_{(2)}, ..., X_{(m)}\right\}$, where j is the queued index by the extent of importance, rather than the value of the information X, because each aspect of information has its own measure scale. Thus, all aspects of the information can be separated into four categories: (1) the information of no materiality, $\overline{\Phi_u \bigcup \Phi_p}$; (2) the information important for underwriter only, $\Phi_u - \Phi_u \bigcap \Phi_p$; (3) the information important for policyholder only, $\Phi_p - \Phi_u \bigcap \Phi_p$ (private information appearing in many studies); (4) the information important for both parties, $\Phi_u \bigcap \Phi_p$. An underwriter would like to investigate any information important for him, so the policyholder has to submit such information contained in Φ_u truly or falsely. As for the information belonging to the third category, policyholder can submit it voluntarily. If a policyholder prefers showing his good reputation and good faith, he discloses, truthfully, the information of materiality in his own understanding. Otherwise, he will conceal the information. There is no necessity for the policyholder to lie, when he decides to disclose. One extreme situation is $\Phi_u \bigcap \Phi_p = \phi$, that is, there is no overlap between the reorganizations of IoM between an underwriter and a policyholder. The policyholder does not consider that the information enquired by the underwriter is too important to conceal. The only decision done by the policyholder is whether to provide the IoM he believes. Another extreme situation is $\Phi_u \bigcap \Phi_p = \Phi_p$, which implies that any IoM known by a policyholder is also the information of materiality for the underwriter. The policyholder has to decide whether to offer the true story, while for the information contained in $\Phi_u \bigcap \overline{\Phi_p}$ the policyholder can directly give the precise value of the information. As for $\Phi_u \bigcap \Phi_p = \Phi_u$, it implies that all information in which an underwriter is interested is also the IoM capturing the attention of the policyholder.

Consider a policyholder has the IoM vector $\left(X_{(1)}, X_{(2)}, ..., X_{(m)}\right)$. The reason why the m^{th} aspect of information is ranked as the most important information is because the value of this piece of information indicates the highest possibility to induce a claim of insured loss, compared with other information. The policyholder also believes any information possible to induce an accident must be contained in the set of IoM Φ_p, while the information out of Φ_p is useless to estimate the actual loss probability and the actual loss size. Since the policyholder is the information advantageous side in an insurance contract, it is rationale to assume that the most precise loss estimation cannot be obtained unless the policyholder offers all information in Φ_p truthfully. However, the disadvantageously informed underwriter cannot affirmatively ask the whole correct questions. Suppose an underwriter involves luckily k questions contained in Φ_p. Then we have

Table 3.1 Information sets and disclosure bias

Section / Items	$\Phi_u - \Phi_p \bigcap \Phi_u$	$\Phi_p \bigcap \Phi_u$	$\Phi_p - \Phi_p \bigcap \Phi_u$
$\in \Phi_u$	√	√	
$\in \Phi_p$		√	√
No. of IoM**	$n-k$	k *	$m-k$
Disclose	Disclose Completely	Disclose with Bias	Disclose with Bias
Bias Ratio	= 0	$\neq 0$	$\neq 0$

* If k = 0, the underwriter ask all questions on the incorrect directions.

Define vector $\delta = (\delta_{[1]},.., \delta_{[n+m-k]})$ as the bias ratios related to the n aspects of information provided by the policyholder. Obviously, Table 3.1 shows that

$$\delta = \begin{cases} 0, & i = 1, 2, ..., n-k \\ \delta_{[i]} > 0, & i = n-k+1, n-k+2, ...n-k+m \end{cases}$$

The information sent by the policyholder and received by an underwriter is the message adjusted by the bias ratios. Each aspect of information value is multiplied by $1 + \delta_{[i]}$, correspondingly, where $\delta_{[i]}$ describes the distance between the disclosed value and the actual value of the same information. The bias ratios $\delta_{[i]}$ is not necessarily positive or negative, because the value of a certain information could be positively or negatively proportional to the loss size. For instance, generally either old or new-built vessel is not security enough to be sea-worthy. The most preferable age of a vessel should be falling into a certain middle range. Thus, if the policyholder wants to insure his old vessel, he intends to offer the age information with a negative bias to say that the age of his vessel is in the middle range, and it is safe and seaworthy. Hereby, the bias ratio is a negative rate. On the contrary, if the vessel is new-built, the bias ratio could be positive.

Accordingly, when a policyholder is asked to provide the IoM by the underwriter, he combines the bias into the disclosed information based on his own understanding. Recall that $j \in \{1, 2, ...m\}$ denotes the importance rank of the j^{th} information with respect to a policyholder. Hereafter, let the superscript (j) indicate that the i^{th} information $X_{[i]}^{(j)}$ is ranked as the j^{th} in the information set of the policyholder. The whole information submitted can be written as the vector

$$\left(\hat{X}_{[1]}, ..., \hat{X}_{[n+m-k]}\right) = \left(X_{[1]} \cdot \left(1 + \delta_{[1]}\right), ..., X_{[n+m-k]}^{(j)} \cdot \left(1 + \delta_{[n+m-k]}\right)\right)$$

where $\delta_{[i]} \propto \beta_{[j]} \cdot j$ for $i = n - k + 1, \ldots, n + m - k$, and the rest bias ratios are nil. The non-nil bias ratios imply that the bias ratio of a disclosed IoM depends on the rank of this piece of information in the evaluation of the policyholder. If the rank is high, the absolute value of the bias ratio is also large.

Let y denote the random loss of a normal distribution, which can be described by a stochastic diffusion process based on the Brownian motion. Let $\mu(\cdot)$ denote the arrival rate and $\sigma^2(\cdot)$ denote the variance rate. The loss y has the normal distribution with the mean $\mu \cdot \Delta t$ and the variance $\sigma^2 \cdot \Delta t$, where Δt denotes the time interval. An underwriter has to estimate the parameters basing on the statement of the policyholder and obtains $\mu_u\left(X_{[1]}\cdot\left(1+\delta_{[1]}\right),\ldots,X_{[n]}^{(\cdot)}\cdot\left(1+\delta_{[n]}\right)\right)$ and $\sigma_u^2\left(X_{[1]}\cdot\left(1+\delta_{[1]}\right),\ldots,X_{[n]}^{(\cdot)}\cdot\left(1+\delta_{[n]}\right)\right)$, while the actual loss is normally distributed with parameters $\mu_p\left(X_{[n-k+1]}^{(\cdot)}\cdot\left(1+\delta_{[n-k+1]}\right),\ldots,X_{[n+m-k]}^{(\cdot)}\cdot\left(1+\delta_{[n+m-k]}\right)\right)$ and $\sigma_p^2\left(X_{[n-k+1]}^{(\cdot)}\cdot\left(1+\delta_{[n-k+1]}\right),\ldots,X_{[n+m-k]}^{(\cdot)}\cdot\left(1+\delta_{[n+m-k]}\right)\right)$.

4. Formation of Insurance Contract

As known, there are two types of underwriter, stock and mutual, differing from each other on ownership. Each of them has his own particular objective. A stock underwriter seeks for the maximal profit to satisfy the requirement of stockholders, while a mutual underwriter is non-profitable so that the mutual maximizes the utilities of his members. Both of stock and mutual underwriters, however, have the common interest, that is, to minimize the ruin probability. At time t, every kind of underwriter collects premium q_t from the policyholders to construct the risk reserve R_t and cover the claims. When the risk reserve is less than nil, the underwriter is ruined. Let τ denote the ruin time, $\tau=\inf\left\{t:R_t\leq 0\right\}$, and the ruin probability is $\Pr\left(\tau<\infty\right)$.

When pricing the insurance contract, an underwriter can only estimate the possible loss distribution depending on the statement done by the policyholder. The underwriter believes that the risk reserve R_t is described by a diffusion process that

$$dR_t^u=\left(q_t-\mu_u\right)dt+\sigma_u dw_t^u$$

where w_t^u is an adapted Brownian motion. Unfortunately, the policyholder provides the IoM with bias of some extent, to which the underwriter mostly underestimates the risk level more or less, while the risk reserve is reduced by rapidly the unexpected claims of normal distribution $N\left(\mu_p,\sigma_p^2\right)$. The actual risk reserve variation is expressed by the following process

$$dR_t^p=\left(q_t-\mu_p\right)dt+\sigma_p dw_t^p$$

where w_t^p is an adapted Brownian motion. Let U denote the concave utility function of the policyholder, denoted with $U'>0$, $U''<0$. Suppose the policyholder has the fixed initial wealth ω. Intuitively, any price strategy of the insurance contract, q_t, has the cumulative value Q_τ satisfying the incentive condition

$$U\left(\omega-Q_\tau\right)\geq EU\left(\omega-y\right)$$

where $Q_\tau=\int_0^\tau q_t e^{-r_f t}dt$, r_f is risk free rate. The risk free rate is the interest rate that it is assured can be obtained by investing in financial instruments without default risk. When default risk is not taken into consideration as the main issue of a study, the assumption of risk free should be adopted without the loss of generality. The right side of inequality is the expected utility of a policyholder suffering the loss y without insurance protection. The cumulative premium collection must be less than a baseline, that is, $Q_\tau\leq\omega-U^{-1}\left[EU\left(\omega-y\right)\right]$. Define *the disclosed baseline* as $\omega-U^{-1}\left[E_uU\left(\omega-y\right)\right]$, estimated basing on the IoM with bias, and analogously define *the actual baseline* as $\omega-U^{-1}\left[E_pU\left(\omega-y\right)\right]$. Thus, the objective function of an underwriter is

$$
\begin{aligned}
&\min_{q_t} \Pr(\tau < \infty) \\
s.t.\quad & dR_t^u = (q_t - \mu_u)\,dt + \sigma_u dw_t^u \\
& R_0^u = x_u \\
& \tau = \inf\{t : R_t^u \le 0\}
\end{aligned} \tag{1}
$$

where x_u denotes the initial risk reserve. Let $V_1(x_u)$ be the value function of the minimal ruin probability, and we have $V_1(x_u) = \exp\left\{-\frac{2(q_t - \mu_u)}{\sigma_u^2} x_u\right\}$. When the initial risk reserve $x_u = 0$, the ruin probability equals to one. On the contrary, if the initial risk reserve $x_u \to \infty$, this underwriter will be never bankrupt. Given a certain level of the initial risk reserve, the ruin probability decreases with an increasing q_t, which implies an underwriter intends to collect premium at a high rate to prevent himself safely from the bankruptcy. The underwriter also knows that there is an upper limit of premium, beyond which the price of the insurance contract is unacceptable for a policyholder with the utility function U. Since the underwriter makes decision depending on the policyholder's statement, the cumulative amount of the collected premium Q_τ must be no more than the disclosed baseline. Suppose the underwriter prefers to dominate the minimal ruin probability at a certain extent v, $0 < v \le 1$. It is easily obtained $q_t = \mu_u - \sigma_u^2 \cdot \frac{\ln v}{2x_u}$. There are two scenarios: (1) $Q_\tau \le \omega - U^{-1}\left[E_u U(\omega - y)\right]$, where an acceptable premium rate q_t must fall in the range satisfying this inequality; (2) $Q_\tau > \omega - U^{-1}\left[E_u U(\omega - y)\right]$, where the contract is unacceptable. In the second scenario, if the policyholder still performs his interests in this insurance contract, the policyholder sends a signal to indicate his concealing in the information disclosure.

Let us further consider the situation without concealing, where all of IoM is disclosed by the policyholder truthfully. Let $\overline{\tau}$ denote the ruin time of the risk reserve with the complete information disclosure. The dynamic programming of the minimized ruin probability of an underwriter is rewritten as

$$
\begin{aligned}
&\min_{q_t} \Pr(\overline{\tau} < \infty) \\
s.t.\quad & dR_t^p = (q_t - \mu_p)\,dt + \sigma_p dw_t^p \\
& R_0^p = x_p \\
& \overline{\tau} = \inf\{t : R_t^p \le 0\}
\end{aligned} \tag{2}
$$

The solution of the stochastic control (2) is denoted by the value function $V_2(x_p)$. $V_2(x_p)$ has the structure similar with V_1 so that let us denote the premium rate as $\overline{q}_t = \mu_p - \sigma_p^2 \cdot \frac{\ln v}{2x_p}$. Correspondingly, the cumulative premium collection can be denoted by $\overline{Q}_{\overline{\tau}} = \int_0^{\overline{\tau}} \overline{q}_t e^{-r_f t} dt$. Actually, a policyholder with the full of IoM would like to accept any premium program $\overline{Q}_{\overline{\tau}} \le \omega - U^{-1}\left[E_p U(\omega - y)\right]$, and the policyholder can always be better off unless the underwriter collects the total amount of the premium exceeding $\omega - U^{-1}\left[E_p U(\omega - y)\right]$.

Hereby, the following two subsections indicate four scenarios contained in $\Phi_p \cup \Phi_u$. The formation of an insurance contract depends on the voluntary behavior of the two parties, demand and supply, both of which have their own target prices of contract respectively. It is subdivided into the following categories to discuss in detail.

(1) No insurance contract is available with truthfully disclosed IoM. In this scenario, the underwriter prefers to collect premium at the rate no less than $\overline{q}_t$, which exceeds the acceptable price of the policyholder. In order to get protection from the underwriter, the policyholder provides the IoM with bias to some extent or conceals the private knowledge of risk components directly.

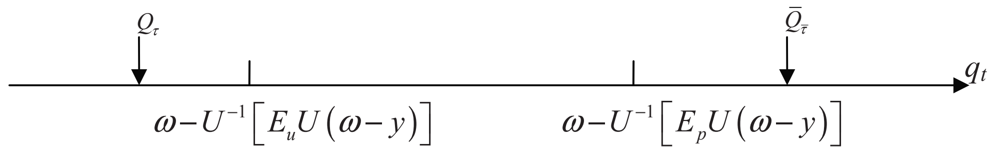


Figure 4.1a

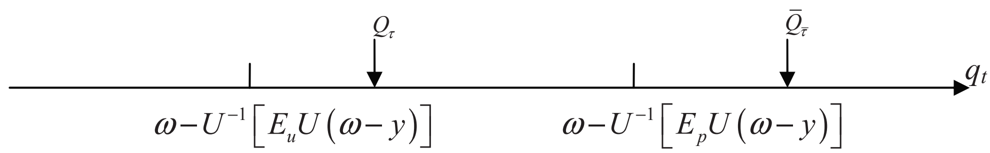


Figure 4.1b

An underwriter estimates the loss probability distribution through analyzing the IoM disclosed by the policyholder. By solving the minimization of the ruin probability, it is found that there is the lowest premium rate for the underwriter to keep the risk reserve at the safe level. If the situation in Figure 4.1a is hold, that is, the cumulative premium collection is lower than $\omega-U^{-1}\left[E_uU\left(\omega-y\right)\right]$, it must be the most favor price for the policyholder, because he obtains *the utility surplus* $U\left(\omega-Q_\tau\right)-E_pU\left(\omega-y\right)$. Especially, the stock underwriter intends to seek for a higher premium rate q_t as close as possible to the disclosed baseline $\omega-U^{-1}\left[E_uU\left(\omega-y\right)\right]$ and furthermore reduce the ruin probability, while the mutual underwriter prefers to adopt the lower limit of q_t and prevent his members from the extra burden of insurance cost.

On the contrary, if the situation in Figure 4.1b is hold, the cumulative premium collection is higher than $\omega-U^{-1}\left[E_uU\left(\omega-y\right)\right]$. Albeit the contract price is still attractive for the policyholder at this moment, it is also dangerous and irrational to accept this contract at the very premium rate. It is easily understood that there is the risk to expose the policyholder's fraud that the policyholder provides the false IoM unless he does not perform any interests in this premium rate. Thus, in the scenario of Figure 3.2b, there is no insurance contract formed.

(2) Insurance contract can be formed with the truthfully disclosed IoM. Albeit a policyholder provides the IoM with bias, the underwriter is still able to find a contract of the premium rate larger than or equal to $\overline{q}_t$, and, simultaneously, the premium rate is less than the actual baseline of the policyholder.

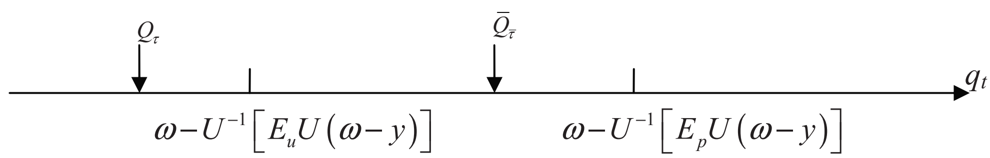


Figure 4.2a

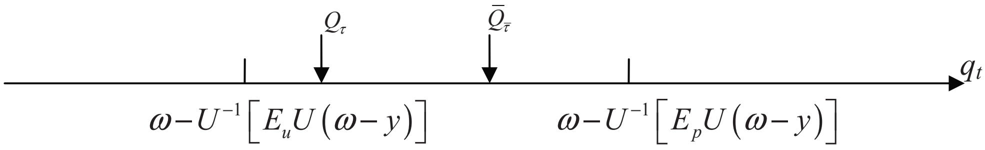


Figure 4.2b

The essential characteristic of these two situations (Figure 4.2a and Figure 4.2b) is that there exists a premium rate priced on the truthfully disclosed IoM. If the situation in Figure 4.2a holds, the insurance contract is constructed at the lower premium, and the policyholder benefits from concealing the true IoM. *The utility surplus of the policyholder is the distance of the utilities under two alternative prices*. In Figure 4.2b, when the policyholder fails to purchase insurance contract priced on the biased IoM, another alternative is available, which is also attractive below the actual baseline of the policyholder. In this situation, the policyholder is unable to obtain any benefit from disclosing the biased IoM, while the premium rate priced on the actual IoM is also acceptable. Thus, the preferred option of this policyholder is to disclose all IoM truthfully.

Briefly, among the four situations, the case in Figure 4.1b is no insurance contract formed, no matter how severe the bias of the information disclosure is. Thus, there is no legal dispute of such a non-existing insurance contract. The case in Figure 4.2b is the insurance contract formed and priced on the actual risk state. This is the most ideal situation for an underwriter, also most security for a policyholder. The policyholder provides the accurate IoM, on which the underwriter bases to price the contract. The policyholder completely performs the duty of disclosure, and the underwriter has unshakable obligations to cover the loss happened to his client.

The case in Figure 4.1a represents the situation that if a policyholder disclosed the unbiased IoM, there would be no underwriter willing to supply the insurance contract. The policyholder, accordingly, has motivation to disclose the biased IoM to make sure the insurance contract is formed at a worthy price. By contrast, the case in Figure 4.2a, essentially, reflects the tradeoff between two premium alternatives. The target of the policyholder, hereby, is not only to find an underwriter offering the loss coverage, but also to reduce the insurance cost through distorting the actual loss distribution intentionally.

5. Social Welfare and Amendment of Marine Insurance Law

Accordingly, there are only two scenarios, which should be regulated precisely by the insurance law, especially, with respect to the duty of disclosure. They are the cases in Figure 4.1a and Figure 4.2a. Under the biased disclosure of IoM, a policyholder can be benefited from the lower premium rate and the indemnity. Since the underwriter is the information disadvantage side, the insurance law regulates a strict remedy to protect the right of the underwriter. If one arbitrarily concluded that the target of the insurance law was to compensate the loss of the less informed underwriter, it ought to be misunderstood the fundamental mission of legislation, that is, to maximize the social welfare, or say, to minimize the loss of the social welfare.

In the pricing model in Section IV, the underwriter presumes the policyholder is trustable and prices a long-run insurance contract with this policyholder at the equilibrium premium rate q_t. The stochastic process of the risk reserve implies connotatively that the premium is paid continuously, where from $t = 0$ to $t = \tau$, it is a long-run picture. This premium rate ensures the risk reserve is worn out with the proposed low probability v. The legal issues faced to the insurance law, however, are usually the disputes in the short run. Since the continuous premium collection only exists in the theoretical circumstance, the premium actually is levied periodically, yearly, monthly, or daily, for instance. Without the loss of the generality, consider the unit time interval randomly, $\left[T, T+1\right]$, where

$0 \le T < T+1 \le \tau$. At time *T*, the premium $Q_T = \int_T^{T+1} q_t e^{-r_f(t-T)} dt$ is collected to cover the claims during *T* to *T* + 1. Suppose at time $T^* \in [T, T+1]$, the loss y_{T^*} comes out which is of the normal distribution $N\left(\mu_p \cdot \left(T^* - T\right), \sigma_p^2 \cdot \left(T^* - T\right)\right)$.

Table 5.1 shows the expected and actual utility surplus of the policyholder before and after the ex post remedy. Vertically, if all of the IoM is disclosed truthfully, we have two categories: (1) no insurance contract can be formed; (2) an insurance contract can be formed at a highly priced premium rate. Horizontally, the ex post utilities are discussed, respectively, with respect to the possible inducements belonging to the two information sets.

Table 5.1 Utility of the policyholder under the current duty of disclosure

		No Insurance Contract with the Fully Disclosed IoM	Highly Priced Insurance Contract with the Fully Disclosed IoM
Ex ante expected utility surplus of the policyholder		$U(\omega - Q_\tau) - E_p U(\omega - y)$	$U(\omega - Q_\tau) - U(\omega - \bar{Q}_{\bar{\tau}})$
Ex post utility	$X_{[n-k+1]}^{(\cdot)}, \dots, X_{[n]}^{(\cdot)}$	$U\left(\omega - Q_T - y_{T^*}\right)$	$U\left(\omega - Q_T - y_{T^*}\right)$
	$X_{[n+1]}^{(\cdot)}, \dots, X_{[n+m-k]}^{(\cdot)}$	$U\left(\omega - Q_T - y_{T^*}\right)$	$U\left(\omega - Q_T - y_{T^*}\right)$

Under the strict remedy, an underwriter has the right to reject to cover the claims and refuse to refund the collected premium, if it is witnessed that the policyholder concealed the IoM or disclosed the biased IoM in the formation of the insurance contract, where the IoM is mostly identified by the statement of the underwriter submitted to the court. Table 5.1 shows that if the policyholder provides the biased information, the strict remedy results in the unique consequence, that is, the policyholder pays the premium (albeit it is cheaper) to the underwriter for a *self-insurance* contract. The weaknesses of the strict remedy with respect to the duty of disclosure are obvious:

(1) The remedy fails to distinguish the situation of highly priced contract from the scenario of no insurance contract formed. From Table 5.1, it is easy to find that the surplus of the former situation is less than the latter one, $U(\omega - Q_\tau) - U(\omega - \bar{Q}_{\bar{\tau}}) < U(\omega - Q_\tau) - E_p U(\omega - y)$. It implies that the policyholder obtains the different levels of the illegal benefit from the violation of the duty of disclosure. But from the ex post remedy, such difference cannot be reflected. Obviously, the motivations of the policyholders in these two scenarios are quite discriminative. If there is no insurance contract available in the normal market, the incentive of the policyholder to conceal or lie is the formation of the insurance contract. On the other hand, if there is a highly priced contract formed, the purpose of providing the biased information is to obtain a relatively low premium rate.

(2) The remedy takes sides with the underwriter through the identification of IoM in the court. The IoM in $X_{[n+1]}^{(\cdot)}, \dots, X_{[n+m-k]}^{(\cdot)}$ is mastered by the policyholder, and the underwriter does not request it during the formation of the contract. Under the current regulation, if the inducement of the loss event does not belong to $X_{[n+1]}^{(\cdot)}, \dots, X_{[n+m-k]}^{(\cdot)}$, but the IoM in this set is concealed by the policyholder and proved to be of materiality for the policyholder, the underwriter still has right to uncover the loss and refuse to refund the premium in the excuse of the duty of disclosure. In another words, it is questioned whether the policyholder should be responsible for the information asymmetry caused by the ignorance of the underwriter. This weakness offers the underwriter an opportunity to deny his liabilities intentionally and maliciously.

(3) The remedy regulates an indifferent penalty rule for policyholders of every variety, albeit a policyholder could be classified in the four types. The loss of the policyholder cannot be covered, and, simultaneously, the extra premium is impossible to be refunded. As a result, a policyholder discloses all of IoM in $\Phi_u \cup \Phi_p$ so that either the insurance contract is priced highly, which induces more insurance costs for the policyholder, or the underwriter cannot provide the coverage, where the policyholder cannot share his risk in the financial market.

Amendment must be done to regulate a more relax remedy about the duty of disclosure. The remedies should be discriminative with respect to the variety of policyholders. Consider the following four scenarios: (1) no insurance contract formed in the normal market and the inducement belonging to $X^{(\cdot)}_{[n-k+1]},\ldots,X^{(\cdot)}_{[n]}$; (2) no insurance contract formed in the normal market and the inducement belonging to $X^{(\cdot)}_{[n+1]},\ldots,X^{(\cdot)}_{[n+m-k]}$; (3) highly priced insurance contract formed in the normal market and the inducement belonging to $X^{(\cdot)}_{[n-k+1]},\ldots,X^{(\cdot)}_{[n]}$; (4) highly priced insurance contract formed in the normal market and the inducement belonging to $X^{(\cdot)}_{[n+1]},\ldots,X^{(\cdot)}_{[n+m-k]}$.

Table 5.2 Utility of the policyholder under the amended duty of disclosure

		Insurance Contract without the Fully Disclosed IoM	Highly Priced Insurance Contract with the Fully Disclosed IoM
Ex ante expected utility surplus of policyholder		$U(\omega-Q_\tau)-E_pU(\omega-y)$	$U(\omega-Q_\tau)-U(\omega-\bar{Q}_{\bar{\tau}})$
Ex post utility	$X^{(\cdot)}_{[n-k+1]},\ldots,X^{(\cdot)}_{[n]}$	$U(\omega-Q_T-y_{T^*})$	$U(\omega-\bar{Q}_{T^*})$ *
	$X^{(\cdot)}_{[n+1]},\ldots,X^{(\cdot)}_{[n+m-k]}$	$U(\omega-\min\{Q_T,\bar{Q}_{T^*}\})$	$U(\omega-Q_{T^*})$ **

* $\bar{Q}_{T^*}=\int_T^{T^*}\bar{q}_t e^{-r_f(t-T)}dt$

** $Q_{T^*}=\int_T^{T^*}q_t e^{-r_f(t-T)}dt$

The general principle, differing from the current duty of disclosure in Marine Insurance Law 1906, is that the utilities of the policyholder after the ex post remedy in these four scenarios are rank in the order (4) > max{(2), (3)} > min{(2), (3)}> (1), and (3) > (1), (4) > (2) as well. A possible amendment suggestion is expressed in Table 5.2.

Scenario (1) The most strict remedy is adopted in this situation. The policyholder provides the biased IoM important for the underwriter intentionally, and purposely through concealment he obtains the formation of the insurance contract, which impossibly exists in the normal market. The underwriter has right to discharge the insurance contract, uncover the loss y_{T^*} and refuse to refund the premium covering time interval $\left[T^*,T+1\right]$.

Scenario (2) Albeit the policyholder fails to provide the related IoM about the loss inducement to the underwriter, the underwriter also has the liability for the underestimation of the loss probability due to the ignorance about the insured bids. The underwriter can discharge the insurance contract but has the obligation to cover the loss. On the other hand, the policyholder should pay the extra premium to fill the gap between Q_{T^*} and the theoretical premium amount $\bar{Q}_{T^*}$. Since the insurance contract cannot be formed in the normal market, the premium rate must be extremely high so that $\bar{Q}_{T^*}$ also might exceed the actual baseline of the policyholder. Thus, the premium compensation is constrained by an upper limit, the collected premium for the whole interval Q_T.

Scenario (3) Albeit the policyholder provides the biased IoM, which is important for the underwriter, the motivation of the policyholder is not as severe and unforgivable as the situation (1). The underwriter has right to discharge the contract. However, if the policyholder would like to pay the extra premium to fill the premium gap, the underwriter has obligation to cover the loss.

In the scenarios (2) and (3), the actual payment for the premium depends on the severity of each scenario and the theoretical premium. If there is still the extra premium collection left after compensation, the extra part should be refunded to the policyholder.

Scenario (4) The underwriter should take charge of the most of the liability in this dispute. The underwriter has right to discharge the insurance contract, but he has obligation to cover the loss and refund the extra premium.

The advised amendment gives a policyholder the more convenient opportunity to pool risk in the insurance market, especially for the policyholder of some particular IoM privately. New penalty rules are regulated with respect to the different scenarios. The underwriter will not suffer the premium loss in the new penalty rules, while he is also unable to earn a lot from rejecting coverage intentionally.

The severity of the information asymmetry is not same as the situation hundred years before. Through the variety of the investigation tools, an underwriter can observe and understand more and more IoM used to be the private one. However, the application of new technology in the shipping industry generates the new private IoM, which is more valuable for the underwriter to estimate the loss distribution. This is a learning process, in which the underwriter discovers the new IoM through covering the loss induced correspondingly. If the remedy is over strict, the insurance contract is difficult to be formed, which objectively blocks the learning process. On the other hand, the policyholder operates in the world full of risk, where any effort on the application of new technology requests the protection from the insurance market. The strict remedy enforces the policyholder to adopt a more traditional technology and delays the development of the shipping industry.

The suggested amendment indicates the whole learning process: (1) collect IoM from the policyholder to form the initial information set; (2) price the insurance contract to obtain the equilibrium premium rate; (3) identify the liabilities of each party, the policyholder and the underwriter, and remedy; (4) collect new IoM and update the initial information set to form a fresh set; (5) re-price the insurance contract to obtain the premium rate for the next period; (6) repeat (3) – (5). If it is found that the policyholder is uninsurable during the re-pricing of the insurance contract, the loop is broken down. The underwriter has to seek for the protection from the reinsurance market.

6. Empirical Evidence

Empirically, the information of materiality with respect to the marine insurance can be classified into two categories, the internal (natural) properties of a vessel and the external influence from the maritime adventure. In this research, the data of the marine loss and the detailed vessel information during 1999-2007 are collected to test the materiality of each piece of information. World Casualty Statistics by Lloyd's Register provides the marine loss data about the detailed information of the total-loss[4] vessels during 1999-2007, including vessel name, vessel type, registry, vessel size (in dwt),

[4] Total loss, defined by the Act 1906, composed of two situations, actual total loss and constructive total loss. Where the subject-matter insured is destroyed, or so damaged as to cease to be a thing of the kind insured, or where the assured is irretrievably deprived thereof, there is an actual total loss. Subject to any express provision in the policy, there is a constructive total loss where the subject-matter insured is reasonably abandoned on

vessel age, incident type, incident zone and incident time (season). Shipping Intelligence Network by Clarkson's Research Service offers the time series data of the sizes and the numbers of non-total-loss vessels.

Table 6.1 Summary of variables in regression

Variable type		Variable name	Variable structure
Independent Variables		Incidence possibility index (IPI)	$\log\left[\frac{\text{Number of the certain type of total-loss vessels in this year}}{\text{Total number of the certain type of vessels in this year}} \cdot \frac{\text{Size of the certain type of total-loss vessels in this year}}{\text{Total size of the certain type of vessels in this year}}\right]$
Dependent Variables	Internal property	Vessel type	Dummy (eight types of vessel)
		Registry	Dummy (FOC or closed registry) Ratio: $\frac{\text{the annual incident number with respect to certain registry}}{\text{the total incident number in this year}}$
		Vessel size	In dwt
		Vessel age	Dummy (four age intervals, 0-16, 17-25, 26-42, larger than 42)
	External influence	Incident type	Ratio: $\frac{\text{the annual incident number with respect to certain incident type}}{\text{the total incident number in this year}}$
		Incident zone	Ratio: $\frac{\text{the annual incident number with respect to certain zone}}{\text{the total incident number in this year}}$
		Incident time	Ratio: $\frac{\text{the annual incident number with respect to certain season}}{\text{the total incident number in this year}}$

In this research, we have totally collected the information of 882 total-loss vessels, distributed to eight types of vessels taken into the considerations, that is, bulk carrier, chemical, container, liquefied gas, MPP, reefer, Ro-Ro and tanker. These eight types of vessels have already occupied the main shares of the global tonnage and dominated the major part of shipping business, especially the international seaborne transportation.

The regression process and the results are reported in Table 6.2 attached in the appendix. After the eight steps of regressions, the insignificant variables are deleted from the regression models one by one based on the *p*-values of the *t* test. Finally, under the confidence level of 90%, the variables, incident zone, vessel size, vessel age interval 26-42, vessel types (bulk carrier, chemical, liquefied gas, MPP, reefer and Ro-Ro) and registry dummy (FOC) are left in the regression, while under the confidence level of 95%, the impact of registry dummy is no longer significant.

The regression steps and significance levels of dependent variables reflect the general sequence of the materiality with respect to each piece of information. Accordingly, the information of materiality can

---

account of its actual total loss appearing to be unavoidable, or because it could not be preserved from actual total loss without an expenditure which would exceed its value when the expenditure had been incurred.

be found, confirmed and ranked from the most important as follow. (1) Vessel type makes significant impact on the incident possibility, albeit it is not hold for every vessel type, for instance, container and tanker. The rest vessels, belonging to MPP, Ro-Ro, bulk carrier, chemical, reefer and liquefied gas, determine significantly the incident possibility correspondingly due to their particular functions on their own business lines. (2) Vessel size has positive impact on the incident possibility index, which implies that the huge-size vessel has great insurance value as well as the destroy capability. (3) Incident zone is the only external factor left in the last regression step. According to World Casualty Statistics 199-2007, about 74% of the total-loss incidents happened on the busiest ship routing connecting Europe and Eastern Asia, which starts from UK, passing through the Mediterranean, Red Sea, Indian Ocean, comes into Southern China Sea by Malacca, and finally ends at Eastern Asian countries, China, Japan and Korea. (4) The vessels in ages 26 – 42 is more dangerous than the ones of the same type in the other age intervals. (5) Open registry has positive impact on the incident possibility index. The vessel registered in a flag-of-convenience (FOC) country has probably greater motivation to loose his security management, which results in an increasing incident possibility.

Among the varieties of information of materiality examined, an underwriter can investigate the information purposely. Obviously, most of the information is difficult to conceal, for instance, vessel type, vessel size and registry. With the development of the specialized-function vessels, the internal properties of an insured vessel are no longer the private information of the policyholder. The strict remedy with respect to duty of disclosure in the Act 1906 fails to coincide with the current development of the shipping industry. The more flexible and loosen remedy rules should be legislated.

On the contrary, the information of materiality, which is most possibly provided with bias, should be closely connected to the business activities, for example, the routing lines (possible incident zones) and cargo states (cargo varieties).

7. Conclusion

In this study, we focus on the duty of disclosure in Marine Insurance Law 1906, which performs outdated and over strict in the remedy rules. Through analyzing the information structures of the underwriter and the policyholder, we separate the whole picture of the IoM into three categories. The policyholder provides the IoM with bias to make the insurance contract formed and obtain the low premium. By the stochastic control model, the optimal premium rates are calculated with respect to the truthfully disclosed IoM and the biased IoM, respectively. The insurance contract is accepted by the both parties.

When a dispute happens, there are four possible scenarios. The current remedy rule is so strict that it fails to distinguish the four scenarios from each other. Our amendment suggestion about the duty of disclosure provides a slightly relaxed remedy, which improves the learning process of the underwriter and encourages the policyholder to develop and apply new technology.

References

Boot, A. W. A., Thakor, A. V., 2001. The many faces of information disclosure. The Review of Financial Studies, 14(4), 1021-1057.

de Garidel-Thoron, T., 2005. Welfare-improving asymmetric information in dynamic insurance markets. Journal of Political Economy, 113(1), 121-149.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., 1998. Law and finance. Journal of Political Economy, 106(6), 1113-1155.

Salanié, B., 1997. The Economics of Contracts: A Primer. The MIT Press, Cambridge, Massachusetts, London, England.

Taksar, M. I., 2000. Optimal risk and dividend control models for an insurance company. Mathematical Methods of Operations Research.

Yan, J., Liu, J. J., Li, K. X., 2008, Threshold control of mutual insurance with limited commitment. Insurance: Mathematics and Economics, article in press and online available.

Appendices

The solution of optimization (1) can be obtained by solving the following HJB equation about the value function $V_1\left(x_u\right)$.

$$\frac{1}{2}\sigma_u^2 V_1''\left(x_u\right)+\left(q_t-\mu_u\right)V_1'\left(x_u\right)=0 \qquad \text{(A1)}$$

It is easy to solve (A1) and obtain $V_1\left(x_u\right)=\exp\left\{-\frac{2\left(q_t-\mu_u\right)}{\sigma_u^2}x_u\right\}$, since $V_1\left(0\right)=1$ and $V_1\left(\infty\right)=0$.

Similarly, optimization (2) can be solved by the HJB equation (A2)

$$\frac{1}{2}\sigma_p^2 V_2''\left(x_p\right)+\left(\overline{q}_t-\mu_p\right)V_2'\left(x_p\right)=0 \qquad \text{(A2)}$$

And we can obtain $V_2\left(x_p\right)=\exp\left\{-\frac{2\left(\overline{q}_t-\mu_p\right)}{\sigma_p^2}x_p\right\}$.

Table 6.2 Regression Result

Regression Steps	IPI							
	1	2	3	4	5	6	7	***8***
Incident Season	0.0615353 0.38 0.706	0.061535 0.38 0.706	0.061535 0.38 0.706	0.064396 0.4 0.691	0.065031 0.4 0.688			
Incident Type	-0.1375355 -1.63 0.104	-0.13754 -1.63 0.104	-0.13754 -1.63 0.104	-0.1386 -1.65 0.1	-0.13859 -1.65 0.1	-0.13846 -1.65 0.1	-0.14017 -1.67 0.096	
Incident Zone	-0.4534743 -2.39 0.017	-0.45347 -2.39 0.017	-0.45347 -2.39 0.017	-0.45187 -2.39 0.017	-0.45241 -2.39 0.017	-0.43441 -2.37 0.018	-0.4497 -2.48 0.013	***-0.47168*** ***-2.6*** ***0.009***
Registry (Ratio)	-0.1250929 -0.5 0.615	-0.12509 -0.5 0.615	-0.12509 -0.5 0.615	-0.12485 -0.5 0.616	-0.12436 -0.5 0.617	-0.13333 -0.54 0.59		
Vessel Size	2.52E-06 3.85 0	2.52E-06 3.85 0	2.52E-06 3.85 0	2.53E-06 3.88 0	2.53E-06 3.9 0	2.53E-06 3.89 0	2.52E-06 3.88 0	***2.57E-06*** ***3.96*** ***0***
Age (0 – 16)	-0.0556855 -0.89 0.375	-0.06623 -1.66 0.097	-0.06623 -1.66 0.097	-0.06484 -1.66 0.098	-0.06524 -1.67 0.095	-0.06417 -1.65 0.1	-0.06479 -1.66 0.096	
Age (17 – 25)	0.0105414 0.19 0.852							
Age (26 – 42)	-0.0600213 -1.09 0.278	-0.07056 -2.43 0.015	-0.07056 -2.43 0.015	-0.06894 -2.49 0.013	-0.06906 -2.5 0.013	-0.06872 -2.49 0.013	-0.06963 -2.53 0.012	***-0.05582*** ***-2.16*** ***0.031***
Age (≥ 43)	(dropped)	-0.01054 -0.19 0.852	-0.01054 -0.19 0.852					
Bulk Carrier	1.155068 9.22 0	1.155068 9.22 0	1.129602 17.88 0	1.130165 17.92 0	1.133677 18.76 0	1.133402 18.76 0	1.130625 18.79 0	***1.125641*** ***18.71*** ***0***
Chemical	1.454247 10.54 0	1.454247 10.54 0	1.428781 16.29 0	1.429418 16.31 0	1.433196 16.78 0	1.433463 16.79 0	1.433391 16.79 0	***1.415162*** ***16.62*** ***0***
Container	(dropped)	(dropped)	-0.02547 -0.2 0.842	-0.02507 -0.2 0.844				
Liquefied Gas	0.8164032 4.6 0	0.816403 4.6 0	0.790937 5.58 0	0.791531 5.59 0	0.795237 5.67 0	0.793351 5.66 0	0.795369 5.68 0	***0.800528*** ***5.71*** ***0***
MPP	3.321799 27.82 0	3.321799 27.82 0	3.296333 62.17 0	3.296487 62.21 0	3.3002 66.69 0	3.29914 66.8 0	3.298857 66.83 0	***3.290416*** ***67.41*** ***0***
Reefer	1.488227 10.49 0	1.488227 10.49 0	1.46276 15.72 0	1.462769 15.72 0	1.466472 16.11 0	1.466027 16.11 0	1.464542 16.11 0	***1.458702*** ***16.02*** ***0***
Ro-Ro	2.31124 18.45 0	2.31124 18.45 0	2.285773 35.02 0	2.285979 35.05 0	2.28971 36.71 0	2.290003 36.74 0	2.287548 36.81 0	***2.275607*** ***36.78*** ***0***
Tanker	0.0254662 0.2 0.842	0.025466 0.2 0.842						
Registry (FOC or not)	0.0524086 1.82 0.069	0.052409 1.82 0.069	0.052409 1.82 0.069	0.05269 1.83 0.067	0.052707 1.83 0.067	0.053741 1.88 0.061	0.04753 1.82 0.07	***0.049676*** ***1.9*** ***0.58***
Constant	-6.707169 -49.62 0	-6.69663 -52.34 0	-6.67116 -94.12 0	-6.67368 -95.97 0	-6.67744 -99.91 0	-6.66427 -114.47 0	-6.66559 -114.64 0	***-6.71222*** ***-123.3*** ***0***
Adjust- R^2	0.8915	0.8915	0.8915	0.8916	0.8917	0.8919	0.8919	***0.8915***

The effect of risk aversion on manufacturer promotion in a two-stage supply chain

Donghan Liang[1a], *Gang Li*[1b], *Jie Gao*[1c], *Linyan Sun*[1d] *and Xinyu Sun*[2]

[1]The Management School of Xi'an Jiaotong University, The State Key Lab for Manufacturing Systems Engineering, Xi'an, 710049, China.
[a] E-mail: dhsweet@yahoo.com.cn
[b] E-mail: glee@mail.xjtu.edu.cn
[c] E-mail: gaoj@mail.xjtu.edu.cn
[d] E-mail: lysun@mail.xjtu.edu.cn
[2]Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hong Kong.
E-mail: lgtxysun@inet.polyu.edu.hk

Abstract

We consider a supply chain system with a risk-neutral manufacturer as the leader and a risk-averse retailer as the follower in the environment with uncertain demand. At the beginning of the game, the manufacturer makes investment on promotion effort and then the retailer decides his ordering quantity before demand realization. The analysis of equilibrium strategies of this Stackelberg game indicates some characteristics which are different from promotion strategies with risk neutral agents. Firstly, there exists an upper bound for the retailer's target profit α, otherwise the equilibrium strategy is unavailable. Secondly, the retailer's risk aversion has direct influence on the manufacturer's promotion investment. In other words, the manufacturer will increase his promotional effort when the retailer has an appropriate degree of risk aversion and cuts down that for a highly risk-averse one. Thirdly, although conventional wisdom suggests that risk-averse retailer definitely reduce his ordering quantity, we find that manufacturer's promotion can effectively prevent the risk-averse retailer from downsizing inventory which is decided by the joint power of the promotion effort ρ and the variable pair(α , β).

Key word: Risk aversion; Promotion effort; Newsvendor

1. Introduction

The power of promotion has been identified in abundant literatures on marketing as well as supply chain management. Promotional activities are implemented by both suppliers and retailers, spreading from media advertising, sponsoring for business events to sending catalogues and salesmen's effort. Typically, promotion activities are simplified to advertising for model analysis and classified into brand advertising and local advertising according to the agent responsible for the matter. Generally speaking, brand advertising is always implemented by the brand owner, widely known as the supplier or manufacturer, to make his product differentiated. In other words, the manufacturer expects to grab potential demand and to develop brand knowledge and preference through brand advertising (Huang and Li, 2001). However, these days, as retailers are getting dominant in supply chain partnerships, famous retailers, such as Wal-mart, Go-me, Suning etc., also make brand advertising for their own good, which is interesting but beyond the range of this paper.

Most marketing studies to date on promotion or advertising have focused on how customers respond to retailers' sales effort, that is, marketing researches mainly concern retailer's local advertising or promotion and the relationship between retailers and consumers. On the other side, in the field of supply chain management, literatures on vertical co-op advertising constitutes mainstream of advertising studies, which typically consider the decision making process arises between

manufacturers and retailers and the equilibrium optimizing the whole supply chain performance. Basically, these papers discuss advertising in general terms, that is, the members involved are assumed identical in behavioral terms such as rational hypothesis, risk attitude etc.. Actually, the way advertising as well as other promotional activities works closely relates to these behavioral features that impact on the players' strategy choice.
Proved to be influential to decision making, risk attitude is a dominant feature of supply chain members and plays an important role in supply chain performance. Most recent studies on optimal stock level take the buyer's risk attitude into account, especially when the buyer is risk averse, to acquire meaningful insights to improve supply chain efficiency. Early risk aversion measures include utility maximization and mean-variance analysis, still in use and illustrative for a number of scenarios. Recently, the introduction of financial measures, the so-called VaR and CVaR, bridging supply chain members' risk aversion with his target profit directly, has greatly enriched the meaning of risk aversion and made analysis more applicable. Likewise, studies with VaR or CVaR approaches on supply chain management also mainly focus on the inventory issue, while the interaction between adverting or promotional activities and risk-averse members is lack of investigation.

This paper follows Gan's (2005) work on supply chain coordination with a risk-averse retailer, in which the concept of downside risk is introduced and employed to illustrate the retailer's inventory strategy and the supplier's coordination contract. The present paper follows most of Gan's (2005) assumptions and takes the view of promotion to investigate supply chain members' equilibrium strategies. Targeting at demonstrating the impact of brand advertising on the retailer's decision making, a Stackelberg game is employed where the supplier act as the leader and the retailer the follower. We find that the risk-neutral manufacturer increases promotional effort in a certain range determined by the retailer's risk-aversion degree and the risk-averse retailer does not necessarily order less than that of a risk-neutral one.

The paper is organized as follows: in section 2 we review literatures concerned; in section 3 we propose our assumptions, model a Stackelberg game with downside risk constraints and report our insights; in section 4 we make the conclusion and discussion part.

2. Literature Review

Conventional wisdom suggests advertising a powerful promotional tool widely investigated in both marketing and SCM criteria. One branch of marketing studies attaches more importance to the advertising attributes such as information, content, formality etc., (e.g. Bloch and Manceau, 1999; Dukes and Gal-Or, 2003); the other branch considers advertising as a principle marketing strategy (e.g. Narasimhan et al. 1996; Drèze and Bell 2002). It is the marketing perspective on advertising that accounts for the customer-oriented research schemes in which the relationship between supply chain members does not concern much, but the insights it provided are meaningful to brand promotion and market attracting for enterprise competition.

On the contrary, SCM papers with promotion scenarios are highly interested in its influence on supply chain members' relationship on cost and profit sharing, which is widely investigated in the criteria of retailer promotion and co-op vertical advertising recently. Cachon (2002) thoroughly reviewed supply chain coordination studies on newsvendor with demand dependent on the retailer's promotional effort and summarized needed conditions under which the supplier would share the retailer's promotion expense and the supply chain can be coordinated. He emphasized that the promotional cost should be observable to the supplier and verifiable to the court and directly benefits the supplier otherwise the cost sharing contract cannot be implemented. Constrained by the rule mentioned above, most studies on promotion either particularly demonstrate the definition of promotional effort in their model or directly choose advertising as promotional parameters for its observable and verifiable cost structure. Generally, the retailer's closeness to end consumers facilitates various promotional activities which have been widely investigated as a significant supply chain phenomenon. Wang and Gerchak (2001) assumed the demand for a certain product is influenced by its display level controlled by the retailer and indicated that the manufacturer should compensate the retailer with an extra holding cost to coordinate the channel and make a profit. Taylor (2002) proposed a supply chain coordination contract for which the retailer has to decide his promotional effort in addition to inventory level. Krishnan et al (2004) extended Taylor's (2002) research and investigated the coordination mechanism when the retailer chooses inventories ex ante and promotional effort ex post. In fact, what the above papers concerns is appropriate compensation mechanism for the retailer who bears the promotional cost beneficial to the supplier as well. The contracts designed for this purpose discreetly split the cost and profit to insure the compensation would reduce the retailer's risk without compromising his effort degree. Although risk attitude was excluded in these studies, they performed as risk-sharing tools in common. Suo et al (2005) explicitly presented a model considering the impacts of the retailer's loss-aversion on his promotional effort. Using a loss-aversion utility function, they found that loss aversion weakens incentives for retailer's sales effort.

Literatures on retailer's promotion mainly establish analysis on the basis of newsvendor problem, while studies on vertical co-op advertising usually develop models with deterministic demand denoted by a function of retailing price and channel members' advertising investment. Jorgensen and Zaccour (1999) proposed a differential game to study channel coordination and channel conflict with channel members' advertising and pricing strategies, proved the existing of closed-form solutions for both scenarios and obtained a global result that the cooperative scenario supports greater level of adverting investments from both members. Thereafter studies on co-op advertising mainly concerns advertisement efforts in dimensions such as national level expenditures, local level expenditures, manufacturer participation rate, sales volume, and brand and store substitutions (Xie and Wei, 2009). Huang and Li (2001) developed three co-op advertising models to explain the cost-sharing issue between the manufacturer and the retailer. For the cooperative model they originally employed Nash bargaining game and took supply chain members' risk attitude into account. Utilizing the Pratt-Arrow risk aversion function they found that the manufacturer shares less of the local advertising cost if the retailer has a higher degree of risk aversion.

Literatures reviewed above typically consider supplier advertising as a supplement to retailer's promotion even with retailer's risk aversion involved. This paper develops a model with a risk-averse retailer and upstream promotion and illustrates the significant role played by supplier promotion in the sense of risk sharing.

3. Model

We now consider a Stackelberg game that consists of a risk-neutral manufacturer *M* and a risk-averse retailer *R*, in which the manufacturer *M* performs as the leader and the retailer plays as the follower. Based on the newsvendor, we suppose the transaction contains a single perishable product with a random market demand *X* (i.e. the deterministic quantity of *X* can not be observed before the selling season) that has a distribution density *f(x)* and distribution function *F(x)* known as common knowledge to both the manufacture and the retailer. The two players moves in following sequence: first the manufacturer promotes his product with effort ρ to enlarging the market demand at an expense $V(\rho)$, increasing on ρ with $V'(\rho) \geq 0, V''(\rho) \geq 0$, extending the original demand X to ρX which realizes when the selling season begins; then the manufacture wholesales products to the retailer at unit cost *c* and receive *w* each unit, and the retailer will sell them on the market at price p per unit; finally the enlarged market demand ρX is observed. For the simplicity of our analysis, we assume the goodwill cost and salvage value of the perishable product is zero for both the manufacturer and the retailer. We also assume that each player targets at optimizing his expected profit within the constraint and there's no information asymmetry.

There are two critical decision variables in the system above-the manufacturer's promotion effort ρ and the retailer's ordering quantity *q*-taking up our main concern in following analysis. The above introduction of our model mostly inherits the traditional newsvendor with promotion problem and the retailer's downside risk constraint is presented in this part. The concept of downside risk was employed by Gan (2005), implying a bearable biggest failure rate describing the probability that the agent can not achieve his target profit, thereby the retailer would reduce his ordering quantity as long as the downside risk exceed his limitation. The constraint is also constructed accordingly: suppose the retailer has a target profit α and downside risk β and his risk constraint is written as:

$$P(\prod_r \le \alpha) \le \beta \tag{1}$$

In which $\prod_r$ represents the retailer's profit and $\prod_r = p\min(q, \rho X) - wq$.
Expected profit functions for the manufacturer, the retailer and the system are determined by following equations:

$$\pi_m = (w-c)q - V(\rho) \tag{2}$$

$$\pi_r = pE\min(q, \rho X) - wq$$
$$s.t.\ \ P(\prod_r \le \alpha) \le \beta \tag{3}$$

$$\pi_s = \pi_m + \pi_r = pE\min(q, \rho X) - cq - V(\rho) \tag{4}$$

Then we solve for the non-cooperative sequential game with the manufacturer as the leader and the retailer as the follower and the result is Stackelberg equilibrium.

We begin with the retailer's ordering strategy considering his downside risk constraint. Let q^* be the optimal ordering quantity of the retailer whose maximization problem is descried in programming (3) and the objective function is $\max \pi_r$. We consider the risk constraint best of all and split it into two scenarios: $q \le \rho X$ and $q > \rho X$ (in which the variable ρ is treated as a known constant because it has been decided by the manufacturer at the first stage of the game). For the first scenario, all the products are sold out and constraint (1) is equal to the expression below:

$$P((p-w)q \le \alpha) \le \beta \tag{5}$$

So we get the lower bound of the retailer's optimal order quantity $q^0 = \dfrac{\alpha}{p-w}$. The retailer make a profit no more than *(p-w)q* given his ordering quantity *q*. Therefore, if the order quantity is less than q^0, the target profit α can never be achieved and the retailer has to order more than q^0 to meet his target profit. When $q^0 < q \le \rho X$, the retailer would deterministically gain more than α and the downside risk is zero, therefore, the constraint binds only if $q > q^0$ and $q > \rho X$. For the second scenario, we have

$$P(p\rho X - wq \le \alpha) = P(X \le \frac{\alpha + wq}{p\rho}) = F(\frac{\alpha + wq}{p\rho}) \le \beta \tag{6}$$

Expression (6) relates the demand distribution function *F(x)* to the downside risk β, which is critical to our further analysis. With some manipulation on expression (6) we get an upper bound for *q*: $q \le \dfrac{p\rho F^{-1}(\beta) - \alpha}{w}$ when the constraint binds. Let ρ^* be the optimal promotion effort invested by the manufacturer in the first stage and ($\overline{\rho}$, $\overline{q}$) be the equilibrium strategy for traditional newsvendor（all the players are risk neutral). Let (ρ_h^*, q_h^*) be the equilibrium strategy when $\beta \ge F(\dfrac{\alpha + w\overline{q}}{p\overline{\rho}})$ and (ρ_l^*, q_l^*) when $F(q^0) < \beta < F(\dfrac{\alpha + w\overline{q}}{p\overline{\rho}})$. Theorem1 describes the equilibrium strategy (ρ^*, q^*) with dynamic α and β value.

Theorem 1:

① *If* $0<\alpha\le\frac{\bar\rho}{\bar\rho-(1-\frac{w}{p})}\cdot\frac{w}{p}(p-w)F^{-1}(\frac{p-w}{p})$, *then*

(1) *when* $\beta\ge F(\frac{\alpha+w\bar q}{p\bar\rho})$, $q_h^*=\bar q=\bar\rho F^{-1}(\frac{p-w}{p})$, $\rho_h^*=\bar\rho$, $V'(\rho)|_{\rho=\bar\rho}=(w-c)F^{-1}(\frac{p-w}{p})$;

(2) *when* $F(q^0)<\beta<F(\frac{\alpha+w\bar q}{p\bar\rho})$, $q_l^*=\frac{p\rho_l^*F^{-1}(\beta)-\alpha}{w}$, $V'(\rho)|_{\rho=\rho_l^*}=\frac{(w-c)p}{w}F^{-1}(\beta)$

(3) *when* $\beta\le F(q^0)$, *there is no equilibrium solution.*

② *If* $\frac{\bar\rho}{\bar\rho-(1-\frac{w}{p})}\cdot\frac{w}{p}(p-w)F^{-1}(\frac{p-w}{p})<\alpha\le\bar\rho(p-w)F^{-1}(\frac{p-w}{p})$, *then the equilibrium strategy is*($\bar\rho$, $\bar q$) *with any* β *that satisfies* $F(q^0)<\beta<1$.

③ *If* $\alpha>\bar\rho(p-w)F^{-1}(\frac{p-w}{p})$, *there is no available solution.*

Proof:

① If $0<\alpha\le\frac{\bar\rho}{\bar\rho-(1-\frac{w}{p})}\cdot\frac{w}{p}(p-w)F^{-1}(\frac{p-w}{p})$, then we have $F(q^0)\le F(\frac{\alpha+w\bar q}{p\bar\rho})$.

(1) When $\beta\ge F(\frac{\alpha+w\bar q}{p\bar\rho})$, the retailer's downside risk constraint does not bind, therefore the retailer's order decision is the same as that of the traditional newsvendor, which is given by $\bar q=\rho F^{-1}(\frac{p-w}{p})$. Then we substitute $\bar q$ into Eq. (2) and solve $\max\limits_{\rho\ge1}\pi_m$ for the manufacturer's optimal promotional effort $\bar\rho$, which can be simply obtained through first derivative condition. The equilibrium strategy($\bar\rho$, $\bar q$) takes the form

$$V'(\rho)|_{\rho=\bar\rho}=(w-c)F^{-1}(\frac{p-w}{p}),\quad \bar q=\bar\rho F^{-1}(\frac{p-w}{p}).$$

(2) when $F(q^0)<\beta<F(\frac{\alpha+w\bar q}{p\bar\rho})$, the downside risk constraint binds and the retailer's maximization problem becomes

$$\max_{q\ge0}\pi_r \qquad s.t.\ P(\Pi_r\le\alpha)=\beta \tag{*}$$

We can now derive the solution to programming (*), as shown below:

$$q^*=\frac{p\rho F^{-1}(\beta)-\alpha}{w} \tag{7}$$

On substituting this equation into Eq. (2) and solving for ρ^*, we obtain(ρ^* , q^*) as follows:

$$V'(\rho)|_{\rho=\rho^*}=\frac{(w-c)p}{w}F^{-1}(\beta),\quad q^*=\frac{p\rho^*F^{-1}(\beta)-\alpha}{w}.$$

(3) When $\beta\le F(q^0)$, there is no such q that achieve the target profit α, making the whole problem unsolvable.

② If $\frac{\bar{\rho}}{\bar{\rho}-(1-\frac{w}{p})}\cdot\frac{w}{p}(p-w)F^{-1}(\frac{p-w}{p})<\alpha\le\bar{\rho}(p-w)F^{-1}(\frac{p-w}{p})$,

then $F(\frac{\alpha+w\bar{q}}{p\bar{\rho}})<F(q^0)<F(\bar{q})$. When $\beta\le F(q^0)$, it is obvious that no appropriate q matches; when $F(q^0)<\beta\le 1$, it can be deduced that we have $\beta>F(\frac{\alpha+w\bar{q}}{p\bar{\rho}})$, so the downside risk constraint does not bind, and we obtain the equilibrium strategy($\bar{\rho}$, $\bar{q}$).

③ If $\alpha>\bar{\rho}(p-w)F^{-1}(\frac{p-w}{p})$, then $F(\frac{\alpha+w\bar{q}}{p\bar{\rho}})<F(\bar{q})<F(q^0)$.When $F(q^0)<\beta\le 1$, $\beta>F(\frac{\alpha+w\bar{q}}{p\bar{\rho}})$, so the optimal order quantity is $\bar{q}$, which contradicts the fact $\bar{q}<q^0$, leaving our problem unsolvable.

Note that the retailer's target profit can not exceed his revenue in risk-neutral setting otherwise there is no appropriate ordering quantity that matches the downside risk constraint. Theorem 1 also shows that when the target profit below the risk-neutral revenue, higher target profit setting (α) requires bigger downside risk (β) for available solutions. When the retailer is highly risk-averse with high target profit and low downside risk, it is almost impossible to give a deterministic estimation for equilibrium solution because of the lower bound constraint $\beta>F(q^0)$. It is interesting that a risk-averse retailer's ordering quantity is not necessarily lower than that of a risk-neutral one due to the impacts of the supplier's promotion effort, a sharp contrast to the situation without supplier's promotion effort in which the risk-averse retailer order strictly less than the risk-neutral one presented in Gan's (2005) work. We start with Theorem 2 on the investigation of the relationship between the two elements ρ^* and q^* of the Stackelberg equilibrium strategy and the comparison of them with different concerning parameters.

Theorem2: *Considering* $0<\alpha\le\frac{\bar{\rho}}{\bar{\rho}-(1-\frac{w}{p})}\cdot\frac{w}{p}(p-w)F^{-1}(\frac{p-w}{p})$ *and* $\beta>F(q^0)$ *only*:

① *If* $\frac{w}{p}(p-w)F^{-1}(\frac{p-w}{p})<\alpha\le\frac{\bar{\rho}}{\bar{\rho}-(1-\frac{w}{p})}\cdot\frac{w}{p}(p-w)F^{-1}(\frac{p-w}{p})$, *then* $\rho^*_{l_1}\ge\rho^*_h$;

② *If* $0<\alpha\le\frac{w}{p}(p-w)F^{-1}(\frac{p-w}{p})$, *then two scenarios are considered*:

(1) *when* $F(\frac{w}{p}F^{-1}(\frac{p-w}{p}))<\beta\le F(\frac{\alpha+w\bar{q}}{p\bar{\rho}})$, *we have* $\rho^*_{l_2}\ge\rho^*_h$;

(2) *when* $F(q^0)<\beta\le F(\frac{w}{p}F^{-1}(\frac{p-w}{p}))$, *we have* $\rho^*_{l_3}\le\rho^*_h$.

Proof:

① If $\frac{w}{p}(p-w)F^{-1}(\frac{p-w}{p})<\alpha\le\frac{\bar{\rho}}{\bar{\rho}-(1-\frac{w}{p})}\cdot\frac{w}{p}(p-w)F^{-1}(\frac{p-w}{p})$, obviously we have $\frac{w}{p}F^{-1}(\frac{p-w}{p})<\frac{\alpha}{p-w}\le\frac{\bar{\rho}}{\bar{\rho}-(1-\frac{w}{p})}\cdot\frac{w}{p}F^{-1}(\frac{p-w}{p})$, which is in equivalence with the following expression:

$$F(\frac{w}{p}F^{-1}(\frac{p-w}{p}))<F(q^0)\le F(\frac{\alpha+w\bar{q}}{p\bar{\rho}}) \tag{8}$$

With $\beta > F(q^0)$ we can deduce that $\beta > F(\frac{w}{p}F^{-1}(\frac{p-w}{p}))$ and $F^{-1}(\beta) > \frac{w}{p}F^{-1}(\frac{p-w}{p})$.

Given that $V'(\rho)|_{\rho=\bar{\rho}} = (w-c)F^{-1}(\frac{p-w}{p})$ and $V'(\rho)|_{\rho=\rho_l^*} = \frac{(w-c)p}{w}F^{-1}(\beta)$, we compare ρ_l^* to $\bar{\rho}$ and find that $\rho_{l_1}^* \geq \rho_h^*$.

② If $0 < \alpha \leq \frac{w}{p}(p-w)F^{-1}(\frac{p-w}{p})$, then $F(q^0) \leq F(\frac{w}{p}F^{-1}(\frac{p-w}{p})) < F(\frac{\alpha + w\bar{q}}{p\bar{\rho}})$.

Therefore two ranges of β are considered with the downside risk constraint binding:

when $F(\frac{w}{p}F^{-1}(\frac{p-w}{p})) < \beta \leq F(\frac{\alpha + w\bar{q}}{p\bar{\rho}})$, we have $V'(\rho_l^*) > V'(\rho_h^*)$ and $\rho_{l_2}^* \geq \rho_h^*$;

when $F(q^0) < \beta \leq F(\frac{w}{p}F^{-1}(\frac{p-w}{p}))$, we have $V'(\rho_l^*) \leq V'(\rho_h^*)$ and $\rho_{l_3}^* \leq \rho_h^*$。

Theorem 2 demonstrates the ρ value under different pairs of α and β, where the manufacturer's willing-to-pay investment in promotion or advertising alters as the retailer's risk attitude changes. It is reasonable that when the retailer becomes increasingly risk-averse-seeking for higher target profit and lower downside risk-the manufacture improves on his promotion effort as a signal of enlarging demand to boost the retailer's ordering quantity. However, the manufacturer would not keep on increasing promotional expense forever: if the retailer is conservative on his downside risk and expects little revenue, mass investment in promotion is improper for the manufacturer and he would cut down advertising cost to avoid potential loss. Moreover, we find that no evidence suggests that the retailer with highest downside risk and lowest target profit ensures the manufacturer's best promotional investment, which implies that the risk-neutral manufacturer prefer moderate risk aversion rather than boldness when dealing with the retailer. The analysis above reveals that the retailer's risk aversion typically influences the manufacturer's promotion decision, however, the manufacturer's promotion ultimately targets at extending consumer demand so that the retailer does not hesitate to increase inventory. We can expect that the manufacturer's promotion decision has an impact on the retailer's ordering decision in reverse. Therefore, the retailer's ordering quantity is influenced by both his own risk attitude and the manufacture's promotion effort, making his inventory decision complicated. Theorem 3 investigates the retailer's inventory decision considering promotion effect.

Theorem 3:

① *If* $\frac{w}{p}(p-w)F^{-1}(\frac{p-w}{p}) < \alpha \leq \frac{\bar{\rho}}{\bar{\rho}-(1-\frac{w}{p})} \cdot \frac{w}{p}(p-w)F^{-1}(\frac{p-w}{p})$ *and* $\rho_{l_1}^* \geq \frac{\bar{\rho}^2}{\bar{\rho}-(1-\frac{w}{p})}$,

then $q_{l_1}^* \geq q_h^*$;

② *If* $0 < \alpha \leq \frac{w}{p}(p-w)F^{-1}(\frac{p-w}{p})$ *and* $\rho_{l_2}^* \geq \frac{\bar{\rho}^2}{\bar{\rho}-(1-\frac{w}{p})}$, *when the range of* β *satisfies*

$$F(\frac{w}{p}F^{-1}(\frac{p-w}{p})) < \beta \leq F(\frac{\alpha + w\bar{q}}{p\bar{\rho}}), \text{ then } q_{l_2}^* \geq q_h^*;$$

③ *If* $0 < \alpha \leq \frac{w}{p}(p-w)F^{-1}(\frac{p-w}{p})$ *and* $F(q^0) < \beta \leq F(\frac{w}{p}F^{-1}(\frac{p-w}{p}))$, $q_{l_3}^* \leq q_h^*$.

Proof:

① If $0<\alpha\le\frac{\bar{\rho}}{\bar{\rho}-(1-\frac{w}{p})}\cdot\frac{w}{p}(p-w)F^{-1}(\frac{p-w}{p})$, then we can deduce that

$$q_l^*-q_h^*\ge\frac{p}{w}\rho_l^*F^{-1}(\beta)-\frac{\bar{\rho}^2}{\bar{\rho}-(1-\frac{w}{p})}F^{-1}(\frac{p-w}{p});$$

Suggested in Theorem 2, there are two scenarios in which $F^{-1}(\beta)\ge\frac{w}{p}F^{-1}(\frac{p-w}{p})$:

when $\frac{w}{p}(p-w)F^{-1}(\frac{p-w}{p})<\alpha\le\frac{\bar{\rho}}{\bar{\rho}-(1-\frac{w}{p})}\cdot\frac{w}{p}(p-w)F^{-1}(\frac{p-w}{p})$ or

when $0<\alpha\le\frac{w}{p}(p-w)F^{-1}(\frac{p-w}{p})$ with $F(\frac{w}{p}F^{-1}(\frac{p-w}{p}))<\beta\le F(\frac{\alpha+w\bar{q}}{p\bar{\rho}})$.

Therefore, for $\rho_l^*\ge\frac{\bar{\rho}^2}{\bar{\rho}-(1-\frac{w}{p})}$, we have $\rho_l^*F^{-1}(\beta)\ge\frac{w}{p}\frac{\bar{\rho}^2}{\bar{\rho}-(1-\frac{w}{p})}F^{-1}(\frac{p-w}{p})$, and the expression $q_l^*-q_h^*\ge 0$ is proved, leading to $q_l^*\ge q_h^*$.

② If $0<\alpha\le\frac{w}{p}(p-w)F^{-1}(\frac{p-w}{p})$ and $F(q^0)<\beta\le F(\frac{w}{p}F^{-1}(\frac{p-w}{p}))$, we directly have

$F^{-1}(\beta)\le\frac{w}{p}F^{-1}(\frac{p-w}{p})$, so that $\rho_l^*\le\rho_h^*$ and $q_l^*-q_h^*\le\frac{p}{w}\rho_l^*F^{-1}(\beta)-\bar{\rho}F^{-1}(\frac{p-w}{p})\le 0$,

therefore $q_l^*\le q_h^*$ is proved.

It is clearly illustrated in Theorem 3 that manufacturer's promotion changes the conventional phenomenon that a risk-averse retailer always ordering less than a risk-neutral one. Certainly the manufacturer has to pay more on promotion as a motivation towards the retailer to boost sales. As a result, the risk-averse retailer is encouraged to buy more from the manufacturer. Note that even if the retailer orders more than a risk-neutral retailer, the manufacturer may not gain extra revenue considering his expense on promotion. This suggests that the manufacturer has to balance his promotion cost and wholesale revenue facing a risk-averse retailer.

According to Gan (2005), the retailer's risk-aversion can be classified with variable pairs$(\alpha，\beta)$, the combination of a higher α and a lower β means increasing risk aversion. However, our analysis indicates that a pair of higher α and lower β probably yields unmatched target profit and downside risk (see Theorem 1). Therefore we cannot advise on the situation that highly risk-averse retailer is involved, and we only concerns appropriate pair of$(\alpha，\beta)$.

4. Conclusion

As discussed in Section 1, this paper is motivated by the desire to explain a common market phenomenon regularly neglected by researchers on marketing or SCM. To extend the scale of our work, we employ methods in different fields including marketing, SCM and finance. This is also similar to the environment where real firms are operated, with various elements impacting on each other. As for the model construction, we follow the form of downside risk analysis with Stackelberge game, that is, a combination of financial measure with game theory which seems to be sustainable in the sense of approach.

Our main concerns are the relationship between supply chain members with different risk attitude when promotion is considered. To address this issue, we compare the equilibrium strategies under different scenarios; although no firm judgment is made, we do obtain some insights about the interaction of the retailer and the manufacturer. Firstly, there exists an upper bound for the retailer's target profit α, otherwise the equilibrium strategy is unavailable. Secondly, the retailer's risk aversion has direct influence on the manufacturer's promotion investment. In other words, the manufacturer will increase his promotional effort when the retailer has an appropriate degree of risk aversion and cuts down that for a highly risk-averse one. Thirdly, although conventional wisdom suggests that risk-averse retailer definitely reduce his ordering quantity, we find that manufacturer's promotion can effectively prevent the risk-averse retailer from downsizing inventory which is decided by the joint power of the promotion effort ρ and the variable pair(α , β).

This work is an exploring job on researches about promotion and risk and there are still many unsolved problems, e.g. designing contracts to coordinate the supply chain and reallocate market risk, the introduction of competition mechanisms, etc.. Our research will go deep into this topic step by step to obtain more insights for both study and real world application.

References

Bloch, F. and Manceau, D., (1999), Persuasive advertising in Hotelling's model of product differentiation, International Journal of Industrial Organization, 17(4):557-574.

Cachon, G. (2002), Supply chain coordination with contracts. http://opim.wharton.upenn.edu/~cachon.

Drèze, X. and Bell, D.R., (2002), Changing the Channel: A Better Way to do Trade Promotions, Sloan Management Review, 43 (2): 42-49.

Dukes, A. and Gal-Or, E., (2003), Minimum Differentiation in Commercial Media Markets, Journal of Economics & Management Strategy, 12(3): 291-325.

Gan, X H., Sethi, s. and Yan, H.M.. (2005), Channel Coordination with a Risk-Neutral Supplier and a Downside-Risk-Averse Retailer, Production and Operation Management 14(1): 80-89.

Huang, Z.M. and Li, S.X., (2001), Co-op advertising models in manufacturer–retailer supply chains: A game theory approach, European Journal of Operational Research, 135(3): 527-544.

Jorgensen, S. and Zaccour, G., (1999), Equilibrium Pricing and Advertising Strategies in A Marketing Channel, Journal of Optimization Theory and Applications, 102 (1): 111–125.

Krishnan, H., Kapuscinski, R. and Butz, D. A. (2004), Coordinating Contracts for Decentralized Supply Chains with Retailer Promotional Effort, Management Science, 50(1): 48–63.

Narasimhan, C., Neslin, S. A. and Sen, S.K., (1996), Promotional Elasticity and Category Characteristics, Journal of Marketing, (60):17-30.

Suo, H.S., Wang, J.C. and Jin, Y.H., (2005), Supply Chain Coordination with Sales Effort Effects and Impact of Loss Aversion on Effort Decision, Tsinghua Science and Technology, 10(1): 102-107

Taylor, T. A. (2002), Supply chain coordination under channel rebates with sales effort effects, Management Science, 48(8): 992-1007.

Wang, Y. Z. and Gerchak, Y., (2001), Supply Chain Coordination When Demand is Shelf-space Dependent, Manufacturing & Service Operations Management, 3(1): 82-87.

Xie, J.X. and Wei, J. C., (2009), Coordinating advertising and pricing in a manufacturer–retailer channel, European Journal of Operational Research, 197(2): 785-791.

Safety culture and safety behaviors in container shipping: a seafarer's perspective

Chin-Shan Lu[1] *and Chaur-Luh Tsai*[2]

[1]Department of Transportation and Communication Management Science, National Cheng Kung University,
Tainan City 701, Taiwan, ROC.
Tel: +886-6-2757575 ext. 53243
Fax: +886-6-2753882
E-mail: lucs@mail.ncku.edu.tw
[2]Department of Transportation and Communication Management Science, National Cheng Kung University,
Tainan City 701, Taiwan, ROC.
Tel: +886-6-2757575 ext. 53271 ext.4050
Fax: +886-6-2753882

Abstract

This study examines safety culture and its effects on safety behaviors from a seafarer's perception in container shipping context. We developed a theoretical model to examine the effect of safety culture dimensions on safety behavior. We also formulated research hypotheses and tested them using survey data collected from 608 seafarers who work on 124 vessels from 13 of the top 20 global container carriers. Three safety culture dimensions are identified based on a factor analysis: (a) supervisor safety, (b) safety policy, and (c) safety management practice. The results suggest that the positive associations between safety culture and safety behavior in the container shipping context. We discuss our findings' contribution to the development of safety culture theory and their managerial implications for work safety in shipping operations.

Key words: Safety culture, safety behavior, container shipping, seafarers.

1. Introduction

Shipping is one of the most dangerous industries in the world. The seafarers are taking higher-level risk than other shore-based workers (Bloor et al., 2000; Roberts and Marlow, 2005). The shipping companies make efforts in an efficiency system such as ship design, technology, and safety equipment to improve safety at sea (IMO, 2007). However, the UK P & I (2004, 2005) reported that accident rate in shipping industry still stands on a high level, and the statistics revealed that average 137 ships claimed total loss and 700 lives lost in accidents per year during 2001 to 2007 (Maritime Knowledge Centre, 2008). Several statistical reports indicated that most maritime accidents were attributed to human errors (MOTC, 2004, 2005; UK P&I Club, 2004, 2005). What contributes to the occurrence of accidents is not always clear. A crucial assumption has been that workers' perceptions of safety culture might affect their behavior so as to decrease the propensity to have accidents (Brown et al., 2000; Fernández-Muniz et al., 2007; Huang et al., 2006; Johnson, 2007; Oliver et al., 2002; Rundmo et al., 1998; Seo, 2005; Siu et al., 2004; Tomás et al., 1999). Fleming (2005) and Nieva and Sorra (2003) pointed out that emphasis on safety culture could improve and influence work safety.

While the antecedents of safety culture have been widely discussed in the safety-related literature, relatively little research except Lu and Tsai (2008) have been conducted on the relationship between safety climate in shipping operations and accidents, an understanding of the perceptions of safety culture and its relationship with safety behavior in a container shipping context is lacking. Filling this research gap, the objective of this study is to investigate the influence of critical safety dimensions on work safety in container shipping operations from the seafarer's perception.

This paper is organized as follows: We introduce the study and discuss the background in the first section. We develop the research model and formulate the hypotheses in the second section. In the third section we discuss development of the research methodology, including the measurement constructs used in the survey, the sampling technique, and the research procedures. In sections four and five we present the empirical results from exploratory factor analysis, confirmatory factor analysis, and hierarchical regression analysis. The research findings and managerial implications were discussed in the final section.

2. Theory and Research Hypotheses

Safety culture is a specific form of organizational culture, which exposes how values, attitudes and beliefs about safety and how they might influence the directions of an organization. Safety culture could create an atmosphere in which employees are aware of the risks in their workplace, are continually on guard against them (Ostrom et al., 1993). The definition of safety culture by the Advisory Committee on the Safety of Nuclear Installations (ACSNI, 1993), which is the product of individual and group values, attitudes, perceptions, competencies, and patterns of behavior that determine the commitment to, and the style and proficiency of, an organization's health and safety management (HSC, 1993: 23). This definition implicit the safety culture relates to personal attitudes and habits of thought and to the style of organizations, which matters are generally intangible. Reason (1998) and Sorensen (2002) indicated that the human performance, management and organization are closely related to the safety culture. The failures arise from the safety culture of an organization have become as the reason why major accidents happened. Hidden (1989) and Clarke (1998) pointed out that a 'positive safety culture' was the key to improving the safety.

The safety culture is a system of values shared by all the members of an organization, by an underlying philosophy that serves and reflects its core mission and its reason to be (Glendon and Stanton, 2000; HSC, 1993; Mitroussi, 2003; Wert, 1986), and will impact on attitudes and behavior related to increasing or decreasing risks (Guldenmund, 2000). The notion of safety culture is the 'engine' that drives the system towards the goal of sustaining the maximum resistance towards its operational hazards, regardless of the leadership's personality or current commercial concerns (Reason, 1998).

The safety culture can be construed to be manifested in shared values and meanings, and in a particular organizational structure and process, safety policies, strategies, goals, practices and leadership styles related to safety management system (Glendon and Stanton, 2000, Guldenmund, 2000, Hale, 2000). There are three general indicators of the organization's safety culture including the managers' commitment to safety, the policies and procedures, and employees' involvement that influence the organizational safety and form the safety management system (Dufort and Infante-Rivard, 1998; Fernández-Muñiz et al., 2007). Those aspects of commitment, policies, and organizational safety management practices direct the approaches of risk prevention. These components are come from underlying patterns of shared meanings and beliefs and could act as valid indicators in a culture evaluation.

The behaviors of managers and supervisors reflect the actions signal of their commitment to safety and the prioritization of safety against other organizational goals (Flin and Yule, 2004). Once top management had provided the initial support for the development of a more advanced safety culture, a number of supporting tools were developed, and a strategy form implementation was developed.

Traditionally, safety has been measured by 'after the loss' type of measurements such as by recordable accidents rates (Cuny and Leijeune, 2003; Hoonakker et al., 2005; Hurst et al., 1996; Rose, 1992), occupational injury rate (Cuny and Leijeune, 2003; Hoonakker et al., 2005; Hurst et al., 1996; Siu et al., 2004), lost time injuries, near miss incidents rate, and days away from work (Hurst et al., 1996). Although the accident rate is the most popular indicator of safety outcome that has been used in the shipping industry, yet it is lack of the reliable statistics. There were no proper records or statistics included all levels of accidents and incidents. Moreover, the formal accident records in shipping industries were made only after a serious accident or disaster occurred, and recorded in different organizations (e.g., P&I Clubs, Port authority, IMO). Objective accident data are not very reliable since near misses are difficult to collect, staffs are reticent about reporting accidents, and an accident free period can suppress accident reporting (Håvold, 2005). In the other words, the formal information could not provide all comprehensive records related to safety in the shipping context. Thus, the lack of an adequate measurement is restricted evaluating the effectiveness of different safety programs (Glendon and Litherland, 2001). Rose (1992) reported that accident rates were not useful in predicting future accidents. Previous researches have produced a consensus about the outcome or consequences of safety culture (Fernández-Muniz et al., 2007; Huang et al., 2006; Johnson, 2007; Oliver et al., 2002; Seo, 2005; Siu et al., 2004). Flin (2007) suggested that the artifact level culture concept such as self-reports on safety, workers' safety behaviors, and comply with the standardizing reporting system can offer an additional safety performance measure.

The Advisory Committee on the Safety of Nuclear Installations (ACSNI, 1993) pointed out that safety culture represented the individual values, attitudes, perceptions, competencies, and patterns of behavior. The safety culture is considered as a system of values that will impact on attitudes and behavior related to increasing or decreasing risks (Guldenmund, 2000). Moreover, the accidents were resulted from workers' expectations, behaviors and errors (Flin, 2007). Zohar (2003) indicated that worker's safety perceptions and the resulting expectations that in turn will motivate and condition safe and unsafe behaviors. Prior studies have reported positive relationships between safety culture and safety behaviors (DeJoy et al., 1995; Neal et al., 2000). Brown et al. (2000) reported that safety culture could influence safety efficacy and cavalier attitudes and lead to safe and unsafe work behaviors.

Accordingly, this study aims to investigate crucial safety culture dimensions, and to examine the effects of safety culture on safety behavior according to seafarers' perceptions. Based on the safety culture dimension literature, dimensions such as supervisor safety, safety policy, and safety management practice are assessed in this research. Thus, this study hypothesized that:
Hypothesis 1: Supervisor safety will be positively related to the seafarers' safety behavior in container shipping.
Hypothesis 2: Safety policy will be positively related to the seafarers' safety behavior in container shipping.
Hypothesis 3: Safety management practice will be positively related to the seafarers' safety behavior in container shipping.

3. Methodology

3.1. Sampling Techniques

The data used in this study were obtained from survey respondents working on 124 vessels from 13 of the top 20 global container carriers in the world. These container carriers included APM Maersk, APL/NOL, CMACGM, Happag Lloyd, MSC, Yang Ming, Evergreen group, Wan Hai Line, K-line, NYK, MOL, Hanjin, and Hyundai Marine. The sample was not stratified according to nationalities, although stratified sampling would have been the preferred approach if the main purpose were to compare national cultures. Container ships were selected randomly from the Port of Kaohsiung in Taiwan. The Port of Kaohsiung, located in the major trade routes - Eastern Asian coastal, Far East/Europe and Transpacific service lines, has been ranked the top 10 container port in the world since 1980 (UNCTAD, 2008). The data collection took place from the beginning of March to the end of May 2008. A total of 2,232 questionnaires were distributed to the target respondents who worked on container vessels. Overall, 773 respondents from 13 countries took part in the study, but only three countries had more than 100 respondents, which included the Philippines (267), Taiwan (208), and China (133). The small samples of respondents from certain countries were excluded in our data analyses because they were far from being effectively representative of the population. These respondents came such countries as Myanmar (29), Korea (24), India (21), Russia (15), Ukraine (12), Japan (3), and others (59). Thus, a total number of 608 usable questionnaire returns were obtained from the survey for data analyses, yielding an effective response rate of 27.7%.

An analysis of non-response bias was conducted to assess the extent of the potential bias in the data collected. A comparison of early (those who were received before April 30, 2008) and late (those who were received behind the cutoff date) respondents was performed to test for a non-response bias (Armstrong and Overton, 1977). The 608 respondents were spilt into two groups according to their returned day, namely, early (n=357, 58.7 %) and late (n=251, 41.3%) respondents. T-tests were carried out on the responses of the two groups. At the 5% significance level, there was one attribute found significant difference between the responses of the two groups. The results do not rule out the possibility of a non-response bias and suggest that non-response may not be a problem to the extent that the late respondents represent the opinions of non-respondents.

There were 43.9% of the respondents in the survey were from Philippines, 34.2% were from Taiwan, and 21.9% were from China. Of which, 23.2% of respondent seafarers were deck officer, 33.6% were deck rating, 19.7% and 23.5% were engineer and engine rating, respectively. Twenty-two percent of respondents were serving for A carrier, whereas 16.9%, 16.3%, 14.3%, 7.6%, and 6.7% for B, C, D, E, and F carriers, respectively. Nearly 15% of respondents were serving for the other carriers. The respondents were also asked to provide their ages and provide of working experience of seafarers to ascertain whether they actually understood or appreciated the safety culture in the container shipping context. There were 32.2% of respondents less than 30 years old, 29.9% were between 31 and 40, 21.4% were between 41 and 50, 15.5% were between 51 and 60, and 1.3% were greater than 60 years old. Results indicated that over 60 per cent of respondents were less than 40 years old. Nearly 36% of respondents had been seafarers less than five years, nearly 20% between 6 and 10 years, and nearly 44% had more than 10 years. These suggested that they had abundant practical experience to answer the questions.

3.2. Measures

The data for this study were collected from a questionnaire survey. The items of safety culture used in this study were developed through the generation of an item pool from the previous studies (Cox and Cheyne, 1998; Gill and Shergill, 2004; Harvey et al., 1999; Hayes et al., 1998; Lu and Tsai, 2008; Naveh et al., 2005). Subsequent screened out 21 items based on the espoused values and the underlying assumptions of safety culture included both positively and negatively items. Items with difficult vocabulary or multiple negatives, and ambiguous means were avoided (DeVellis, 2003). There were four items employed to measure the seafarers' safety behavior. The validity of the contents of the questionnaire in current study was based on a comprehensive review of the literature and interviews and discussions with a number of seafarers in container shipping.

We developed the questionnaire for this study in both Chinese and English. The original version was in English, which was translated into an equivalent Chinese version. The questionnaire was then back translated into English. The respondents were asked to rate the agreement level of the measurement of both the safety culture and safety behavior in the container shipping context. These measures were rated using a five-point Likert scale, where 1 corresponds to "strongly disagree" and 5 to "strongly agree."

4. Results of empirical analyses

4.1. Exploratory Factor Analysis and Reliability Test

In order to test the posited structural model, measurements for each construct have to be obtained. The exploratory factor analysis (EFA), item-total correlations (or corrected item-total correlations), and estimations of reliability using Cronbach's alpha were conducted to develop and evaluate measurement scales. These techniques are useful in the early stages of empirical analysis, where theoretical models do not exist and the basic purpose in exploration. An EFA analysis was used to reduce the attributes of safety culture into a smaller, more manageable set of underlying factors.

There were 18 indicators used to measure safety culture in the container shipping context. Exploratory factor analysis with VARIMAX rotation was employed to reduce these attributes of safety culture into a smaller and manageable set of underlying dimensions. The Kaiser-Meyer-Olkin value of 0.919 indicated that the data were suitable conducting factor analysis, and the Bartlett Test of Sphericity [χ^2=6816.6, P < 0.000] suggested that correlations existed among some of the response categories. Eigenvalues greater than one were used to determine the number of factors in each data set (Churchill and Iacobucci, 2004; Gorsuch, 1983). Results shown in Table 1 indicate that three factors accounted for approximately 62.91% of the total variance and thus represented the safety culture. Moreover, an examination of loading factor in Table 1 reveals all items on each of the factors at 0.5 or higher. Consequently, there were three factors, found to underlie safety culture sets in container shipping context based on the seafarers' responses.

Factor 1, a safety management dimension, comprised nine items with factor loading ranging from 0.82 to 0.51. These items were related to safety management activities. "The safety programs in my company are good" and "Safety audit in my company is really practical" had highest factor loading on this factor. Factor 1 had an eigenvalue of 8.52 and accounted for 28.95% of the total variance.

Factor 2, a safety policy dimension, contained five items with factor loading ranging from 0.84 to 0.65. These five items were related to the attitude of safety policy. "Everybody attributes to the safety responsibility" had highest factor loading on this factor, followed by "all crew need adequate safety training", "safety is the top priority when I completing a job", "safety rules and procedures need to be followed to get the job done safely", and "my company encourages all people to raise safety concerns". This factor had an eigenvalue of 1.79 and accounted for 19.43% of the total variance.

Factor 3, a supervisor safety dimension, comprised four items with factor loading ranging from 0.74 to 0.63. The item, "My supervisor rewards for reporting unsafe condition" had highest factor loading on this factor, followed by "my supervisor halts operations when crew's safety is at risk", "my supervisor praises safety behavior" and "my supervisor often informs me of safety concerns and issues". Factor 3 had an eigenvalue of 1.01 and accounted for 14.53% of the total variance. The reliability values as shown in Table 1 were well above the suggested threshold of 0.7. These results suggest that all constructs are acceptable in exploratory analysis stage.

Table 1: Exploratory factor analysis of safety culture

Safety culture attributes	F1	F2	F3
The safety programs in my company are good.	0.82	0.19	0.22
Safety audit in my company is really practical.	0.80	0.23	0.18
The safety rules and procedures in my company are really practical.	0.79	0.23	0.28
The operational targets always comply with safety needs.	0.77	0.32	0.20
My company carries out safety audit periodically.	0.74	0.30	0.15
The safety training programs in my ship are useful.	0.61	0.32	0.27
The safety training programs in my ship help prevent accidents.	0.59	0.38	0.17
We always have enough people to get the job done safely.	0.57	0.11	0.38
My company allocates resources to meet safety needs.	0.51	0.10	0.37
Everybody attributes to the safety responsibility.	0.20	0.84	0.00
All crew need adequate safety training.	0.13	0.83	-0.03
Safety is the top priority when I completing a job.	0.25	0.70	0.27
Safety rules and procedures need to be followed to get the job done safely.	0.29	0.66	0.24
My company encourages all people to raise safety concerns.	0.33	0.65	0.21
My supervisor rewards for reporting unsafe condition.	0.07	0.01	0.74
My supervisor halts operations when crew's safety is at risk.	0.35	0.15	0.69
My supervisor praises safety behavior.	0.46	0.19	0.64
My supervisor often informs me of safety concerns and issues.	0.42	0.28	0.63
Eigenvalues	8.52	1.79	1.01
Percentage of Variance	28.95	19.43	14.53
Mean	4.03	4.35	3.76
S.D.	0.13	0.03	0.12
Cronbach Alpha	0.91	0.85	0.79

4.2. Confirmatory Factor Analysis

These traditional techniques (EFA, and Cronbach's alpha) do not assess unidimensionality (O'Leary-Kelly and Vokurka, 1998; Segars, 1997), nor can unidimensionality be explained by either mathematical or practical examinations (Gerbing and Anderson, 1988; Koufteros, 1999). Confirmatory factors analysis (CFA) with multiple-indicator measurement models was suggested to assess unidimensionality (Anderson et al., 1987; Segars, 1997).

Confirmatory factor analysis (CFA) was conducted on the entire set of items simultaneously (Anderson et al., 1987). Examining the significance of item loadings through t-tests to assess convergent validity. The overall fit of a hypothesized model can be assessed using fit indices such as the ratio of chi-square to degrees of freedom (χ^2/df), goodness-of-index (GFI), adjusted goodness-of-index (AGFI), comparative fit index (CFI), Tucker Lewis Index (TLI), root mean square residual (RMR), and root mean square error of approximation (RMSEA) (Byrne, 2001; Hu and Bentler, 1998). Potential misspecifications in the measurement model can be examined by reviewing standardized residuals, modification indices, and completely standardized expected change in Λx. Subsequent to the assessment of the measurement model, a structural model was evaluated in order to test the hypotheses that were postulated earlier. If the structural model fits the data adequately, the research hypotheses can be tested using the t-values of the structural coefficients.

The process of evaluating the measurement model resulted dropping two items (SP1 and SP3) from safety policy dimension, and the other two items (SM2 and SM7) from safety management dimension. These items were deleted iteratively based on the criteria such as large standardized residuals and modification indices. Before each deletion, the particular item and respective construct were evaluated from substantive point of view to assure that the loss of the item would not jeopardize the integrity of the construct in review.

The resulting measurement model had an acceptable model-to-data fit. The value of the Chi-square (χ^2) is 253.42 at 74 degree of freedom, which has a statistical significance at 0.00, below the minimum level of 0.05. Because the Chi-square value is sensitive to sample size (Bentler and Bonnet, 1980; Shah and Goldstein, 2006), the model fit was combined examination the other indices. The normed Chi-square ratio (χ^2/df) has a value of 3.42 in an acceptable level. The RMR indicates that the average residual correlation is 0.019, which is smaller than the threshold value of 0.05. The RMSEA is a reasonable value of 0.063. The GFI is 0.943, while AGFI is 0.919, which are good model fitness values. Both measures of incremental fit, the CFI is 0.961 and the TLI is 0.952, which are well greater than the recommended level of 0.90. Thus, acceptance can be given on this measure. Accompany the modification index statistic with the completely expected changes in the loading with other latent variables, found that no items exhibited change in Λx greater than 0.3 (Koufteros, 1999). These results did not justify an alternative specification.

Convergent validity can be tested by t-values that are all statistically significant on the factor loading (Dunn et al., 1994). The t-value is the critical ratio (C.R.) in the AMOS text output file, which represents the parameter estimate divided by its standard errors. A t-value greater than 1.96 or smaller than -1.96 implies statistical significance (Byrne, 2001; Segars, 1997). The larger the factor loading or coefficients, the stronger evidence that the measured variables or factor represent the underlying constructs (Koufteros, 1999). The completely standardized coefficients and t-values for the measurement model are all statistically significant at $p<0.01$. Therefore, all indicators are significantly related to their specified constructs, verifying the posited relationships among the indicators and constructs.

4.3. Results of Testing the Hypotheses Model

After confirming and establishing a good model fit for the measurement model of safety culture, this study proceeded to assess the proposed model and examine the hypothesized relationships. Regression analysis was conducted to assess the influence of the safety culture dimensions on safety behavior.

Before examining the effects of safety culture on the safety behavior, the influences of controlling demographic variables were first examined. The results revealed that no demographic variables had significant effect on the safety behavior. Subsequent, regressed the safety behavior on safety management, safety policy, and supervisor safety. Results of regression analysis in Table 2 demonstrated that the model had F value of 48.62 was at 0.00 significance level, implying that it fitted the data. VIF for safety management was 2.25, for safety policy 1.49, and for supervisor 1.85, respectively. These suggested that the three safety culture dimensions were orthogonal. The results in Table 2 also indicated that the t-value of each dimension was significant at 0.05 level. The safety policy (estimation=0.357) had strongest effect on the safety behavior, followed by supervisor safety (estimation=0.189), and safety management (estimation=0.174). These results provided evidences support for hypotheses H1, H2, and H3, that the dimensions safety culture (safety management, safety policy, and supervisor safety) had a positive significant effect on the safety behavior. The regression model with significant variables as follows:

G(x)=0.988 + 0.174(safety management) + 0.357(safety policy) + 0.189(supervisor safety)

Table 2: Results of regression analysis

Safety culture dimensions	Estimation of β	S.E.	t-value	P-value	VIF
Constant	0.988	0.248	3.991	0.00	
Safety management	0.174	0.078	2.243	0.03	2.25
Safety policy	0.357	0.062	5.756	0.00	1.49
Supervisor safety	0.189	0.064	2.967	0.00	1.85

Note: (1) S.E.= standard error. (2) R^2=0.195 significant at 0.00 level.

5. Conclusions and Implications

The safety culture has been recognized as an essential element in the organization’s efforts to prevent accidents in the workplace. But few researchers reached a consensus about its content, or how the safety culture is formulated within the organization. This study provides a theoretical rationale for a set of constructs (basic underlying assumptions and espoused value) of safety culture and showed adequate levels of reliability about it. The results indicate that the safety culture in the container shipping context includes three dimensions, namely, supervisor safety, safety policy, and safety management practice. These findings revealed that an organization has a policy to be safety and the supervisors could have a strong commitment toward safety formulate the safety management in container shipping context. The safety value and attitude will be shared to all members within the organization. The supervisors show a continuing interest in their employees’ working conditions, personally involve in the safety activities, and create a positive safety management system to improve safety.

This study not only identifies the underlying dimensions of safety culture in container shipping context, but also examines the relationships between safety culture and safety behavior. The results indicate that the safety culture has a positive significant effect on seafarers’ safety behavior. These findings provided evidences supporting that safety behaviors are influenced by the prevailing safety culture and are consistent with the previous studies that safety culture is influenced a safety behavior (Cooper and Philipps, 2004; Fernández-Muniz et al., 2007; Huang et al., 2006; Johnson, 2007; Seo, 2005; Siu et al., 2004).

This study explored the safety culture in the container shipping context. The results reveal that the artifacts phenomena (i.e., safety behavior) are significant affected by the underlying assumption and espoused value of safety culture. Future research may specify related performance variables such as vessel accidents rate, cargo damage ratio or crew injury frequency as dependent variables. It is possible that the safety culture may have highly related to the actual accident outcome. Researches also can emphasis on whether the organizational culture and national culture affect on the safety behavior. Future research also can extend the safety culture in other types of shipping context such as bulk carrier, general cargo, and tanker shipping context.

References

Advisory Committee on the Safety of Nuclear Installations (ACSNI) (1993), Study Group on Human Factors, Third Report: Organising for Safety. HSMO, London.

Anderson, J.C., Gerbing, D.W. and Hunter, J.E. (1987), On the assessment of unidimensional measurement: internal and external consistency, and overall consistency criteria. Journal of Marketing Research 24, 432-437.

Armstrong, S.J. and Overton, T.S. (1977), Estimating nonresponse bias in mail survey. Journal of Marketing Research 14(6), 396-402.

Bentler, P.M. and Bonnet, D.G. (1980), Significance tests and goodness-of-fit in the analysis of covariance structures. Psychological Bulletin 88, 588-606.

Bloor, M., Thomas, M. and Lane, T. (2000), Health risks in the global shipping industry: an overview. Health, Risk and Society 2(3), 329-340.

Brown, K.A., Willis, P.G. and Prussia, G.E. (2000), Predicting safe employee behavior in the steel industry: development and test of a sociotechnical model. Journal of Operations Management 18 (4), 445–465.

Byrne, B.M. (2001), Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming. New Jersey: Lawrence Erlbaum Associates.

Churchill, G.A. and Iacobucci, D. (2004), Marketing Research: Methodological Foundations, 9th edition. South-Western, USA.

Clarke, S. (1998), Safety culture on the UK railway network. Work & Stress 12(3), 285-292.

Cooper, M.D. and Phillips, R.A. (2004), Exploratory analysis of the safety climate and safety behavior relationship. Journal of Safety Research 35 (5), 497-512.

Cox, S. and Cheyne, A. (1998), Measuring safety culture in offshore environments offshore technology report. London: Health and Safety Executive.

Cuny, X. and Lejeune, M. (2003), Statistical modeling and risk assessment. Safety Science 41, 29-51.

DeJoy, D., Murphy, L. and Gershon, R. (1995), The in.uence of employee, job/task, and organizational factors on adherence to universal precautions among nurses. International Journal of Industrial Ergonomics 16, 43–55.

DeVellis, R.F. (2003), Scale Development: Theory and Applications. Thousand Oaks, CA.: Sage.

Dufort, V.M. and Infante-Rivard, C. (1998), Housekeeping and safety: an epidemiological review. Safety Science 28, 127-138.

Dunn, S.C., Seaker, R.F. and Waller, M.A. (1994), Latent variables in business logistics research: scale development and validation. Journal of Business Logistics 15, 145-172.

Fernández-Muniz, B., Montes-Peón, J.M. and Vazquez-Ordás, C.J. (2007), Safety culture: Analysis of the casual relationships between its key dimensions. Journal of Safety Research 38, 627-641.

Fleming, M. (2005), Patient safety culture measurement and improvement: A ‘How To’ guide. Healthcare Quarterly 8, 14-25.

Flin, R. (2007), Measuring safety culture in healthcare: A case for accurate diagnosis. Safety Science 45, 653-667.

Flin, R. and Yule, S. (2004), Leadership for safety: industrial experience. Quality and Safety in Health Care 13 (Suppl. II), ii45–ii51.

Gerbing, D.W. and Anderson, J.C. (1988), An updated paradigm for scale development incorporating unidimensionality and its assessment. Journal of Marketing Research 25, 186-192.

Gill, G.K. and Shergill, G.S. (2004), Perceptions of safety management and safety culture in the aviation industry in New Zealand. Journal of Air Transport Management 10, 233-239.

Glendon, A.I. and Litherland, D.K. (2001), Safety climate factors, group difference and safety behavior in road construction. Safety Science 39, 157-188.

Glendon, A.I. and Stanton, N.A. (2000), Perspective on safety culture. Safety Science 34, 193-214.

Gorsuch, R.L. (1983), Factor analysis (2nd ed.). Hillsdale, NJ: Erlbaum.

Guldenmund, F.W. (2000), The nature of safety culture: A review of theory and research. Safety Science 34(1-3), 215-257.

Hale, A.R. (2000), Culture's confusions. Safety Science 34(1-3), 1-14.

Harvey, J., Bolam, H. and Gregory, D. (1999), How many safety cultures are there? The safety and Health Practitioner 17(12), 9-12.

Håvold, J.I. (2005), Safety-culture in a Norwegian shipping company. Journal of Safety Research 36, 441-458.

Hayes, B.E., Perander, J., Smecko, T. and Trask, J. (1998), Measuring perceptions of workplace safety: development and validation of the work safety scale. Journal of Safety Research 29, 145-161.

Health and Safety Commission (HSC) (1993), Third Report: Organizing for Safety. ACSNI Study Group on Human Factors (London: HMSO).

Hidden, A. (1989), Investigation into the Clapham Junction Railway Accident. Department of Transport, CM 820. London: HMSO.

Hoonakker, P., Loushine, T., Carayon, P., Kallman, J., Andrew, K.A., Michael, J. and Smith, M.J. (2005), The effect of safety initiatives on safety performance: a longitudinal study. Applied Ergonomics 36, 461-469.

Hu, L. and Bentler, P.M. (1998), Fit indices in covariance structure modeling: sensitivity to under-parameterized model misspecification. Psychological methods 3(4), 424-453.

Huang, Y.H., Ho, M., Smith, G.S. and Chen, P.Y. (2006), Safety climate and self-reported injury: Assessing the mediating role of employee safety control. Accident Analysis and Prevention 38, 425-433.

Hurst, N.W., Young, S., Donald, I., Gibson, H. and Muyselaar, A. (1996), Measures of safety management performance and attitudes to safety at major hazard sites. Journal Loss Prevention Process Industry 9(2), 161-172.

International Maritime Organization (IMO) (2007), Consolidated Text of the International Convention for the Safety of Life at Sea, 1974, and its Protocol of 1988: Articles, Annexes and Certificates. 4th edition, London.

Johnson, S.E. (2007), The predictive validity of safety climate. Journal of Safety Research 38, 511-521.

Koufteros, X.A. (1999), Testing a model of pull production: a paradigm for manufacturing research using structural equation modeling. Journal of Operations Management 17, 467-488.

Lu. C.S. and Tsai, C.L. (2008), The effects of safety climate on vessel accidents in the container shipping context. Accidents Analysis and Prevention 40(2), 594-601.

Maritime Knowledge Centre (2008), International Shipping and World Trade Facts and Figures.

Ministry of Transportation and Communication (MOTC) (Taiwan) (2004), Statistical Abstract of Transportation and Communication. http//www.motc.gov.tw.

Ministry of Transportation and Communication (MOTC) (Taiwan) (2006), Statistical Abstract of Transportation and Communication. http//www.motc.gov.tw.

Mitroussi, K. (2003), The evolution of the safety culture of IMO: a case of organizational culture change. Disaster Prevention and Management 12(1), 16-23.

Naveh, E., Navon, T.K. and Stern, Z. (2005), Treatment errors in healthcare: a safety climate approach. Management Science 51(6), 948-960.

Neal, A., Griffin, M.A. and Hart, P.M. (2000), The impact of organizational climate on safety climate and individual behavior. Safety Science 34(1–3), 99-109.

Nieva, V. and Sorra, J. (2003), Safety culture assessment: a tool for improving patient safety in healthcare organizations. Quality and Safety in Health Care 12 (Suppl. II), ii17–ii23.

O'Leary-Kelly, S.W. and Vokurka, R.J. (1998), The empirical assessment of construct validity. Journal of Operations Management 16, 387-405.

Oliver, A., Cheyne, A., Tomás, J.M. and Cox, S. (2002), The effects of organizational and individual factors on occupational accidents. Journal of Occupational and Organizational Psychology 75, 473–488.

Ostrom, L., Wilhelmsen, C. and Daplan, B. (1993), Assessing safety culture. Nuclear Safety 34, 163-172.

Reason, J. (1998), Achiving a safe culture: theory and practice. Work and Stress 12(3), 293-306.

Roberts, S.E. and Marlow, P.B. (2005), Traumatic work related mortality among seafarers employed in British merchant shipping, 1976-2002, Occupation Environment & Medicine 62, 172-180.

Rose, N.L. (1992), Fear of flying: economic analyses of airline safety. Journal of Economic Perspectives 6, 75–94.

Rundmo, T., Hestad, H. and Ulleberg, P. (1998), Organisational factors, safety attitudes and workload among offshore oil personnel. Safety Science 29 (2), 75–87.

Segars, A. (1997), Assessing the unidimensionality of measurement: a paradigm and illustration within the context of information systems research. Omega-International Journal of Management Science 25(1), 107-121.

Seo, D.H. (2005), An explicative model of unsafe work behavior. Safety Science 43, 187-211.

Shah, R. and Goldstein, S.M. (2006), Use of structural equation modeling in operations management research: looking back and forward. Journal of Operations Management 24, 148-169.

Siu, O.L., Phillips, D.R. and Leung, T.W. (2004), Safety climate and safety performance among construction workers in Hong Kong: the role of psychological strains as mediators. Accident Analysis and Prevention 36, 359-366.

Sorensen, J.N. (2002), Safety culture: a survey of the state-of-the-art. Reliability Engineer and System Safety 76, 189-204.

Tomás, J.M., Melia, J.L. and Oliver, A. (1999), A cross-validation of a structural equation model of accidents: organizational and psychological variables as predictors of work safety. Work and Stress 13 (1), 49–58.

UK P & I Club (2004), Loss Prevention Claims Statistics, http://www.ukpandi.com/ukpandi/infoppol.nsf.

UK P & I Club (2005), Annual Report 2004, http://www.ukpandi.com/.

United Nations Conference on Trade and Development (UNCTAD) 2008, Review of Maritime Transport 2007. http://www.unctad.org/en/docs/rmt2007_en.pdf.

Wert, J. (1986), Safety culture at nuclear plants. An Unpublished Essay, as cited in O'Hara, P. (2003). 'Nuclear Safety Culture: An Organisational Development Model'. At http://nucelarsafetyculture.freeyellow.com.

Zohar, D. (2003), Safety climate: conceptual and measurement issues. In: Quick, J., Tetrick, L. (Eds.), Handbook of Occupational Health Psychology. American Psychological Association, Washington, DC, pp. 123–142.

Senior managers' safety leadership and safety performance in container terminal operations

Chin-Shan Lu[1] *and Chung-Shan Yang*[2]

[1]*Department of Transportation and Communication Management Science, National Cheng Kung University, Tainan City 701, Taiwan, ROC.*
Tel: +886-6-2757575 ext. 53243
Fax: +886-6-2753882
E-mail: lucs@mail.ncku.edu.tw.
[2]*Department of Transportation and Communication Management Science, National Cheng Kung University, Tainan City 701, Taiwan, ROC.*
Tel: +886-6-2757575 ext. 53271 ext.4050
Fax: +886-6-2753882
E-mail : r5895102@mail.ncku.edu.tw.

Abstract

This paper empirically evaluates safety leadership and safety performance in the context of container terminal operations. Results indicated that terminal workers agreed most that *my senior managers have established a safety responsibility system*, followed by *my senior managers show an interest in acting on safety policies*, *my senior managers are concerned about safety improvement, my senior managers establish clear safety goals* and *my senior managers coordinate with other departments to solve safety issues*. Based on respondents' agreement with safety leadership attributes, cluster analysis was subsequently employed to classify terminal workers into three groups, namely, safety responsibility and mission leadership oriented workers, safety reward leadership oriented workers, and safety support leadership oriented workers. This study found safety support leadership oriented workers had the best safety performance. Theoretical and practical implications of the research findings are discussed.

Keywords: safety leadership; safety performance; container terminal; cluster analysis.

1. Introduction

Evidence emerging from safety climate research clearly implicates the leadership process in the formation and maintenance of safety climate and the reduction of accidents (Ostrom et al., 1993; Rundmo and Hale, 2003). Studies have been conducted in high hazard industries such as construction sites (Dedobbeleer and Beland, 1991), chemical industry (Carder and Ragan, 2003), the nuclear power industry (Lee and Harrison, 2000), the aviation industry (Edkins, 1998), and the offshore oil and gas industry (O'Dea and Flin, 2003). These studies emphasize the importance of management, or factors which are under the direct control of management, in formulating a positive safety climate and good safety performance.

While many of the studies have examined managers' leadership behaviors (often called safety leadership) and safety, and focused on the supervisory level, few have investigated the role of senior level managers on safety outcomes. With a few notable exceptions (Hopkins, 1999; Carroll and Hatakenaka, 2001; Smallman and John, 2001; O'Dea and Flin, 2003), there is scant research on senior managers' leadership and safety performance in the context of container terminal operations.

Container terminals are key nodes in the sea transportation network and places where containers are loaded on to and discharged from containerships. Handling equipment such as the chassis-based transporter, straddle carrier, rubber tired gantry crane, rail mounted gantry crane or reach stacker are common in container terminals. Operations in container terminals are therefore very dangerous. According to the Occupational Safety and Health Administration (OSHA) (2001) in the United States, over 100 workers' deaths and nearly 95,000 injuries occur each year because of improper use of equipment in container terminal operations. Statistics from the UK P & I Club (1997, 2004), indicate that of the loss prevention claims made in recent years as a result of marine and port accidents, more than 53% were attributable to human error. Of these, 21%, 16%, 11%, 4% and 2% were deck officer, crew, shore person, pilot and engineering officer errors, respectively. A number of transport studies and reports similarly state that 60–90 percent of all accidents at sea or in the air can be attributed to the ''human factor'' (Mars, 1996; Sherry, 1992; Zohar, 1980).

One way to decrease human error or incidents is to have effective safety leadership. The United States Occupational Safety and Health Administration (OSHA, 1996) has tilized d the power of leadership and defined management leadership as a key element in safety system design. Safety leadership that motivates team members to work harder, to work efficiently, and to take ownership of safety is encouraged (O'Dea and Flin, 2001). The Health and Safety Executive (HSE, 2003) stated that without effective leadership one cannot have good safety performance. The Federal Safety Commissioner (2006) encouraged all senior managers to lead by example and to be consistent in the way they behave in relation to safety as senior managers' safety commitment is a key step in achieving a safety culture. The increasing number of safety leadership training programs instituted in various industries are evidence of the assumption that safety leadership will result in increased organizational safety effectiveness. Developing and sustaining safety leadership is increasingly important to reduce accidents and to promote safety among managers and general employees. Safety in container terminals starts with senior managers' safety leadership since their actions can help broaden safety awareness throughout the organization.

Due to its crucial importance for long-term safety success, this study aims to evaluate workers' perceptions of safety leadership and its impact on safety performance in the context of container terminal operations. This study is divided into five sections. The next section provides a review of the literature on safety leadership and safety performance which is followed by an explanation of the methodology employed to fulfill the research aim. The analysis of data obtained from a questionnaire survey of workers' perception is detailed in the fourth section. Conclusions drawn from the analysis and the associated practical implications are presented in the final section.

2. Literature Review

2.1. Safety Leadership

Leadership is crucial to improve safety performance (Mearns et al., 1997; O'Dea and Flin, 2001). Safety leadership is a sub-system of organizational leadership (Pater, 2001). Wu et al., (2007) defined safety leadership as "the process of interaction between leaders and followers, through which leaders can exert their influence on followers to achieve organizational safety goals under the circumstances of organizational and individual factors". Such visible leadership behaviors provide front line workers with an opportunity to meet leaders and discuss issues of concern at the workplace. Leaders in turn are provided with valuable insight on the safety issues and on how their guidance and direction support a safe workplace. Bryden (2002) identified a number of critical senior managers' behaviors for safety in an oil company, namely: articulating an attainable vision of future safety performance; demonstrating personal commitment to safety symbolically; engaging everyone with relevant experience in decision making; and being clear and transparent when dealing with safety issues.

Yule (2003) reported that leaders in the UK energy sector rated as transformational by their subordinates led business units with significantly lower injury rates. Stimulating, individually considerate, and rewarding styles were particularly influential for safety. Yule, Flin and Murdy (2007) investigated the association between senior managers' perceived leadership style and the safety performance of their units. Leadership behaviors, such as intellectual stimulation, idealized consideration and contingent reward, were found to be significantly associated with lower accident rates. The OSHA (1996) has emphasized that senior managers' most important behaviors relate to the value of safety. Cooper (1998) verified two important leadership behaviors: caring and controlling. Caring behavior refers to paying attention to the welfare of the organization's members, providing help when needed, establishing harmonious relationship with subordinates, maintaining effective communication channels, and offering anything beneficial to subordinates. Controlling behavior includes setting up goals, maintaining performance level, clarifying members' role, expectation, and duty, and encouraging members to follow regulations and procedures.

Carrillo and Simon (1999) proposed the Safety Culture Leadership Inventory (SCLI) which comprises six critical leadership practices: to make the case for change, to create a shared vision, to build trust and open communication, to develop capabilities, to monitor progress, and to recognize accomplishments. According to Bass and Avolio's (1990) study, the considerate leader expresses appreciation for good work, stresses the importance of job satisfaction, maintains and strengthens self-esteem of subordinates, is easy to approach, and takes subordinates' suggestions seriously.

O'Dea and Flin (2001) surveyed 200 offshore installation managers (OIMs) in charge of 157 production platforms and drilling rigs in the UK oil industry and identified four important issues concerning safety leadership: (1) visibility at the worksite and leading by example; (2) developing open, honest and trusting relationships with the workforce; (3) workforce involvement and empowerment in planning and decision-making, thereby increasing workforce ownership and responsibility of safety performance, and (4) being proactive about safety.

Because research on safety and accident prevention in the container terminal context is lacking, this study will identify crucial safety leadership attributes in order to examine their association with safety performance in container terminal operations through a review of the safety performance literature.

2.2. Safety Performance

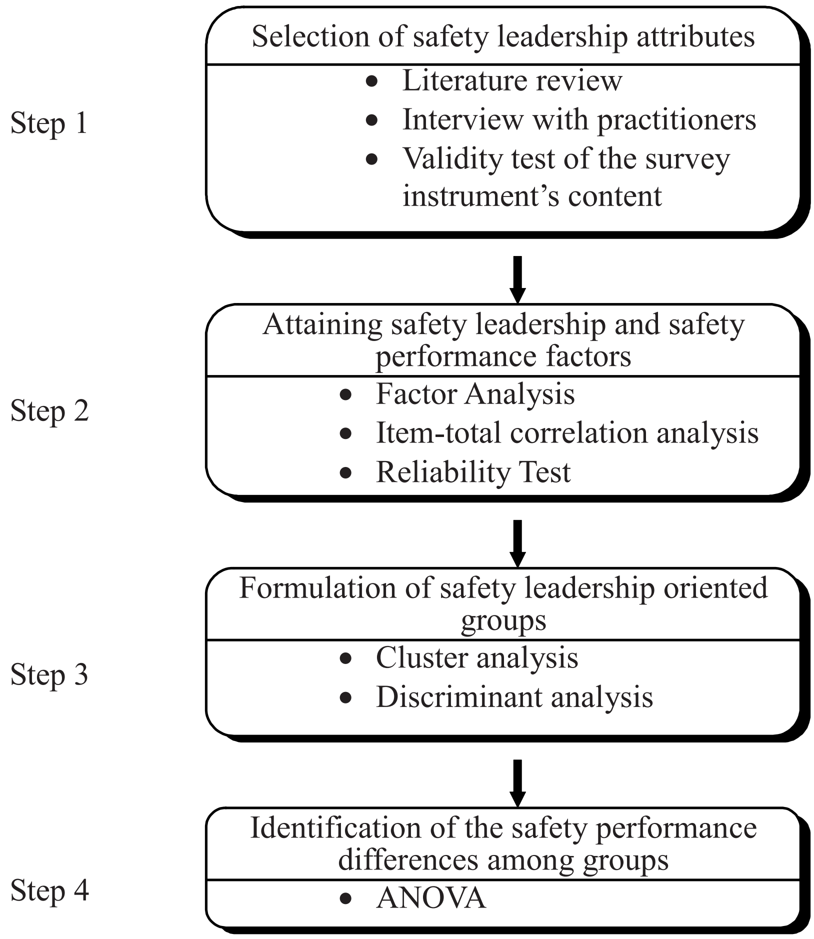
Safety performance can be measured in two ways: hard (objective) performance and soft (perceptual or responsive) performance (Dalton et al., 1980; Chow et al., 1994; Maltz and Maltz, 1998). Hard performance includes raw financial statistics, cost statistics, commissions, and services rendered, whereas soft performance involves supervisor appraisals and self-perceptions. In safety studies, the evaluation of safety performance is difficult and a major drawback in port safety research is the non-reporting of incidents, particularly safety failure events with no injury or fatality outcome. Moreover, accident or injury data (hard performance) are, according to several researchers (Glendon and Mckenna, 1995), "…insufficiently sensitive, of dubious accuracy, retrospective, and ignore risk exposure" (Glendon and Litherland, 2001: 161).

The Health and Safety Executive (HSE, 2001) pointed out that organizations need to recognize that there is no single reliable measure of safety performance. Because of the drawbacks associated with the use of injury data alone as a means of measuring performance, some researchers have recognized the need for more proactive measures of performance. Safety performance has been conceptualized as workers' behaviors that are intended to promote individual, organizational, and environmental safety (Burke et al., 2002). Perceptual-based measures (soft performance) have been recommended by several researchers (Neal et al., 2000) and are used in this study.

3. Methodology

3.1. Research Methods

The research was accomplished by carrying out a questionnaire survey. The research steps are illustrated below.



Step 1: Selection of safety leadership attributes

The first step was selecting safety leadership attributes by reviewing safety leadership and safety performance related literature, followed by designing the questionnaire through personal interviews with container terminal practitioners, and a content validity test. The questionnaire design followed the stages outlined by Churchill and Iacobucci (2002). Information to be sought was first specified, and then the following issues were settled: type of questionnaire and its method of administration, the content of individual questions, form of response to and wording of each question, sequence of questions, and physical characteristics of the questionnaire. When determining questionnaire items, it is crucial to ensure the validity of their content, since this is an important measure of a survey instrument's accuracy. Content validity is the extent to which a test measures what it is intended to measure (Cooper and Emory, 1995). The content validity of the questionnaire tilized in this study was tested through a literature review and interviews with practitioners, that is to say, questions in the questionnaire were based on previous studies and discussions with a number of container terminal executives and experts who included the chairman and safety supervisor of a Kaochun stevedoring company, the junior vice president of China Container Terminal Corporation, the harbor master and director of the stevedoring and warehousing department of Kaohsiung harbor bureau, the senior vice president of Evergreen International Storage and Transport Corporation, and the assistant vice president of the Taichung terminal of Wan Hai lines. Interviews resulted in minor modifications to the wording and examples provided in some measurement items, which were finally accepted as possessing content validity.

Step 2: Attaining safety leadership and safety performance factors

In the second step, *factor analysis* and *item total correlation analysis* were carried out in order to identify and summarize a large number of safety leadership and safety performance attributes into a smaller, manageable set of underlying factors or dimensions. A reliability test was conducted to assess whether the safety leadership and safety performance factors were reliable.

Step 3: Formulation of safety leadership oriented groups

In the third step, cluster analysis was performed. Through cluster analysis, three groups of terminal workers were identified based on their level of agreement with safety leadership factors. Ward's hierarchical technique using squared Euclidean distances was chosen to form clusters. Respondents were categorized into three groups on the basis of their factor scores.

Step 4: Identification of safety performance differences among groups

In the final step, ANOVA was used to test whether significant differences existed among the three terminal workers' groups' attitude toward the three safety leadership factors. Analysis was carried out using *SPSS 12.0* and the results are discussed in the next section.

3.2. Sampling Technique

This study focused on container terminal operators in Taiwan. One thousand and one hundred questionnaires were sent to workers at container terminal operators. The initial mailing elicited 238 usable responses. A follow-up mailing was sent two weeks after the initial mailing. An additional 98 usable responses were returned. The total number of usable responses was 336, therefore an overall response rate of 30.5 percent (336/1100). Each of the safety leadership and safety performance attributes used in the questionnaire were assessed by means of a five-point Likert scale, anchored by the level of agreement ranging from "1=strongly disagree" to "5=strongly agree."

3.3. Respondents' Characteristics

Respondents' profiles and characteristics are shown in Table 1. The majority of respondents (72.3%) were aged between 31 and 50 years. Nearly nineteen percent (18.5%) were aged 51 or more, and 9% were less than 30 years old. Table 1 also shows that 11.9% of survey participants were managers or above, and 20.6% were supervisors. Nearly three-quarters (68%) of respondents were workers. Since this study sought primarily to evaluate the perceptions of senior managers' safety leadership in container terminals, the views of workers were considered particularly useful.

Table 1 also indicates that 85% of respondents had attained senior high school and bachelor or above educational level. Just under a fifth (14.6%) of them had attained only a junior high school or below educational level. Respondents' work departments are also shown in Table 1. Results show operational was the main department (61.6%), followed by maintenance (26.5%), administration (10.1%), and others (1.8%). Results also indicated that over half of respondents (65.2%) had received safety training between once and three times, nearly 26% had received safety training more than four times, while 8.9% had never received any safety training at all. Table 1 also shows the type of company respondents worked in. More than half (62.2%) worked for a terminal operator, whereas 21.1% and 16.7% worked for a stevedoring and tally company, respectively.

Table 1: Profile of respondents (n=336)

Characteristics of respondents		Frequency	%
Age	Less than 30	31	9.2
	31~50	243	72.3
	More than 51	62	18.5
Job title	Manager or above	40	11.9
	Supervisor	69	20.6
	Worker	226	67.5
Educational level	Junior high school or below	49	14.6
	Senior high school	143	42.6
	Bachelor degree or above	144	42.9
Department	Operational	207	61.6
	Administration	34	10.1
	Maintenance	89	26.5
	Other	6	1.8
Frequency of safety training	Never	30	8.9
	1~3 times	219	65.2
	4~6 times	55	16.4
	7 times or more	32	9.5
Type of company	Terminal operator	209	62.2
	Stevedoring	56	16.7
	Tally	71	21.1

An evaluation of the aggregated perceptions of terminal workers of each safety leadership attribute revealed that their level of agreement with all 16 safety leadership attributes was at the upper end of the five-point interval scale, where 1 represented strongly disagree and 5 signified strongly agree. Table 2 shows respondents' agreement level with the safety leadership attributes in descending order of agreement. Notably, there were three safety leadership attributes with which respondents agreed most (their mean scores were all over 4.2): *my senior managers have established a safety responsibility system; my senior managers express an interest in acting on safety policies* and *my senior managers are concerned about safety improvement*. Respondents least agreed with the safety leadership attribute: *my senior managers have set up a safety incentive system* (its mean score was below 4.0).

Table 2: Respondents' agreement with safety leadership attributes

Ranking	Safety leadership attributes	Mean	SD
1	My senior managers have established a safety responsibility system.	4.33	0.69
2	My senior managers express an interest in acting on safety policies.	4.26	0.65
3	My senior managers are concerned about safety improvement.	4.21	0.67
4	My senior managers establish clear safety goals.	4.18	0.72
5	My senior managers coordinate with other departments to solve safety issues.	4.18	0.68
6	My senior managers explain the safety mission clearly.	4.16	0.62
7	My senior managers encourage workers to provide safety suggestions.	4.16	0.70
8	My senior managers emphasize worksite safety.	4.14	0.70
9	My senior managers stress the importance of wearing personal protective equipment.	4.12	0.70
10	My senior managers encourage workers' participation in safety decision-making.	4.07	0.75
11	My senior managers encourage workers to report potential incidents without punishment.	4.06	0.81
12	My senior managers show consideration for workers.	4.06	0.83
13	My senior managers trust workers.	4.05	0.81
14	My senior managers praise workers' safety behavior.	4.01	0.86
15	My senior managers reward those who set an example in safety behavior.	4.01	0.83
16	My senior managers have set up a safety incentive system.	3.95	0.92

Note: The mean scores are based on a five-point scale (1= strongly disagree to 5= strongly agree); SD = standard deviation.

Respondents were also asked to indicate their level of agreement with 18 safety performance attributes. An evaluation of the aggregated perceptions of each safety performance attribute revealed that respondents' level of agreement with all 18 safety performance attributes was at the upper end of the five point scale (their mean scores were all over 3.89 or above).

Table 3: Respondents' agreement with safety performance attributes

Ranking	Safety performance attributes	Mean	SD
1	I maintain safety awareness at work.	4.38	0.53
2	I comply with safety rules and standard operational procedures.	4.31	0.54
3	I do not neglect safety even when in a rush.	4.26	0.62
4	I wear personal protective equipment at work.	4.21	0.62
5	My company carries on safety policies.	4.18	0.69
6	My workplace has good illumination and ventilation.	4.14	0.70
7	My company has written safety policies.	4.14	0.71
8	I participate in setting safety goals.	4.12	0.64
9	My company addresses safety rules frequently.	4.12	0.70
10	The number of personal injuries is reducing.	4.09	0.74
11	My workplace is neat and tidy.	4.08	0.73
12	I would like to provide safety improvement suggestions.	4.07	0.67
13	The frequency of accidents is reducing.	4.05	0.74
14	I actively participate in safety meetings.	4.04	0.66
15	Equipment and tools in my workplace are deployed in a safe manner.	4.01	0.67
16	My workplace layout is designed with safety in mind.	3.96	0.75
17	The value of cargo loss and damage is reducing.	3.92	0.75
18	The frequency of equipment failure is reducing.	3.89	0.75

Note: Mean scores are based on a five-point scale (1= strongly disagree to 5= strongly agree); SD = standard deviation.

Table 3 shows respondents' level of agreement with safety performance attributes in descending order of agreement. Notably, there were four safety performance attributes with which respondents agreed most (their mean scores were all over 4.2): *I maintain safety awareness at work, I comply with safety rules and standard operation procedures, I do not neglect safety even when in a rush* and *I wear personal protective equipment at work*. In contrast, respondents least agreed with the following three safety performance attributes: *the frequency of equipment failure is reducing, the value of cargo loss and damage is reducing*, and *my workplace layout is designed with safety in mind* (their mean scores were all below 4).

4. Results of Empirical Analyses

4.1. Factor Analysis

Factor analysis was used to reduce the 16 safety leadership attributes to a smaller, manageable set of underlying factors. In order to detect the presence of meaningful patterns among the original variables and extract the main service factors. Principal component analysis with VARIMAX rotation was employed to identify safety leadership factors. To aid interpretation, only variables with a factor loading of greater than 0.50 were extracted, a conservative criterion based on Hair, et al. (1995). Eigenvalues greater than one were used to determine the number of factors in each data set (Churchill and Iacobucci, 2002). The percentage of variance for each of the three factors identified is shown in Table 4. The total percentage of variance can be used as an index to determine how well a particular factor solution accounts for what all the variables together represent. If the variables fall into one or more highly redundant or related groups, and if the extracted factors account for all the groups, the index will approach 100 percent. Factor analysis showed that approximately 75 percent of the total variance was represented by the information contained in the factor matrix, thus could be considered to represent all the safety leadership attributes (Hair, et al., 1995). Three factors were found to underline the 16 safety leadership attributes of container terminal operations based on respondents' responses to the survey. The three safety leadership factors were labeled and are described below:

Factor 1 (SL1) was called the *safety reward factor* since it comprised the following seven items: *my senior managers trust workers, my senior managers reward those who set an example in safety behavior, my senior managers praise workers' safety behavior, my senior managers have set up a safety incentive system, my senior managers encourage workers to report potential incidents without punishment, my senior managers encourage workers to provide safety suggestions,* and *my senior managers encourage workers' participation in safety decision-making.* This factor accounted for 59.66 percent of the total variance. Further, *my senior managers praise workers' safety behavior* and *my senior managers reward those who set an example in safety behavior* had the highest factor loadings on this factor.

Table 4: Exploratory factor analysis of safety leadership attributes

Safety leadership attributes	SL1	SL2	SL3
My senior managers trust workers.	<u>0.672</u>	0.379	0.318
My senior managers reward those who set an example in safety behavior.	<u>0.765</u>	0.362	0.213
My senior managers praise workers' safety behaviors.	<u>0.816</u>	0.209	0.262
My senior managers have set up a safety incentive system.	<u>0.766</u>	0.292	0.289
My senior managers encourage workers to report potential incidents without punishment.	<u>0.673</u>	0.309	0.384
My senior managers encourage workers to provide safety suggestions.	<u>0.739</u>	0.133	0.403
My senior managers encourage workers' participation in safety decision-making.	<u>0.730</u>	0.237	0.381
My senior managers explain the safety mission clearly.	0.216	<u>0.797</u>	0.298
My senior managers emphasize worksite safety.	0.241	<u>0.784</u>	0.203
My senior managers have established a safety responsibility system.	0.282	<u>0.819</u>	0.167
My senior managers establish clear safety goals.	0.295	<u>0.765</u>	0.284
My senior managers stress the importance of wearing personal protective equipment.	0.436	0.242	<u>0.679</u>
My senior managers express an interest in acting on safety policies.	0.261	0.237	<u>0.815</u>
My senior managers are concerned about safety improvement.	0.358	0.254	<u>0.797</u>
My senior managers coordinate with other departments to solve safety issues.	0.288	0.245	<u>0.821</u>
My senior managers show consideration for workers.	0.392	0.276	<u>0.697</u>
Eigenvalues	9.545	1.384	1.091
Percentage variance	59.655	8.648	6.821
Cumulative variance	59.655	68.303	75.124

Factor 2 (SL2) was designated the *safety responsibility and mission factor* since it consisted of the following four items: *my senior managers explain the safety mission clearly, my senior managers emphasize worksite safety, my senior managers have established a safety responsibility system*, and *my senior managers establish clear safety goals*. *My senior managers have established a safety responsibility system* had the highest factor loading on this factor, followed by *my senior managers explain the safety mission clearly*. Factor 2 accounted for 8.648 percent of the total variance.

Factor 3 (SL3) was called the *safety support factor* since it comprised the following five items: *my senior managers stress the importance of wearing personal protective equipment, my senior managers express an interest in acting on safety policies, my senior managers are concerned about safety improvement, my senior managers coordinate with other departments to solve safety issues,* and *my senior managers show consideration for workers*. The item *my senior managers coordinate the departments to solve safety problems* had the highest factor loading on this factor, followed by *my senior managers express to act on safety policies*. Factor 3 accounted for 6.821 percent of the total variance.

Factor analysis was also used to detect the presence of meaningful patterns among the original 18 safety performance attributes. Results showed that approximately 78 percent of the total variance was represented by the information contained in the factor matrix as shown in Table 5. Five factors were found to underlie the various sets of safety performance attributes of container terminal operations in the survey. The five safety performance factors are labeled and described below:

Factor 1 (SP1) was labeled the *loss and fatality reducing factor* since it comprised the following four items, namely *the frequency of accidents is reducing, the frequency of equipment failure is reducing, the value of cargo loss and damage is reducing,* and *the number of personal injuries is reducing.* This factor was ranked the most important factor by container terminal workers (see Table 5) and accounted for 48.03 percent of the total variance. *The frequency of equipment failure is reducing* and *the value of cargo loss and damage is reducing* had the highest factor loadings on this factor.

Factor 2 (SP2) was termed the *safety compliance factor* since it consisted of the following four items: *I maintain safety awareness at work, I do not neglect safety even when in a rush, I comply with safety rules and standard operational procedures*, and *I wear personal protective equipment at work.* These items represent attributes that are interactive in nature with safety compliance. *I maintain safety awareness at work* had the highest factor loading on this factor, followed by *I comply with safety rules and standard operational procedures*. Factor 2 accounted for 11.17 percent of the total variance.

Table 5: Exploratory factor analysis of safety performance attributes

Safety performance attributes	SP1	SP2	SP3	SP4	SP5
The frequency of accidents is reducing.	0.771	0.124	0.235	0.244	0.239
The frequency of equipment failure is reducing.	0.816	0.035	0.241	0.129	0.268
The value of cargo loss and damage is reducing.	0.832	0.171	0.279	0.151	0.128
The number of personal injuries is reducing.	0.780	0.176	0.162	0.323	0.094
I maintain safety awareness at work.	0.091	0.842	0.065	0.090	0.051
I do not neglect safety even when in a rush.	0.115	0.793	0.067	0.193	0.098
I comply with safety rules and standard operational procedures.	0.047	0.805	0.217	0.150	0.180
I wear personal protective equipment at work.	0.189	0.688	0.263	-0.027	0.299
My workplace has good illumination and ventilation.	0.218	0.168	0.790	0.208	0.031
My workplace is neat and tidy.	0.260	0.174	0.754	0.254	0.216
My workplace layout is designed with safety in mind.	0.252	0.114	0.839	0.171	0.181
Equipment and tools in my workplace are deployed in a safe manner.	0.180	0.169	0.726	0.222	0.244
My company has written safety policies.	0.235	0.150	0.287	0.743	0.282
My company carries on safety policies.	0.249	0.183	0.290	0.831	0.146
My company addresses safety rules frequently.	0.324	0.156	0.269	0.775	0.203
I participate in setting safety goals.	0.212	0.434	0.103	0.257	0.673
I would like to provide safety improvement suggestions.	0.245	0.123	0.163	0.244	0.811
I actively participate in safety meetings.	0.217	0.212	0.334	0.134	0.757
Eigenvalues	8.645	2.010	1.294	1.069	1.007
Percentage variance	48.027	11.167	7.186	5.938	5.593
Cumulative variance	48.027	59.194	66.380	72.319	77.912

Factor 3 (SP3) was called the *safety workplace factor* since it comprised the following four items: *my workplace has good illumination and ventilation, my workplace is neat and tidy, my workplace layout is designed with safety in mind,* and *equipment and tools in my workplace are deployed in a safe manner*. The item *my workplace layout is designed with safety in mind* had the highest factor loading on this factor, followed by *my workplace has good illumination and ventilation*. Factor 3 accounted for 7.186 percent of the total variance.

Factor 4 (SP4) was labeled the *safety policies factor* as it consisted of the following three items: *my company has written safety policies, my company carries on safety policies,* and *my company addresses safety rules frequently.* This factor accounted for 5.938 percent of the total variance. *My company carries on safety policies* had the highest factor loading on this factor.

Factor 5 (SP5) was called the *safety participation factor* since it comprised the following three items: *I participate in setting safety goals, I would like to provide safety improvement suggestions,* and *I actively participate in safety meetings.* These items represent attributes that are interactive in nature with safety participation. It accounted for 5.593 percent of the total variance. The item *I would like to provide safety improvement suggestions* had the highest factor loading on this factor, followed by *I actively participate in safety meetings*.

4.2. Item-total Correlation

Item-total correlation refers to the correlation of an item or indicator with the composite score of all the items forming the same set. Items from a given scale exhibiting item-total correlations less than 0.50 are usually candidates for elimination. If the items in a measure are drawn from the domain of a single construct, responses to those items should be highly intercorelated (Churchill and Iacobucci, 2002). As can be seen in Table 6, the item-total correlation of the 16 items safety leadership belonging to the three different safety leadership factors ranged in values from 0.725 to as high as 0.843.

A reliability test based on the Cronbach Alpha statistic was employed to test whether the items under the factors were consistent and reliable. The Cronbach Alpha statistic for each of the three factors was well above 0.70, considered a satisfactory level of reliability in basic research (Nunnally, 1978; Carmines and Zeller, 1979; Sekaran, 1992; Churchill and Iacobucci, 2002; Litwin, 1995). Table 6 also shows that respondents agreed most with items under the *safety responsibility and mission factor* (Factor 2), followed by those under the *safety support factor* (Factor 3) and then the *safety reward factor* (Factor 1).

Table 6: Reliability test of safety leadership factor items

Factors	Cronbach Alpha	Corrected item-total correlation	Alpha if item deleted
SL1: Safety reward (Mean=4.044; S.D.=0.692)	0.935		
My senior managers trust workers.		0.768	0.927
My senior managers reward those who set an example in safety behavior.		0.801	0.924
My senior managers praise workers' safety behaviors.		0.812	0.923
My senior managers have set up a safety incentive system.		0.810	0.924
My senior managers encourage workers to report potential incidents without punishment.		0.776	0.926
My senior managers encourage workers to provide safety suggestions.		0.779	0.927
My senior managers encourage workers' participation in safety decision-making.		0.800	0.925
SL2: Safety responsibility and mission (Mean=4.204; S.D.=0.591)	0.889		
My senior managers explain the safety mission clearly.		0.775	0.853
My senior managers emphasize worksite safety.		0.725	0.869
My senior managers have established a safety responsibility system.		0.768	0.853
My senior managers establish clear safety goals.		0.765	0.854
SL3: Safety support (Mean=4.166; S.D.=0.619)	0.921		
My senior managers stress the importance of wearing personal protective equipment.		0.760	0.910
My senior managers express an interest in acting on safety policies.		0.794	0.905
My senior managers are concerned about safety improvement.		0.843	0.895
My senior managers coordinate with other departments to solve safety issues.		0.841	0.895
My senior managers show consideration for workers.		0.771	0.913

As can be seen in Table 7, the item-total correlation of the 18 items belonging to the five different safety performance factors ranged in value from 0.648 to as high as 0.854. The Cronbach Alpha statistic for each of the five factors was well above 0.70, regarded as a satisfactory level of reliability in basic research as previously stated.

Table 7: Reliability test of safety performance factor items

Factors	Cronbach Alpha	Corrected item-total correlation	Alpha if item deleted
SP1: Loss and fatality reducing (Mean=3.996; S.D.=0.660)	0.910		
The frequency of accidents is reducing.		0.797	0.884
The frequency of equipment failure is reducing.		0.796	0.884
The value of cargo loss and damage is reducing.		0.827	0.873
The number of personal injuries is reducing.		0.765	0.895
SP2: Safety compliance (Mean=4.291; S.D.=0.479)	0.845		
I maintain safety awareness at work.		0.683	0.804
I do not neglect safety even when in a rush.		0.665	0.812
I comply with safety rules and standard operational procedures.		0.742	0.779
I wear personal protective equipment at work.		0.648	0.819
SP3: Safety workplace (Mean=4.049; S.D.=0.623)	0.895		
My workplace has good illumination and ventilation.		0.730	0.878
My workplace is neat and tidy.		0.796	0.853
My workplace layout is designed with safety in mind.		0.836	0.837
Equipment and tools in my workplace are deployed in a safe manner.		0.710	0.885
SP4: Safety policies (Mean=4.077; S.D.=0.575)	0.852		
My company has written safety policies.		0.693	0.822
My company carries on safety policies.		0.748	0.770
My company addresses safety rules frequently.		0.729	0.788
SP5: Safety participation (Mean=4.145; S.D.=0.646)	0.909		
I participate in the safety goals setting.		0.772	0.907
I would like to provide safety improvement suggestions.		0.854	0.838
I participate in safety meeting actively.		0.826	0.861

Table 7 also shows that respondents agreed most with item under the *safety compliance factor* (Factor 2), followed by those under the *safety participation factor* (Factor 5), *safety policies factor* (Factor 4), *safety workplace factor* (Factor 3), and *loss and fatality reducing factor* (Factor 1).

4.3. Cluster Analysis Results

Cluster analysis with a dendogram using the Ward method was conducted. The clustering procedure began with the 336 respondents and continued until all the respondents became again an undifferentiated group as shown in Figure 2. The solid horizontal line shows the point at which the clustering solution best represents all respondents (Hair, et al., 1995). The 336 respondents were assigned to three groups: 125 in group 1, 141 in group 2, and 70 in group 3.

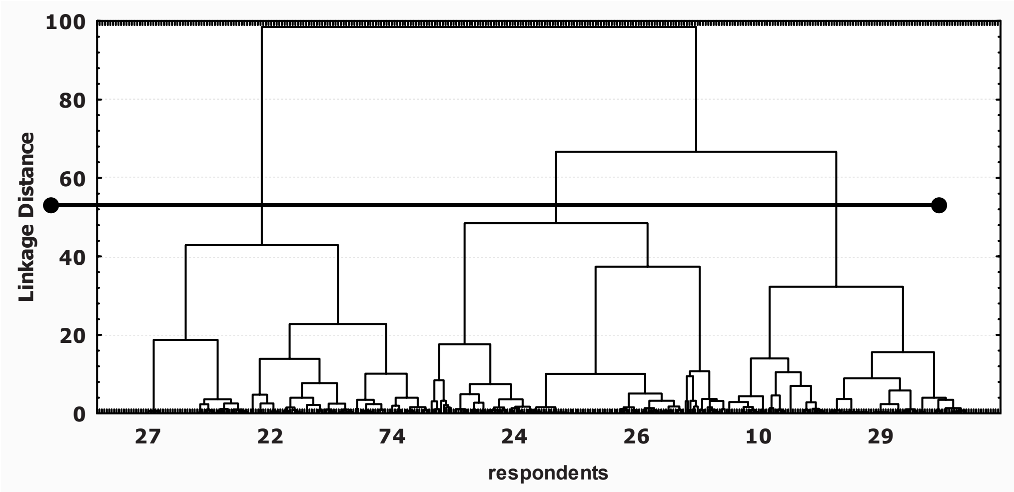


Figure 2: Tree diagram for 336 respondents using the Ward method

A canonical discriminant function (Klecka, 1980) was conducted to demonstrate the nature of group differences. As can be seen in Table 8, the two functions can explain 100% of the variance. It should be noted that the two discriminant functions represent the following:

$Y1 = 0.620\ SL1 - 0.794\ SL2 - 0.602\ SL3$

$Y2 = -0.344\ SL1 + 0.578\ SL2 - 0.800\ SL3$

Where Y is the discriminant score.

Table 8: Canonical discriminant function coefficients

Factors	1	2
SL1: safety reward	0.620	-0.344
SL2: safety responsibility and mission	-0.794	0.578
SL3: safety support	-0.602	-0.800
% of variance	54.8	45.2
Cumulative %	54.8	100

Table 8 indicates that the first function can explain 54.8 percent of variance. Figure 3 presents the plot of the two discriminant functions to visually display differences among the three groups.

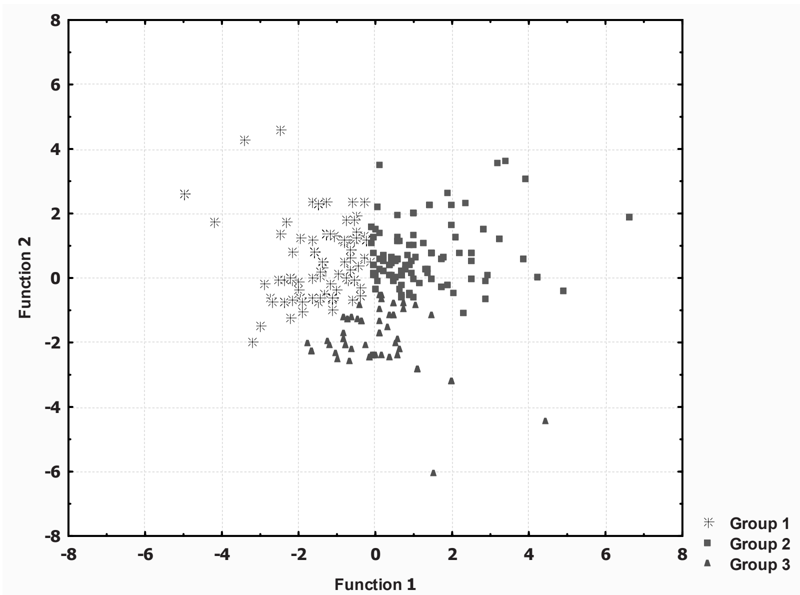


Figure 3: Canonical discriminant functions

4.4. Interpretation of Clusters

One-way analysis of variance was then used to examine whether the three safety leadership factors differed among the three container terminal workers' groups. Table 9 shows the ANOVA test results in terms of factor score coefficients. The three safety leadership factors were found to significantly differ among the three groups at the $p < 0.05$ significance level.

Table 9: ANOVA of safety leadership factors' differences among the three groups

Factors	group1	group 2	group 3	F Value	F Prob.	Tukey's HSD test
Safety reward	-0.489	<u>0.238</u>	0.394	28.404	0.000*	(1,2)(1,3)
Safety responsibility and mission	<u>0.785</u>	-0.312	-0.772	109.005	0.000*	(1,2)(1,3)(2,3)
Safety support	0.265	-0.773	<u>1.083</u>	87.527	0.000*	(1,2)(1,3)(2,3)

Note: a. The description of groups is based on factor scores with a mean of zero and standard deviation of one.
b. Factor scores were derived from data pooled across the three groups.
c. *Significance level $p < 0.05$

Table 9 also shows the three safety leadership factors significantly differed between the three groups based on Tukey's Honestly Significant Difference (Tukey's HSD) tests. Factor 2 (safety responsibility and mission) and the factor 3 (safety support) significantly differed between groups 1 and 2, groups 1 and 3, and groups 2 and 3. Factor 1 (safety reward) significantly differed between groups 1 and 2 and groups 1 and 3. A comparison of factor score coefficients showed that group 1 had its highest centroid scores on *safety responsibility and mission* (mean=0.785), while group 2 had its highest positive score on *safety reward* (mean=0.238). Group 3 had its highest score on *safety support* (mean=1.083). However, it had a negative score on *safety responsibility and mission* as shown in Table 9. From Tukey's HSD test results and the comparison of factor score coefficients, three groups of terminal workers emerged based on the three safety leadership factors: *safety responsibility and mission leadership oriented workers, safety reward leadership oriented workers,* and *safety support leadership oriented workers.*

Table 10: ANOVA of safety performance differences among the three groups

Safety performance	group1	group 2	group 3	F Value	F Prob.	Tukey HSD test
SP1: Loss and fatality reducing	0.093	-0.120	0.076	1.263	0.284	NA
SP2: Safety compliance	0.079	-0.213	<u>0.287</u>	6.666	0.001*	(1,2)(2,3)
SP3: Safety workplace	-0.020	-0.150	<u>0.338</u>	5.883	0.003*	(1,3)(2,3)
SP4: Safety policies	<u>0.194</u>	-0.155	-0.034	4.371	0.013*	(1,2)
SP5: Safety participation	0.026	-0.120	<u>0.195</u>	2.551	0.080	NA

Note: a. The description of groups is based on factor scores with a mean of zero and standard deviation of one.
b. Factor scores were derived from data pooled across the three groups.
c. *Significance level $p < 0.05$

Table 10 shows the differences in safety performance factors among the three groups based on Tukey's Honestly Significant Difference (Tukey's HSD) tests. With the exception of SP1 (*loss and fatality reducing*) and SP5 (*safety participation*), significant differences were found between safety performance factors in groups 1, 2, and 3. SP4 (*safety policies*) was, however, found to significantly differ between groups 1 and 2. Comparing Table 9 and 10, group 3 workers mostly emphasized on *safety support* (mean=1.083) and had the best perceived safety performance. Conversely, group 2 workers emphasized *safety reward* (mean=0.238) but negatively emphasized *safety responsibility and mission* (mean= -0.312) and *safety support* (mean= -0.773) and had the worst perceived safety performance.

Table 11 shows the characteristics of the three groups of respondents in terms of age, job title, educational level, department, frequency of safety training, and type of company in container terminal industry. Twenty-seven percent, 31% and 14% of respondents were in groups 1, 2 and 3 reported they were aged between 31 and 50 years. Further, whereas in group 1, 6.57% of respondents were managers or above, only 2.69% of the respondents in groups 2 and 3 were managers or above. In addition, more than half of respondents in groups 1, 2 and 3 had attained senior high school or a Bachelor degree or above educational level. Less than 6% of the respondents in each group had attained junior high school level only. Table 11 shows the work departments of respondents in the three groups. Almost a quarter of respondents in group 1 (22.62%) worked in the operational department, followed by 9.52% in maintenance, 4.17% in administration, and 0.89% in others. Almost a quarter of respondents in group 2 (24.12%) worked in the operational department, followed by 13.10% in maintenance, 3.87% in administration, and 0.89% in other department. Regarding group 3, 14.88% of respondents worked in the operational department, followed by 3.87% in maintenance, and 2.08% in administration. As regards safety training, 1.19%, 2.98% and 4.76% of respondents in Group 3, 1 and 2 had never received it. Results also indicate that 25.00% of the respondents in groups 1 and 2 worked for terminal operator in contrast to 12.20% in group 3. Chi-square analysis was performed to test if the three groups differed in their characteristics with respect to their age, job title, educational level, work department, safety training and company type. Results indicated that *job title, safety training,* and *company type* differed significantly among the three groups at the $p < 0.05$ significance level.

Table 11: Chi-square tests of respondents profile among the three groups

	Groups						Chi-square test		
	(1)		(2)		(3)				
	N=125		N=141		N=70				
	No.	%	No.	%	No.	%	Value	(d.f).	P
Age									
Less than 30 years old	10	2.98	12	3.57	9	2.68			
31~50	92	27.38	104	30.95	47	13.99			
51 or more	23	6.85	25	7.44	14	4.17	1.74	(4)	0.783
Job title									
Manager or above	22	6.57	9	2.69	9	2.69			
Supervisor	29	8.66	33	9.85	7	2.09			
General employee	73	21.79	99	29.55	54	16.12	14.50	(4)	0.006*
Educational level									
Junior high school or below	19	5.65	14	4.17	16	4.76			
Senior high school	48	14.29	65	19.35	30	8.93			
Bachelor degree or above	58	17.26	62	18.45	24	7.14	7.94	(4)	0.093
Department									
Operational	76	22.62	81	24.12	50	14.88			
Administration	14	4.17	13	3.87	7	2.08			
Maintenance	32	9.52	44	13.10	13	3.87			
Others	3	0.89	3	0.89	0	0	6.25	(6)	0.396
Safety training									
Never	10	2.98	16	4.76	4	1.19			
1~3 times	82	24.40	100	29.76	37	11.01			
4~6 times	19	5.65	15	4.46	21	6.25			
7 times or more	14	4.17	10	2.98	8	2.38	16.59	(6)	0.011*
Type of company									
Terminal operator	84	25.00	84	25.00	41	12.20			
Stevedoring	25	7.44	24	7.14	7	2.08			
Tally	16	4.76	33	9.82	22	6.55	11.48	(4)	0.021*

Note: a. % = the number of each item / total number of responses.
b. *Significance level $p < 0.05$

5. Discussion and Conclusions

This study examined container terminal workers' level of agreement with 16 safety leadership attributes and 18 safety performance attributes. Main findings of this research based on a questionnaire survey are summarized below.

Respondents agreed most with the following three safety leadership attributes: *my senior managers have established a safety responsibility system; my senior managers express an interest in acting on safety policies* and *my senior managers are concerned about safety improvement*. They least agreed with the safety leadership attribute: *my senior managers have set up a safety incentive system*. The research thus suggests that senior managers should continue to focus on safety responsibility, safety policies and safety improvement and also consider setting up a safety incentive system to maintain safety leadership. Factor analysis confirmed the aforementioned findings and suggestions since the following three safety leadership factors were identified: *safety reward*, *safety responsibility and mission*, and *safety support*. According to item-total correlation analysis, factor 2 (*safety responsibility and mission*) was the most important safety leadership factor, followed by factor 3 (*safety support*) and factor 1 (*safety reward*).

Cluster analysis distinguished three groups of terminal workers based on their factor scores in the three safety leadership dimensions: *safety responsibility and mission leadership oriented workers*, *safety reward leadership oriented workers*, and *safety support leadership oriented workers*. Group 3 workers who emphasized safety support had the best perceived safety performance. Safety performance factors (safety compliance, safety workplace, and safety policies) significantly differed among the three terminal workers' groups. The results thus suggest that senior managers' ability to improve and promote safety performance is to great extent dependent on safety leadership.

One managerial implication of the findings is that senior managers should pay close attention to *safety support* related leadership to improve safety performance in container terminal operations, that is to say, senior managers should focus on "stressing the importance of wearing personal protective equipment", "expressing an interest in acting on safety policies", "showing concern about safety improvement", "coordinating with other departments to solve safety issues", and "showing consideration for workers". The second implication is that the key to enhancing safety performance is safety leadership training courses for senior managers. Senior managers can benefit from understanding what baseline leadership behaviors and practices are most likely to translate to positive safety performance.

To the author's knowledge, this study is one of the first to identify safety leadership attributes and evaluate safety performance factors among different groups in the container terminal context. It thereby offers an understanding of safety leadership from container terminal workers' perspective and also provides a general framework for senior managers to evaluate their safety leadership behaviors and their relative strengths and weaknesses. While this is a valuable study in the context of safety research, it has a number of limitations. First, this research limited itself to examining safety leadership within the container terminal industry in Taiwan. Future research could undertake the same investigation scope but perform a cross-industry study. Next, the study did not address the issue of cause and effect. The analysis of variance was adequate to identify a significant relationship between the various variables, however, structural equation modeling could be used to examine whether there are any cause and effect relationships between safety leadership and safety performance. Finally, this study was undertaken within a one year period. It would be helpful to undertake a longitudinal study and hence make comparisons over time.

References

Bass, B. and Avolio, B. (Eds.). (1994), Improving organizational effectiveness through transformational leadership, New York: Sage.

Bass, B. and Avolio, B. (1990), The implications of transactional and transformational leadership for individual, team and organizational development. Research in Organizational Change and Development 4, 231–72.

Bryden, R. (2002), Getting serious about safety: accountability and leadership—the forgotten elements. In: Proceedings of the Society of Petroleum Engineers' Health, Safety, Environment conference, Kuala Lumpur, March. (SPE 73940). Texas: SPE.

Burke, M.J., Sarpy, S.A., Tesluk, P.A. and Smith-Crowe, K. (2002), General safety performance: A test of a grounded theoretical model. Personnel Psychology 55, 429–457.

Carder, B. and Ragan, P.W. (2003), A survey-based system for safety measurement and improvement. Journal of Safety Research 34, 157-165.

Carmines, E.G. and Zeller, R.A. (1979), Reliability and Validity Assessment. Sage Publication, U.S.A.

Carrillo, R.A. and Simon, S.I. (1999), Leadership skills that shape and keep world-class safety cultures. In Proceedings of the 38th annual professional development conference (337–344). Baltimore, MD: ASSE.

Carroll, J. and Hatakenaka, S. (2001), Driving organizational change in the midst of crisis. Sloan Management Review, Spring, 70-79.

Chow, G., Heaver, T.D. and Henriksson, L.E. (1994), Logistics performance: definition and measurement. International Journal of Physical Distribution & Logistics Management 24, 17-28.

Churchill, G.A. and Iacobucci, D. (2002), Marketing Research: Methodological Foundation. 8th edn., South-Western, USA.

Cooper, D. (1998), Improving safety culture: A practical guide. Wiley, England.

Cooper, D.R. and Emory, C.W. (1995), Business Research Methods, 5th edn., Irwin, U.S.A.

Dalton, D.R., Todor, W.D., Spendolini, M.J., Fielding, G.J. and Porter, L.W. (1980), Organization structure and performance: a critical review. Academy of Management Review 5, 49-64.

Dedobbeleer, N. and Beland, F. (1991), A safety climate measure for construction sites. Journal of Safety Research, 22, 97-103.

Edkins, G.D. (1998), The INDICATE safety program: evaluation of a method to proactively improve airline safety performance. Safety Science 30, 275-279.

Glendon, A.I. and Litherland, D.K. (2001), Safety climate factors, group differences and safety behaviour in road construction. Safety Science 39, 157-188.

Glendon, A.I., and McKenna, E.F. (1995), Human Safety and Risk Management, Chapman and Hall, London.

Hair, J., Anderson, R., Tatham, R. and Black, W. (1995), Multivariate data analysis with readings. 4th edn, Englewood Cliffs, NJ: Prentice Hall International.

Health and Safety Executive (HSE), (2001), A guide to measuring health and safety performance.

Health and Safety Executive (HSE), (2003), The role of managerial leadership in determining workplace safety outcomes (No. 044), Norwich: Health and Safety Executive.

Hopkins, A. (1999), For whom does safety pay? The case of major accidents. Safety Science 32(2), 143-153.

Klecka, W.R. (1980), Discriminant Analysis, Series: Quantitative Applications in the Social Science. Sage University Paper, London.

Lee, T. and Harrison, K. (2000), Assessing safety culture in nuclear power stations. Safety Science 34, 61-97.

Litwin, M.S. (1995), How to Measure Survey Reliability and Validity, London: Sage Publications Ltd.

Maltz, A., and Maltz, E. (1998), Customer service in the distributor channel empirical findings. Journal of Business Logistics 19, 103-130.

Mars, G. (1996), Human factor failure and the comparative structure of jobs: the implications for risk management. Journal of Managerial Psychology 11, 4–11.

Mearns, K., Flin, R., Fleming, M. and Gordon, R. (1997), Human and organisational factors, offshore safety report (OTH 543), Suffolk: HSE Books.

Neal, A., Griffin, M.A. and Hart, P.M. (2000), The impact of organizational climate on safety climate and individual behavior. Safety Science 34, 99-109.

Neal, A. and Griffin, M. (2002), Safety climate and safety behaviour. Australian Journal of Management 27, 67-76.

Nunnally, J.C. (1978), Psychometric Theory. 2nd edn, New York: McGraw-Hill.

O'Dea, A. and Flin, R. (2003), The Role of Managerial Leadership in Determining Workplace Safety Outcomes. Research Report 044, Health & Safety Executive, UK.

O'Dea, A. and Flin, R. (2001), Site managers and safety leadership in the offshore oil and gas industry. Safety Science 37, 39–57.

Occupational Safety and Health Administration (OSHA), (1996), Program Evaluation Profile (PEP). http://www.osha.gov/SLTC/safetyhealth/pep.html

Occupational Safety and Health Administration (OSHA), (2001), News Release: Improper forklift training contributes to injury of cosmetic worker. http://www.osha.gov/pls/oshaweb/owadisp.show_document?p_id=385&p_table=NEWS_RELEASES

Federal Safety Commissioner, (2006), Leaders in Safety: A guide to developing senior management safety behaviours in the building and construction industry. http://www.fsc.gov.au/ofsc/Forindustry/Leadership/Leadership.htm

Ostrom, L., Wilhelmsen, C. and Daplan, B. (1993), Assessing safety culture, Nuclear Safety 34 (2), 163-172.

Pater, R. (2001), Leadership skills for the 21st century. In Proceedings of 2001 ASSE professional development conference and exposition (CD-ROM). Anaheim, CA, Session 631.

Rundmo, T. and Hale, A.R. (2003), Managers' attitudes towards safety and accident prevention. Safety Science 41, 557-574.

Sekaran, U. (1992), Research Methods for Business. 2nd edn, New York: John Wiley & Sons.

Sherry, P. (1992), Peer involvement and behavioral safety: A case study. Paper at the Annual Meeting of the American Association of Railroads, Safety Conference, 8, June, Little Rock, AR.

Smallman, C. and John, G. (2001), British directors perspectives on the impact of health and safety on corporate performance. Safety Science 38(3), 227-239.

UK P & I Club. (1997), Analysis of major claims: Ten-year trends in maritime risk. Thomas Miller P. and I. Ltd.

UK P & I Club. (2004), Loss Prevention Claims Statistics, http://www.ukpandi.com/ukpandi/infoppol.nsf

Wu, T.C., Chen, C.H., and Li, C.C. (2007), A correlation among safety leadership, safety climate and safety performance. Journal of Loss Prevention in the Process Industries 6(3), 261-272.

Yule, S. (2003), Senior management influence on safety in the UK and US energy sectors. PhD Thesis, University of Aberdeen.

Yule, S., Flin, R. and Murdy, A. (2007), The role of management and safety climate in preventing risk-taking at work. Int. J. Risk Assessment and Management 7(2), 137–151.

Zohar, D. (1980), Safety climate in industrial organizations: theoretical and applied implications. Journal of Applied Psychology 65, 96 –102.

Analyzing the trade transport and the demand of multi-mode transport between China and Koreas

Jing Lu and Zhongzhen Yang

College of Transport Management, Dalian Maritime University, Dalian, 116026, China.

Abstract

Since the un-solved political issues between the South and North Koreas, the Sino Koreas trade goods are mainly transported by vessels. This leads to a simplex mode and the multi-mode cannot jointly work to serve the trade transport between China and Koreas. Thus, it is necessary to consider the potential of the improvement of the relationship in Korea peninsular proactively and study the land transport issues between China and Koreas. This paper firstly analyzes the Sino-Koreas bilateral trade and studies the corresponding supply and demand of the shipping lines. Secondly, questionnaire surveys are done on the companies that involve in the bilateral trade or trade transport between China and Koreas, and the surveyed data are used to analyze the mode choice behaviors of the shippers to establish a modal-split model. At last, the necessity of opening the land transport corridor is discussed and the modal splits of all modes are estimated.

Keyword: Korea peninsular, multi-mode transport, bilateral trade, shipping and transport

1. Trade between China and S. Korea

The trade volume between China and S. Korea increased from 5.03 billion USD in 1992 to 111.93 billion USD in 2005. The projected volume of 100 billion USD for 2008 was realized three years ahead. At present, S. Korea is the six biggest trade partner of China following USA, Japan, Hong Kong and ASEAN. In 2007, the trade volume further increased to 145.02 billion USD. Table 1 shows the trade situation between China and S. Korea. It can be seen that Chinese side has a big adverse balance, however, as the economic growth and construction of many Korean factories in China, the adverse balance decreases continually from 2005.

Table 1: Trade situation between China and S. Korea

Year	Trade Volume		Export from China		Import of China		Balance
	Billion USD	Increment (%)	Billion USD	Increment (%)	Billion USD	Increment (%)	
1989	8.9	214.0	4.7	437.9	4.2	114.7	0.5
1991	32.5	67.0	21.8	72.9	10.7	55.8	11.1
1993	82.2	63.5	28.6	18.9	53.6	104.3	-25.0
1995	169.8	44.8	66.9	51.9	102.9	40.6	-36.0
1997	240.5	20.3	91.2	21.5	149.3	19.6	-58.1
1999	250.4	17.4	78.1	24.9	172.3	14.7	-94.2
2000	345	37.8	112.9	44.6	232.1	34.7	-119.2
2001	359.1	34.1	125.2	10.9	233.9	26.8	-108.7
2002	440.71	22.8	154.97	23.8	285.74	22.2	-130.8
2003	632.3	43.4	201.0	29.4	431.3	50.1	-230.3
2004	900.7	42.5	278.2	38.4	622.5	44.3	-344.3
2005	1119.3	24.3	351.1	26.2	768.2	23.4	-417.1
2006	1180.3	5.4	485.71	38.3	694.59	-9.6	-208.9
2007	1450.2	22.9	630.27	29.8	819.85	18	-189.6

Sources: Homepage of Ministry of Commerce of China.

The "Research Report on Long and Medium Term Development Planning of the Cooperation between China and S. Korea", which is published in November in 2005, projected the cooperative objectives and the key fields. It forecasted that the trade volume in 2010 would reach 150 billion USD, and in 2012 (twenty years anniversary of the normalization of the bilateral relationship) would be 200 billion USD. In fact, in 2007 the trade volume has reached 145.02 billion USD, thus is can be said that the actual volumes in the future will overtake the projected ones.

From above analyses, it can be known that trade and economic cooperation will increase further between China and S. Korea and the good bilateral relationship will induce large amount of transport demand. Therefore, sea and land transport will have a big market.

2. Shipping Market between China and S. Korea

In 2005, 2.35 million TEU were shipped by the shipping lines between China and S. Korea. Among them, 1.35 million are east bound and 1.0 million are west bound, and 0.68 million are transshipped. Also in 2005, 1.086 million passengers were transported by the passenger-cargo vessel between China and S. Korea, which is 0.27 million more than that of the 2004. At present, 71 container vessels and 14 passenger-cargo vessels are being operated between China and S. Korea, 79 and 35 voyages are arranged per week respectively (YANG, 2008).

2.1. Container Liners between China and S. Korea

China has 16 companies to operate container liners between China and S. Korea, and with 38 vessels they operate 41 voyages and supply the capacity of 160,000 TEU weekly. S. Korea has 14 companies to operate container liners, and with 31 vessels they operate 36 voyages and supply the capacity of 150,000 TEU weekly. There are only 2 vessels of the third party that operate 2 voyages weekly.

The shipped volumes in 2004 and 2005 are 1.774 and 1.837 millions TEU respectively. Ports of Tianjin, Qingdao and Shanghai are the top three in terms of the transported containers. In 2005, they handled 0.395 million, 0.391, million and 0.374 million TEU respectively (YANG, 2008).

2.2. Passenger-Cargo Vessel between China and S. Korea

There are 14 Sino-Korea joint-ventures to operate passenger-cargo liner between China and S. Korea whose shipping lines are shown in Figure 1. Total capacities for the containers, passengers and vehicles are 2,909TEU, 7,796 persons and 373 vehicles respectively. The shipped containers and passengers increases continually, the former increased from 0.148 million TEU in 2000 to 0.316 million TEU in 2005, while, the later increased from 0.485 million in 2000 to 1.086 million persons in 2005. Total passengers from 2000 to 2005 are 3.883 million persons.

Figure 1: Lines of Passenger-cargo Vessel

3. Trade and the Transport between China and N. Korea

3.1. Trade between China and N. Korea

China is the top country in terms of the foreign trade volume with N. Korea. Bilateral trade volume of China and N. Korea is 1.6 billion USD in 2005 and increased 14% compared with the former year. In 2005, N. Korea imported 1.0 billion USD of oil and corn from China and exported 0.5 billion USD of coal and iron ore. Since the economic policy adjustment of “7.1” in 2005 of N. Korea, lots of Chinese enterprises began to invest in N. Korea. In 2005, N. Korea attracted about 60 million USD foreign investments, while 52 million are from China.

3.2. Trade Transport between China and N. Korea

There are two land routes between China and N. Korea, which are located at Hunchun in Jilin province and Dandoing in Liaoning province respectively. Because the truck of Chinese side can only reach the supervision zones of N. Korea near its border, most of trucks used for trade transport are from N. Korea. Thus the land transport is very inconvenient for China, and the shippers in China have to use shipping lines.

Shipping lines between China and N. Korea are mainly operated between the ports in Liaoning and Nampo port. For example, a vessel with capacity of 400 TEU jointly own by Liaoning Danxing, Korean Trans and Germany Mill Cooperation runs between Dalian and Nampo twice a week. Another line between Dandong and Nampo is operated by Shifeng Shipping of N. Korea twice a week with an 180TEU vessel. In order to quicken the delivery speed and improve the logistic system in Dadong, and alleviate the congestion in Yalujiang Pass, Donggang Developing Zone Shipping operates a line from Dadong to Nampo in every Thursday.

Liaoning and Sinuiju are separated only by a small river (YaluJiang), in recent years trade transport across the river reaches 3 million tons, while 1.7 million tons are passed by vessels. After about twenty year’s construction, the two sides have 168 vessels, 6 docks for boundary trade delivery. Shipping lines of boundary trade delivery have the attributes of fixed lines, fixed vessel, short distance (4-40 nautical mile), short voyage time (0.5-5 hours), small capacity (capacity of 10-100 tons) and short annual operating period (no operation during January to March)

4. Demand Analyses of Multi-Mode Transport between China and Koreas

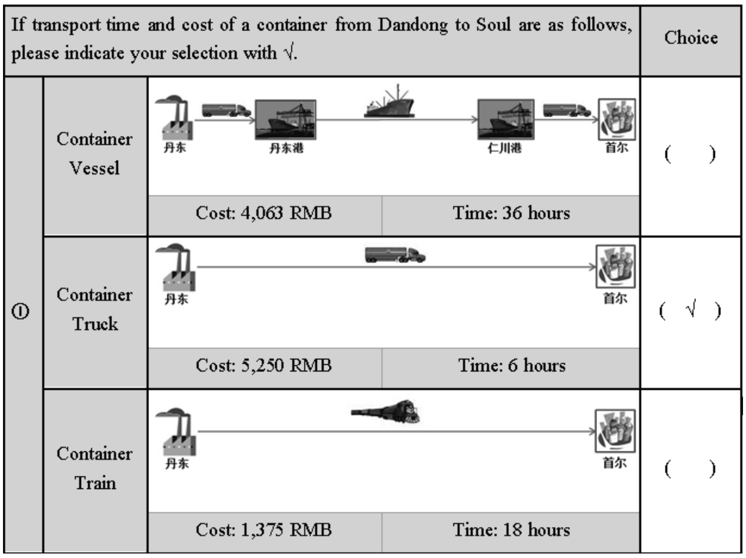
4.1. Construction of Multi-Mode Transport between China and Koreas

Sea transport is the main mode for long distance inter-continental transport, while it has some obvious disadvantages, such as over long needed time, easily affected by weather, needing transshipping at the terminals. Especially the transport cost of passenger-cargo vessels is expensive due to the adoption of some safety equipments and measures. Land transport has the advantages of rapid speed, high safety and feasibility, especially the road transport can offer door to door service. Therefore, the short distance and small batch delivery are mainly served with road transport. Moreover, because the rail transport is very safe and weather-resisting, it is at an advantageous position for the transport between 300 and 800 km (SHEN, 2003).

From Dandong to Pyongyang is about 200 km long, and from Pyongyang to Soul is also about 200 km long. If the political issues could be solved, multi-mode transport network should be formed in this area. Because the situation in Korea peninsular tends to be improved and N. Korea hopes to develop its economy through "Open and Reform", we think it is necessary to study on the construction of the multi-mode transport network. The integrated network can make full use of all transport modes and change the situation of over relying on the shipping mode. To construction land transport network, firstly the relationship between the South and the North must be improved. Without a normal relationship, N. Korea will not open more land passes to offer transfer points. And only with a normal relationship, the safety of rail and road transports between China and S. Korea can be guaranteed. In addition to the political issues, we should also analyze the behaviors of the shippers and the substitute affects of the land transport on the water one.

4.2. Modal Split of Trade Transport between China and Koreas

In order to analyze the modal splits in the context of a good land transport system and offer evidence for the construction of an integrated transport network between China and Koreas, we did questionnaire surveys at Dalian, Dandong and Yanji in China on 40 companies which are doing trade business with Koreas. During the survey, under the assumption of existence of a land transport system and unchanged water transport condition, the surveyed persons were asked to answer their choices when transporting containers from China to Koreas corresponding to different land transport prices and times. Part of the used questionnaire is shown in Figure 2.



If transport time and cost of a container from Dandong to Soul are as follows, please indicate your selection with √.				Choice
①	Container Vessel	丹东 → 丹东港 → 仁川港 → 首尔 Cost: 4,063 RMB	Time: 36 hours	( )
	Container Truck	丹东 → 首尔 Cost: 5,250 RMB	Time: 6 hours	(√)
	Container Train	丹东 → 首尔 Cost: 1,375 RMB	Time: 18 hours	( )

Figure 2: A Part of the Contents on the Questionnaire

The survey aims to understand the evaluation of the shippers on the modes. The evaluation may be determined by the price, timeliness, convenience and safety of the mode. And it also relates to the perception of the users on the attributes of the modes. It means that the utility of a mode is determined by the attributes of the mode and the users and can be expressed with Eq. 1.

$$U_{ij} = V_{ij} + \varepsilon_{ij}$$
$$V_{ij} = \alpha_{ij} + \sum_k \beta_{ijk} x_{ijk} + \sum_s r_{is} z_{is} \quad (1)$$

Where, U_{ij}=utility of service *j* perceived by user *i*, V_{ij}=deterministic term of U_{ij}, α_{ij}= constant term, x_{ijk}=preference of client *i* to attribute *k* of service *j*, z_{is}= attribute *s* of user *i*, β_{ijk}, r_{is}= parameter, ε_{ij}= random term (Ben-Akiva, Lerman, 1985).

The utility determines the chance that a mode is chosen, and the greater the utility the bigger the chance. When the random term ε_{ij} follows the distribution of Gumble, the chosen probability can be calculated with Logit model as shown in Eq. 2 (Ben-Akiva, Lerman, 1985).

$$P_{ij} = \mathrm{E}xp\,(V_{ij}) / \left(\sum_{k=1,\dots,K} \mathrm{E}xp\,(V_{ik})\right) \quad (2)$$

In this study, the size of the companies, the price and time of the mode are used as the viarables in the utility function and its detailed formation is as Eq. 3.

$$V_{ij} = \alpha_i + \beta_1 S_j + \beta_1 C_i + \beta_2 T_i \quad (3)$$

Where, S_j= scale of company *j* (classified by the number of the employees). C_i=price of mode *i* (RMB/TEU), T_i=time of mode *i* (hour). With the surveyed data, the utility model is calibrated and the results are shown in Table 2. It can be seen that the parameters have good statistic qualities and the calirated model can be used to analyze the modal splits between China and Koreas.

Table 2: Calibrated Results of the Model

	China - S. Korea			China - N. Korea			Sample Number
	Parameter	T-Value	R^2	Parameter	T-Value	R^2	
Constant	-0.3215049	-1.57	0.65	0.0177014	0.84	0.71	35
Size	-0.4512816	-3.54		-0.9329969	-4.42		
Cost	-0.0000970	-2.28		-0.0000295	-3.96		
Time	-0.0025686	-1.81		0.0321566	1.29		

Here we use the averaged values of the supposed land transport cost, time and company size in the questionnaires to analyze the modal splits. The detailed values and the estimated modal splits are listed in Table 3. It means that when a Chinese company with the size of 1.5 (1=with less than 100 employee, 2=with 100 to 500 employees) will send its 19% trade goods, 49% trade goods and 32% trade goods to S. Korea by road, sea and rail respectively, while it will send its 30% trade goods, 30% trade goods and 40% trade goods to N. Korea by road, sea and rail respectively, the prices and times of transporting a container to Koreas by road, sea and rail are as shown in Table 3.

Table 3: Average Value of the Price, Time and Company Size

	China - S. Korea			China - N. Korea		
	Road	Sea	Rail	Road	Sea	Rail
Price (RMB/TEU)	4,860	4,063	1,448	3,388	4,063	907
Time (Hour)	8	36	23	5	36	19
Company Size	1.5			1.5		
Modal Splits	19%	49%	32%	30%	30%	40%

5. Conclusion

From the analyses, it can be said that there are great trade transport demand between China and Koreas. However, at present almost all cargos have been transported by waterway. This kind of transport pattern is not reasonable. From the survey on the companies in China, it is found that shippers demand the multi-mode transport system, and many shippers will select road and rail transport if a suitable road or rail transport is available. Especially, between China and N. Korea the road, rail and water transport may be used evenly, while rail transport may play a more important role in the trade transport between China and S. Korea.

References

Ben-Akiva, Lerman Steven (1985). Discrete Choice Analysis: Theory and Application to Travel Demand, MIT Press.

Ministry of Commerce of China's website: http://english.mofcom.gov.cn/

SHEN, Z. (2003), Transportation Engineering, China Communication Press, Beijing China.

YANG, Z. (2008), Sea Transport between China and Korea in Pan Yellow Sea, Proceedings of Symposium of Sino-Korea Logistics Cooperative Research, Dalian, China, October 2008.

A dynamic-economic model for container freight market

Meifeng Luo, Lixian Fan and Liming Liu

Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hong Kong.

Abstract

This paper presents a dynamic-economic model analyzing the fluctuation of container freight rate due to the interactions between the demand for container transportation services and the container fleet capacity. The demand for container transportation services is derived from international trade and is assumed exogenous. The container fleet capacity increases with new orders made two years ago, proportional to the industrial profit. Assume market clears each year, the shipping freight rate will change with relative magnitude of demand and supply shifts.

The dynamic model is estimated using the world container shipping market statistics from 1980 to 2008, applying the three-stage least square method. The estimated parameters of the model have high statistical significance, and the overall explanatory power of the model is above 90%. The short-term in-sample prediction of the model can largely replicate the container shipping market fluctuation in terms of the fleet size dynamics and the freight rate fluctuation in the past 20 years. The prediction of the future market trend reveals that the container freight rate would continue decreasing in the coming three years if the demand for container transportation services grows less than 8%.

Keyword: container freight, economic-dynamic model, Empirical analysis, market forecast

1. Introduction

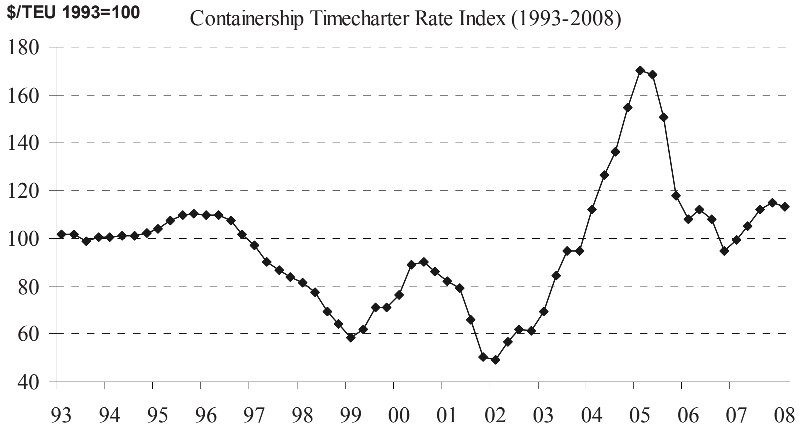
Within a short history of containerization, transportation of containerized goods by sea has thoroughly harnessed the possibility of trade among nations with different comparative economic advantages. Continued specialization and technological progress have boosted the efficiency in global shipping and port operation in the past two decades, making container transportation the indispensable soil for global trading firms to thrive in the increasingly competitive economic environment. Recent statistics shows that while world seaborne trade has been doubled from 3631 million tons in 1985 to 7852 million tons in 2007, containerized trade has increased almost 8 times within the same period from 160 million tons to 1257 million tons[1]. This demonstrates the increasing role of container transportation in the world seaborne trade, and its contribution to the global economy.

Regardless of the booming global seaborne trade and containerization, the fluctuated container freight rate, as shown by the container time charter rate in recent two decades (
Figure 1), unhinged the profitability in container shipping industry. The high demand for container shipping services hastened shipping companies to order new container vessels with bigger size and higher efficiency, to attract global customers with better services at lower cost, and to gain larger shares in the global competitive shipping market. Companies with the most up-to-date container vessels can out-perform the others with faster and reliable services at a lower unit cost. While reducing sea transportation cost can induce additional demand, it also agonize the companies with less efficient fleets. The usual practice of 'passing the rent to shippers' can decrease the freight rate, dissipate rent, diminish profit at the industry level, and even make some companies bankrupt. According to Drewry Shipping Consultants Ltd., comparing with 2006, global carriers moved 14.7% more cargo, but earned1.2 percent less revenue in 2008. On the main east-west trade routes, aggregated losses of the carriers amounted to

[1] Data source: Shipping Review and Outlook 2008

$2.4 billion, an 8% net loss. Maersk Line, the world largest shipping line with more than 16% of the world's liner fleet, has suffered a $568 million loss in 2006.

Figure 1: Container time charter rate index 1993-2008



The low freight rate in shipping cycle not only has significant negative impact on business operations and investment decision, but also brought extensive concerns at both national and international level. Bankers, who financed the building or purchasing of ships, bear high financial risks due to the insolvency of the ship-owners at low freight rate. According to Volk (1984), most of the ship investment activities are concurrent to the high freight rate. Goulielmos and Psifia (2006) point out that bankers financed 75-80% of the ship construction cost. Therefore, it is very important for the bankers to understand the shipping cycle and take it into consideration in making loan decisions.

Low freight rate and thin profit in shipping industry can create extensive concerns in maritime policy and administration. 'Safer Shipping and Cleaner Ocean', once was a mission statement for the International Maritime Organization, resonates the wide concerns over the substandard vessels and crews, two critical factors for maritime accidents that caused the loss of lives and properties at sea, as well as marine environmental pollution. These undesirable incidents most likely follow when ship-owners have insufficient earnings to keep regular maintenances and continued trainings for the crew. To stay in business when the freight rate is low, ship operators have to reduce operation cost in vessel maintenance and manning, even replacing the qualified crew with inexperienced, low salary ones. This can multiply substandard vessels, impair maritime safety, heighten maritime casualty, and undermine sustainability in maritime shipping. According to a report prepared by SSY Consultancy and Research Ltd. for OECD Maritime Transportation Committee, low freight rate in the past 30 years is the most important factor in substandard shipping, which has caused huge economic losses.

From the perspective of national and regional public policy, perhaps the major concern is the mass layoff from shipping industry at low freight rate. When the freight revenue cannot cover its operating cost, a shipping company has to layup a ship and layoff its employees. This is particularly harmful to those developing countries supplying a large maritime work force or providing various kinds of services to the shipping industry. The massive layoffs from shipping industry facing low freight rate can significantly increase the unemployment rate in these countries. In January 9, 2008, having suffered huge losses in 2006 and very low profit in 2007, Maersk Line announced in Los Angeles Times that it plans to layoff as many as 3,000 people from 25,000 employees in its container division. On November 6, 2008, as part of the global layoff plan, Maersk A/S announced that it will cut 700 positions in the Chinese market by 2009, and shut down the global services center in Guangzhou. This province has already suffered massive layoffs recently from the shutting down of many manufacturers facing the weak exporting demand. Further layoff from the shipping company would further exacerbate the economic situation in this region.

The importance of shipping cycle in both private business operation and public sectors has, unsurprisingly, motivated numerous efforts to understand, describe, model, and predict the fluctuation of shipping freight rate. Martin Stopford (2009), for example, described the shipping cycle in the past 266 years, discussed its characteristics, frequency, and difficulties in prediction. Freight market analysis is the first area for applied econometrics. Tinbergen and Koopmans, two well-known pioneers in the econometrics, actually started their econometric analysis in shipping (Beenstock and Vergottis, 1993). Tinbergen investigated the sensitivity of freight rates to changes in demand and supply. Koopmans proposed the first theory to forecast tanker freight rates, assuming market equilibrium between demand and supply. He explained the dynamic behavior of the tanker market by investigate the interrelationship between the market size, freight rate, and shipyard's activity. Since then, many different models have been developed for the tanker and bulk carrier market analysis. Beenstock and Vergottis (1993) developed a market equilibrium model assuming explicitly profit optimization in supply side, and perfect competition on the demand. They tested the model for tanker and dry bulk shipping market using annual data. This work is recognized as a milestone in econometric analysis of shipping market that "heavily influenced" the modern analysis of bulk shipping markets (Glen, 2006). The most recent work that follows BV's model is Tvedt (2003), who combined structural and econometric stochastic methods, and built a continuous stochastic partial equilibrium model for the freight and new building market. He found that the equilibrium freight rate process is close to that of a standard geometric mean reversion process.

Despite the significant contribution of container shipping in world seaborne trade, literature on economic modeling and statistical analysis of the container shipping market is scarce. This paper fills the gap by building a dynamic-economic model for container shipping market and testing it using annual data in recent 28 years. Furthermore, without assuming individual behavior in ship investment and operation, this paper reveals the significance of collective market adjustment principles using the observed data, without involving complexities in individual behavior analysis, such as market competition strategies, speculation, and hedging.

This paper lay out as follows. It first describes the theoretic model on the container shipping freight rate and container fleet dynamics. Then, it explains the econometric process for estimating the structural model, the data, the regression results, as well as the stability test of the model. After that, the paper presents an in-sample model prediction to compare with the actual freight data in the study period, and a validity test by calculating the forecast errors for 2007 and 2008 using different estimated models. As an application, it presents the model predictions for future container shipping market between 2009-2013, under different assumptions on the increasing rate of future container shipping demand, and possible cancellation on the new orders. The purpose of this prediction is to alert the decision makers on the possible risks and short-term market trends in container shipping sector. The last section is summary and conclusion.

2. Theoretical model

As in dry bulk and tanker market, container shipping market also includes second-hand market, new-building market and scraping market. Several assumptions are made to simplify the model and focus on the freight market.

First, as container shipping industry is relatively new and the life time of a container vessel is usually more than 30 years, the scraping activity only starts recently and the size of the scraping is just a small fraction of the total fleet size. The average proportion of demolition to world container fleet capacity is only 0.593% from 1994 to 2007. Thus we can ignore the impact of scraping on container fleet capacity.

Secondly, we assume that the second hand market will not affect the container freight market. As trade in the second hand market does not change the usage of a container vessel, it does not affect the world container fleet capacity.

To further simplify the model, we assume new building market will not affect the container freight market. When a shipping company considers placing a new order, the main decision variable is the freight rate, not the new building price. Statistics show that there is a high positive correlation between new building order and freight rate, as depicted in Figure 2. Most of the new orders are made when the freight rate is high, which is often the time when the price of new building is high.

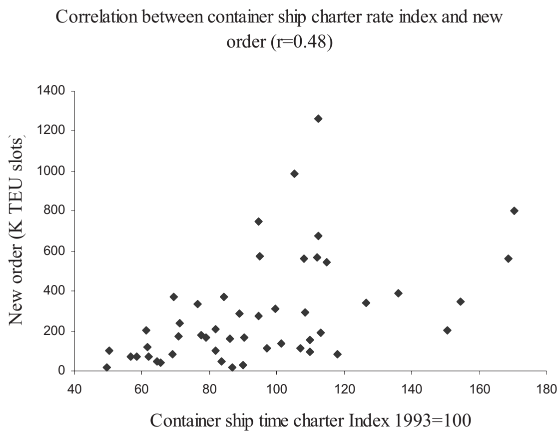
The above assumptions enable us to focus on freight market, i.e., model the fluctuation in freight rate from the interaction between demand for container shipping services, and its supply measured by the total world fleet capacity (in TEU slots). We first model the change in world container fleet capacity with the industrial profit. As high freight rate is a good indicator of high industrial profit, we first postulate that the total new order N_t at year t is proportional to the overall industrial profit of that year, i.e.,

$$N_t=\eta \cdot Profit_t \,, \tag{1}$$

where η is the average proportion of profits spent on purchasing new ships, and the $Profit_t$ follows the common definition:

$$Profit_t=P_t Y_t - c_1 X_t - c_2 OIL_t \,, \tag{2}$$

Figure 2: Correlation between container ship charter rate index and new order (1996-2008)



where Y_t is the total number of containers carried, P_t the market freight rate per TEU, X_t the world container fleet capacity (in TEU slots), OIL_t the bunker price, $c_1>0$ the constant marginal/average cost per fleet capacity (in TEU slots), and $c_2>0$ the profit adjustment factor for bunker price.

For simplicity, we use the average lag (θ) to represent the time from new order to delivery. Then the change of the world container fleet capacity can be expressed as:

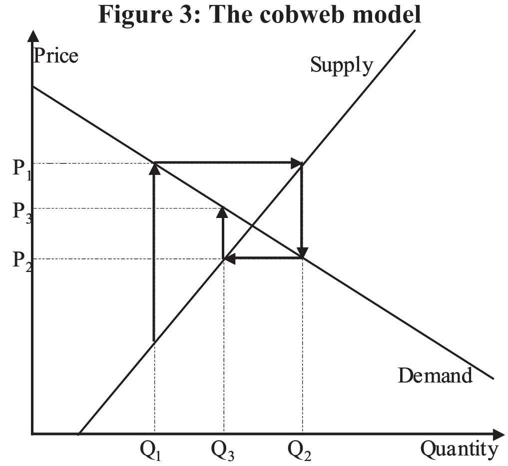
$$\dot{X}_t=N_{t-\theta} \,, \tag{3}$$

where $\dot{X}_t=X_t-X_{t-1}$. Put equation 1-3 together, the world shipping fleet dynamic equation can be specified as:

$$\dot{X}_t= \eta \cdot (P_{t-\theta} Y_{t-\theta} - c_1 X_{t-\theta} - c_2 OIL_{t-\theta}) \quad . \tag{4}$$

Next we describe how freight rate changes with the demand for container shipping, and the world fleet capacity. The change of market price due to the change in demand and supply, a fundamental economic problem, has been well studied in the literature. When there is no short-term flexibility in demand and supply, the delayed response to the excessive demand or supply can also result in price oscillation. Kaldor (1934) used the well-known Cobweb model (Figure 3) to describe the price change with alternative excessive demand and excessive supply.

Figure 3: The cobweb model



When the market price is high at time 1, the quantity demand (Q_1) at P_1 is lower than the quantity supplied (Q_2). The excessive supply (Q_2-Q_1) will reduce the price down to P_2 in the next period. At this price level, the quantity demanded (Q_2) is higher than the quantity supplied (Q_3). This excessive demand (Q_2-Q_3) will increase the price. The stability of the market price in the long run will depend on the relative price sensitivity in demand and supply. According to this theory, the change of market price can be written as:

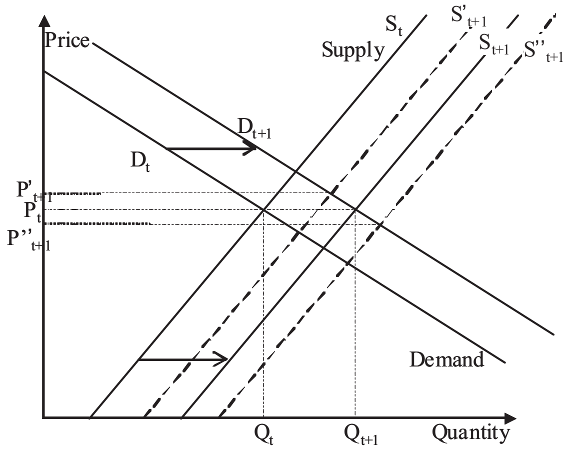
$$\dot{P}_t = \delta(Y_t - \varphi X_t), \quad (5)$$

where $\dot{P}_t = P_t - P_{t-1}$, the price change in year t, and φ the reuse rate of a TEU slot. This equation states that price will increase when there is excessive demand, and drops with excessive supply.

Consider the nature of maritime transportation for the containerized goods. First, shipping freight rate is flexible and negotiable between the shipper and carrier. Second, it is well-known that the marginal cost for additional container, especially in liner services, is very low. Container carriers can always accept one more box as long as it covers the marginal cost. Third, there are many ways to provide short-term shipping services facing sudden demand increase, including increase loading factors and increase cruise speed. Demand and supply are both flexible enough in container shipping industry, especially on annual level. This conforms to BV's model assumption in market equilibrium in his econometric analysis for dry bulk and tanker market (Beenstock and Vergottis, 1993).

Assume market clears each year, market freight rate changes with exogenous demand shift caused by the exogenous change in international trade, and the supply shift as more container vessels add to the world container fleet capacity. From the demand side, with the increase in international trade, the demand for container shipping will increase even when the market freight rate is constant. On the supply side, when more capacity is added to the industry, more container ships are available in the market to provide more services even with the same market price. An illustration of how market price changes with relative shifts in demand and supply are given in Figure 4.

Figure 4: Illustrated price dynamics with demand and supply shift



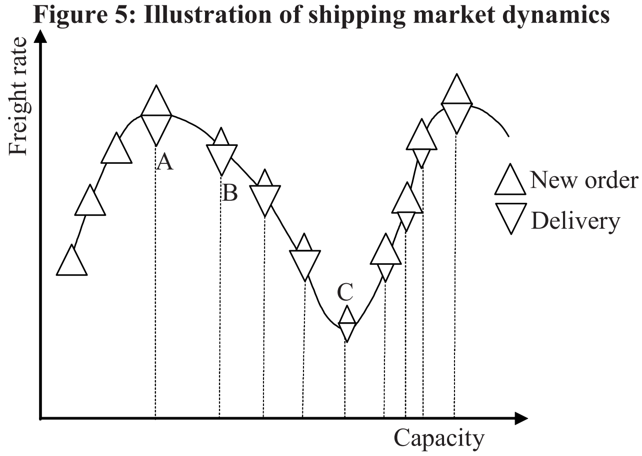
Assuming at time *t*, the market clearing price and quantity, the intersection of demand D_t and supply S_t, are (P_t, Q_t). If there are equal amount of supply and demand shifts ($D_t \rightarrow D_{t+1}$, $S_t \rightarrow S_{t+1}$), The new market clearing price will remain unchanged, while the quantity will be Q_{t+1}. This confirms to the description by Tvedt (2003). If the supply only shift to S'_{t+1}, less than the demand shift, the market clearing price will increase to P'_{t+1}. On the other hand, if the supply shift to S''_{t+1}, more than demand shift, the market clearing price will be drop to P''_{t+1}. Applying this to the container market freight rate with respect to the supply and demand change, we postulate:

$$\dot{P}_t = \delta(\Delta Y_t - \varphi \Delta X_t), \qquad (6)$$

where ΔY_t and ΔX_t are the changes in the total number of containers handled and the fleet capacity, respectively, $\varphi>0$ is a constant representing *average annual container slot reuse rate*, and $\delta>0$ is the price adjustment factor due to the demand and supply shifts.

Equation (4) and (6) are the two dynamic equations that describe the two major forces in container shipping market. The interaction of these two forces can be illustrated in Figure 5. Assume that market demand for container shipping increases exogenously. When the price is high (at A in Figure 5), the high industry profit will bring up the number of new orders (denoted by larger upper triangles). If the delivery of new containers ships resulted in a larger increase in capacity than that in demand, the market price will fall. When this happens (at B), there will be very few new orders (denoted by smaller upper triangle) by the speculators, but the ships ordered in the previous two or three years when the freight rate is increasing will keep adding to the existing fleet, which will accelerate the decreasing rate of freight rate. This downward trend in the market freight rate will end when the capacity increase slower than the demand increase (at point C where the delivery is very small). Because of the few new orders during the previous three or four years, the low supply in shipping capacity will push up the market price. When the market price is increasing, there will be a stronger incentive to order more new container vessels again, which will lead to a new shipping cycle.

Figure 5: Illustration of shipping market dynamics



To test the theory, using the annual data for container market freight rate, the total number of containers handled, the world fleet capacity in TEU slots, and bulker price from 1980 to 2006, we estimated the parameters in the statistical model using the above data, which will be explained in next section.

3. Quantitative analysis of the dynamic model

We construct the statistical model by transfer the equation (4) and (6) into linear forms as follows:

$$\Delta X_t = \eta\, P_{t-\theta} Y_{t-\theta} - \eta\, c_1 X_{t-\theta} - \eta\, c_2 OIL_{t-\theta} + \varepsilon_{1t} = \alpha_1 P_{t-\theta} Y_{t-\theta} - \alpha_2 X_{t-\theta} - \alpha_3 OIL_{t-\theta} + \varepsilon_{1t}\,, \tag{7}$$
$$\Delta P_t = \delta \Delta Y_t - \delta\varphi \Delta X_t + \varepsilon_{2t} = \alpha_4 \Delta Y_t - \alpha_5 \Delta X_t + \varepsilon_{2t}\,. \tag{8}$$

The last term ε_{it} in each equation is the error term. Although $Y_{t-\theta}$ appears in the first equation and Y_t appears in the second, they are not contemporaneously correlated. The first equation can be estimated by itself. As ΔX_t appears on the left-hand side of the first equation and the right-hand side of the second equation, the error terms are not independent ($cov(\Delta X_t, \varepsilon_{2t})=\sigma_{12}$). Thus we apply Simultaneous Equation (SE) method to estimate the coefficients in the system. The first step in SE is to rewrite equation (7) and (8) into reduced form by substituting ΔX_t into the second equation:

$$\Delta X_t = \pi_1 P_{t-\theta} Y_{t-\theta} - \pi_2 X_{t-\theta} - \pi_3 OIL_{t-\theta} + \varepsilon_{1t}$$
$$\Delta P_t = \alpha_4 \Delta Y_t - \alpha_5 \alpha_1 P_{t-\theta} Y_{t-\theta} + \alpha_5 \alpha_2 X_{t-\theta} + \alpha_5 \alpha_3 OIL_{t-\theta} + \alpha_5 \varepsilon_{1t} + \varepsilon_{2t}$$
$$= \pi_4 \Delta Y_t - \pi_5 P_{t-\theta} Y_{t-\theta} + \pi_6 X_{t-\theta} + \pi_7 OIL_{t-\theta} + \pi_8 \varepsilon_{1t} + \varepsilon_{2t}$$

As can be seen from the reduced form, the two equations are not independent. Thus, two stage least square (2SLS) is not sufficient to make full use of the correlation between error terms. Therefore, 3SLS method was applied in the estimation process for the coefficients in the reduced form. The estimated parameters are transferred back to the structural equation. The instrument variables used in 3SLS include all the exogenous variables and predetermined variables. The estimation process follows the standard treatment as specified in Judge *et al* (1988), so it will not be included in this paper.

3.1. Data

Demand for container transportation services is derived demand from global trade, which is determined by the comparative advantage of individual countries. We take demand as given to avoid modeling the global trade. Besides, as unsatisfied demands are not observable, the assumption on market clearance each year enables us to use the container throughput as quantity demanded. The data used in this study and their sources are included in

Table 1.

The world container throughput, from the Drewry Annual Container Market Review and Forecast, is the total port throughput, including the empties and transshipment. We use container throughput, not the world trade volume, as the demand for container shipping services, for the following two reasons. First, the world trade volume includes many commodities that are not carried by ships. Further, not all the seaborne trade is containerized. The containerization rate is changing. To convert world trade volume of different commodities to number of TEUs is not currently feasible. Secondly, container throughputs are a more appropriate data to use. Although there are empties, transshipments and possible double counting, they are actually part of the demand for container transportation services. Thus, we used container throughput, rather than the world trade volume, as the demand for container transportation services.

The same report from Drewry also provides the container freight rate, which is the weighted average of Transpacific, Europe-Far East and Transatlantic trades, inclusive of THCs (Terminal Handing Charge) and intermodal rates. This variable is a synthetic index, representing the average level of container freight rate. This can be an index for shipowners' unit revenue. As Drewry only reported freight rates from 1994-2006, we have to calculate the missing part (1980-1993) from General Freight Index in Shipping Statistics Yearbook 2007, using a simple statistical equation between container freight rate and the general freight index during 1994 and 2006. The container fleet capacity data are also from the Drewry Annual Container Market Review and Forecast.

On the supply side, we use the data from Clarkson Research Services Limited 2008, which include the new order, delivery, and scrap data in TEU slots and bunker prices.   Although some of these data are not used in estimate the main model, they are used in determine whether to include the scrapping market and the shipbuilding lag. Therefore, they are also included in the table.

Table 1: Data used in this study and sources

Year	Container Throughput (Yt; K TEU)*	Freight rate (Pt; $/TEU)*	Fleet Capacity (X_t; K TEU)*	Bunker Price (OILt; $/ton)%	Delivery (N_t; K TEU) %	Scrap (S_t; K TEU) %	New Order (O_t; K TEU) %
1980	38,821	*1,762*	665.0	307.0	115.8		
1981	41,900	*1,644*	702.0	288.3	38.3		
1982	43,800	*1,449*	745.0	284.8	72.6		
1983	47,600	*1,441*	799.0	243.7	100.6		
1984	54,600	*1,451*	883.0	229.6	130.3		
1985	56,170	*1,420*	1012.0	222.8	131.1		
1986	62,200	*1,355*	1189.0	142.1	140.1		
1987	68,300	*1,455*	1276.3	144.1	92.7		
1988	75,500	*1,630*	1384.7	124.7	116.4		
1989	82,100	*1,632*	1487.9	144.1	102.3		
1990	88,049	*1,544*	1613.2	191.2	133.6		
1991	95,910	*1,544*	1756.0	170.8	152.1		
1992	105,060	*1,471*	1916.3	161.9	167.6		
1993	114,920	*1,480*	2101.3	150.5	200.1		115.9
1994	129,380	1,466	2370.7	133.2	268.8	2.8	472.5
1995	144,045	1,519	2684.0	140.6	330.1	10.9	597.4
1996	156,168	1,434	3048.1	175.1	408.1	21.5	501.5
1997	175,763	1,282	3553.0	157.2	523.1	25.0	203.6
1998	190,258	1,267	4031.5	112.2	529.6	87.3	414.5
1999	210,072	1,385	4335.2	133.0	257.1	51.7	555.2
2000	236,173	1,421	4799.1	231.6	449.1	15.5	956.9
2001	248,143	1,269	5311.0	192.4	623.2	36.1	519.1
2002	277,262	1,155	5968.2	188.2	642.8	66.5	414.2
2003	316,814	1,351	6528.6	230.4	560.7	25.7	2057.0
2004	362,161	1,453	7162.8	313.4	643.0	4.0	1652.9
2005	397,895	1,491	8117.0	458.4	941.5	0.3	1644.3
2006	441,231	1,391	9472.0	524.1	1366.6	20.4	1784.6
2007	496,625	1,435	10805.0	571.3	1321.1	23.8	3060.1
2008	540,611	*1,375*	12126.0	850.7			

Freight rates (Italic part) computed from general freight index (from shipping statistics yearbook)

*Source: * The Drewry Annual Container Market Review and Forecast 2000-2008*

% Clarkson Research Services Limited 2008

3.2. Specification of θ

As a key factor in shipping market analysis, the shipbuilding lag is ubiquitous in almost all the econometric analysis in this field. Binkley and Bessler (1983) found that shipbuilding construction lag, ranging from eight months to around two years, is one of the most important market features in the bulk shipping market analysis. In our study, we assume constant construction lag during the study period. Further, as we are using annual data, we require the lag to be rounded to an integer. Therefore, we constructed 6 statistical equations between the delivery and the new order data, and selected the most significant one to use in our model.

The regress results of the 6 equations are listed in

Table 2. The 2-year lag and 3-year lag are all significant in model 1-6, but R^2 in model 5 is much bigger than that in model 6, so we choose θ=2. This means on average it takes two years to build a container vessel, although bigger ships may take longer and smaller ones may only need several months.

Table 2: Modeling of construction lags (p-value in parenthesis)

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
b_0	255.5099 (0.0436)	257.6291 (0.0174)	175.4902 (0.0061)	218.035 (0.0035)	297.841 (0.0022)	356.7358 (0.004)
$order_t$	-0.00475 (0.9475)					
$order_{t-1}$	0.167043 (0.0596)	**0.166467 (0.0327)**	0.112564 (0.0444)			
$order_{t-2}$	**0.211811 (0.0361)**	**0.210334 (0.0152)**	**0.251027 (0.0021)**	**0.313896 (0.0007)**	**0.434595 (0.0003)**	
$order_{t-3}$	**0.27146 (0.0156)**	**0.26916 (0.0037)**	**0.267692 (0.0013)**	**0.279653 (0.0028)**		**0.441905 (0.0043)**
$order_{t-4}$	-0.20405 (0.3479)	-0.212 (0.1867)				
R-squared	0.966885	0.966844	0.954026	0.914724	**0.745311**	0.615101

3.3. 3SLS results of the system parameter

The regression result from the 3SLS process for the structure equation, including the estimates of the structural parameters and their corresponding t-values, R^2 and adjust R^2 are listed below:

$\Delta X_t = 0.0000034 P_{t-2} Y_{t-2} - 0.06411 X_{t-2} - 0.438215 OIL_{t-2}$
(5.00) (-1.52) (-2.72) t-value
$R^2=0.95$, Adjusted $R^2=0.947$

$\Delta P_t = 0.00894 \Delta Y_t - 0.378085 \Delta X_t$
(3.96) (-4.01) t-value
$R^2=0.353$, Adjusted $R^2=0.328$

All the estimated coefficients (a_1 through a_5 in equation 7 and 8) are significant at least at 90% confidence level, and the coefficient estimates on revenue (a_1), bunker (a_3), demand (a_4) and supply (a_5) are all significant at 99% confidence level.

To evaluate the overall explanatory power of the whole system, we use the error sum of squares (SSE) and total variance (SST) in respective regression equation to construct the overall coefficient of determination for the system. From first equation, we have $SSE_1=200197$, $SST_1=4125572$. From the second equation, we get $SSE_2=158424$, and $SST_2=245222$. So the overall R^2 could be written as:

$$R^2 = 1 - \frac{SSE_1 + SSE_2}{SST_1 + SST_2} = 0.9179$$

which indicates the over explanatory power of the system is about 92%.

3.4. Explanation of the regression results

To understand the regulation results, we first transform them into the dynamic equations in equation (4) and (6):

$$\Delta X_t = 0.0000034 P_{t-2} Y_{t-2} - 0.06411 X_{t-2} - 0.438215 OIL_{t-2} = 0.0000034(P_{t-1} Y_{t-2} - 19080 X_{t-2} - 130421 OIL_{t-2}) \quad (9)$$

$$\Delta P_t = 0.00894 \Delta Y_t - 0.378085 \Delta X_t = 0.00894(\Delta Y_t - 42.27 \Delta X_t) \quad (10)$$

Thus the coefficients in the equation (4) and (6) can be obtained from equations (9) and (10), which are listed in Table 3.

Table 3 Summary of the parameters

Parameters	estimates	Meaning
η	0.0000034	Propensity for new order per dollar industrial profit (in TEU)
c_1	19080	Average annual cost to operate one TEU slot (in US$)
c_2	130421	Cost adjustment per unit increase in bunker price (in K US$)
δ	0.00894	Price adjustment factor for Demand-Supply change ($/K TEUs)
φ	42.27	Annual productivity per TEU slot

The economic meanings of the estimated parameters are explained here. Parameter η is the propensity to order new ships or the increase rate of container capacity per one dollar increase in industrial profit. The estimation indicates that 34 TEU slots will be added to the capacity per 10-million US dollar profit in the industry. Considering that the cost for 3,500TEU container vessel is about US$63 million in 2007 (Clarkson, Shipping Review Database, 2009). Our result shows that to order a container ship with that size, the total industrial profit have to be $1,029 million. This implies around 6.2% of the earnings are used for building new ships. c_1 is the annual average cost to own and operate one TEU slot. It is the total cost paid by the shipping company in the transportation process, as long as that process is covered by the freight rate. c_2 is the gross cost adjustment factor per dollar increase in bunker price for the whole industry, in thousand dollars. It indicates for one dollar increase in bunker price, the operation cost for the whole industry will increase about US $130 million. δ represents the price sensitivity for relative annual increment in demand and supply. The result shows that if demand shifts one hundred thousand more in TEU slots than the shift in fleet capacity, there will be 89 cents increase in freight rate. The higher is the δ, the more sensitive is the freight rate to the relative magnitude in demand and capacity change. φ is the capacity utilization factor, representing annual reusing rate per unit TEU slot.

To test the stability of the model, we conducted another two regression analyses using the same model applied to data from different time period. The regression period is from 1980 to 2006 and 1980 to 2007, respectively. The comparisons of the two additional regression results together with the regression on all observations (1980 to 2008) are shown in
Table 4.

Table 4: Comparison of parameter estimates using different observation range

	80-06		80-07		80-08	
	Parameter value	t-value	Parameter value	t-value	Parameter value	t-value
η	0.00000312	4.41	0.00000362	5.10	0.00000336	4.97
c_1	-17003	-1.21	-22683	-1.85	-19080	-1.53
c_2	-141064	-2.71	-134830	-2.96	-130421	-2.72
δ	0.009386	3.88	0.009093	3.78	0.008944	3.96
φ	-45	-3.93	-43	-3.74	-42	-4.01

Table 4 indicates the parameter estimates are stable, and the t-values for all the parameters are all significant. Hence, first, it is predicable the estimated parameters in the model will not change much when there are more data observations in the future. Secondly, the stability in the parameter estimation and the t-value indicate the stability of the whole model and is free from the autocorrelation and heteroskedasticity. Therefore, as we will show next, the prediction using the estimated model can also be assured reliable.

4. Prediction of the estimated model

When market demand is exogenous, container shipping fleet capacity and the market freight rate are the two most important variables in the container shipping market analysis. The specification of the model enables us to predict the fleet capacity increases in two years based on the current container throughput, freight rate, and bunker price. The relative capacity increase, determined endogenously from the first dynamic equation, can then be used to predict the adjustment in freight rate, for given container transportation demand.

4.1. In-sample prediction

To demonstrate the explanatory power of our model, we first compare the model prediction with the actual date. An in-sample prediction for the market fleet capacity and the freight rate, together with the actual freight rate and fleet capacity from 1980 to 2008 (called 80-08 model), are provided in Figure 6.

Figure 6: In-sample prediction of fleet capacity and freight rate (1980-2008)

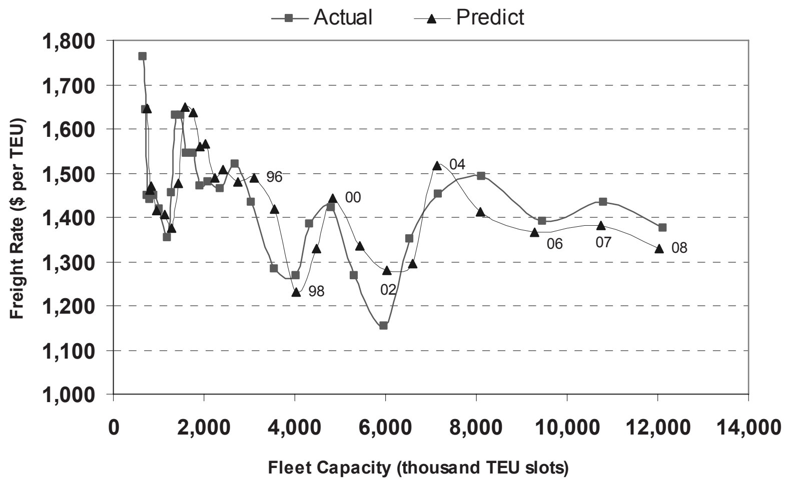


Figure 6 exhibits that the fleet capacity increases faster in recent years than that of the earlier years, as the horizontal distances between each pair of dots are wider in recent years. The freight rate is generally decreasing over time, and it is oscillating around the US$1,400 in recent years. Our prediction can largely replicate the trend of the actual freight rate change. The predicted fleet capacity each year is very close to the real fleet capacity (They are basically in the same vertical line). The predicted fleet capacity and the freight rate for 2008 (12,030 thousand TEU slots, US$1,329 per TEU) are very close to the real capacity and freight rate (12,126 thousand TEU slots, US$1,375 per TEU).

To check the stability of model prediction, we estimated the model using the observation from 1980 to 2006 to predicate the freight rate and fleet capacity for 2007 and 2008. The actual value of fleet capacity and freight rate, and the predicated ones with 80-06 model and 80-08 model are shown in Table 5.

Table 5: Comparison of predicted values using different model results

Actuall value			Predicted value			
			80-06 model		80-08 model	
Year	Fleet Capacity	Freight Rate	Fleet Capacity (error)	Freight Rate (error)	Fleet Capacity (error)	Freight Rate (error)
2007	10805	$1,435	10691	$1,399	10744	$1,382
			1.06%	2.52%	0.56%	3.66%
2008	12126	$1,375	11873	$1,318	12030	$1,329
			2.09%	4.14%	0.79%	3.36%

Note: Fleet Capacity in thousand TEUs

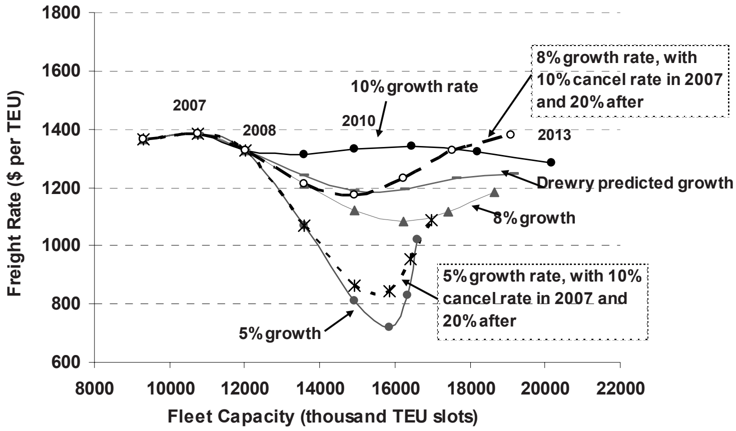
Comparing the predicted results from two different models, it is obvious that both models can predicate fleet capacity in relatively smaller error than the prediction of freight rate. The error margin for freight rate prediction ranges from 2.52% to 4.14%, while prediction for fleet capacity is within 2.10%. In either model, the prediction values are within 5% error margin, which indicate a good fix, even for the out of sample prediction. Therefore, in out-sample prediction for future container market, the 80-08 model prediction result will be presented.

4.2. Prediction for the future container shipping market.

The purpose for this dynamic-economic model is to predict the future market situation, so that the decision makers could anticipate and respond possible market changes. The first necessary step is to assume the future container demand growth rate based on the past information. Our data shows that average increasing rate of the container throughput in the past 27 years from 1981 to 2007 is about 9.94%, with highest 14.31% in 2004, and lowest 2.88% in 1985. Considering the possible range of container transportation demand in the coming years, we assume three different growth rates (5%, 8% and 10%) for the year from 2009 to 2013.

Under the current global financial crisis, the shipping sector not only refrained from ordering new ships, but also motivated to cancel existing orders. According to recent statistics from Lloyd's register, total new orders in October 2008 have been dropped by 90% comparing with the same period in 2007. According to Clarkson, there are totally 94 new order cancellations. Cancellations can reduce the number of new deliveries to the market and slow down the freight rate decrease. As recent new orders are easier to cancel, we assume 10% cancellation rate for the new orders made in 2007, and 20% cancellation rate for the new orders made in following years. The continued cancellation after 2008 reflects the change in industry behavior for making new orders, a more prudent measure facing the financial crisis. The prediction of the market freight rate and the fleet capacity from year 2009 to 2013 are shown in
Figure 7. Actual data for 2007 and 2008 are included in the figure just to show the trend.

Figure 7: Forecast of future container shipping market from 2007 to 2013



Drewry predicted growth rate (8.6%, 8.7%, 9.1%, 8.9%, 8.7% from 2009 to 2013 respectively)

Figure 7 also includes the market prediction based on Drewry's forecast on possible growth rate of the future container transportation demand. According to that, our model shows that freight rate will continue decreasing until 2010, then increase slowly until 2013. This reveals the excessive capacity in the world container fleet. Our prediction base on the 10% growth rate is more optimistic than the Drewry, reflecting the best situation for quick recovery. However, high freight rate can encourage larger new orders, causing an earlier decreasing market in 2011. Our prediction using 8% growth rate in future container throughput represents a more conservative prediction than the Drewry's. The freight rate will be below US$1,200 from 2010 to 2013, and will recover after 2011. In this case, the new order activity will decrease, as the net profit will decrease in the industry. Considering the current financial crisis and the low demand in the container shipping, if the future demand increasing rate was only 5%, then the freight rate will be below US$1,100 in 2009, close to US$800 in 2010, and further below US$800 in 2011.

Cancellations of new orders could slow down the dropping of the freight rate. For 8% future growth rate, if 10% of the new orders made in 2007 and 20% of new orders made afterwards were canceled, the freight rate would stop decreasing in 2010 when it was slightly lower than US$1,200, and it could better than predicated rate based on the Drewry's prediction later. For 5% growth rate, under the same cancellation scheme, the predicted freight rates were higher than that in no cancellation case, and would never below US$800. This shows that cancellations are beneficial to prevent further drop in freight rate, if the current financial crisis has a serious negative impact on the world economy and international trade. Because of the comparatively higher freight rate, when there is cancelation, the new order would be higher than the case when there is no cancellation; therefore, the capacity with cancellation is larger than the capacity without cancellation.

5. Summary and Conclusion

We presented a dynamic-economic model for container shipping market characterized by the container shipping freight rate and the global container fleet capacity. The model postulates the changing of equilibrium freight rate under demand and supply shifts in the container shipping market. The world container fleet capacity is augmented by the number of new orders which is proportional to the industrial profit earned two years ago. The quantity demanded for container transportation services, as a derived demand from international trade, is assumed to be exogenously determined in the model.

The model parameters were estimated using the global container shipping market data from 1980 to 2008, based on the available data from Drewry and Clarkson. Considering the interdependency of the two dynamic equations, a three-stage least square method was adopted in the regression analysis. The estimated results are quite stable, provided a high goodness of fit, and the parameter estimates are significant above 90% confidence level. The overall model can explain more than 90% of the variations in fleet capacity and freight rate, and the in-sample prediction of the model can largely replicate the actual data within the research period. The errors of in-sample prediction and out-sample prediction for the previous two years, using model estimated using date with different number of observations, reveals that the stability of prediction is within 5% error margin.

As an application of the research, we predicted the future container market from 2009 to 2013, based on different assumptions on the future growth rate in container transportation demand. The result shows that if the world financial crisis were continue to decrease the international trade, the container freight rate could drop to below US$800 in 2011. With decreasing rate of new ordering and the cancellation of existing order, the market freight rate could be saved from reaching such a low level, although the one who cancelled the new order would suffer some immediate losses.

In conclusion, the model can provide information for decision makers in both public policy and private business. The maritime agencies or organization at regional, national and international level can use this information to stabilize the market freight rate, so as to mitigate the negative impact of the recent financial crisis on maritime industry, marine environment, maritime safety, and national, regional and local economy. The bankers can use this information in ship financing decision, to minimize the possible risks caused by the low freight rate. Shipowners and ship operators can also use this method to setup their respective best strategies to prevent or reduce possible losses in the coming several years.

6. Acknowledgements

The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. B-Q08K).

References

Beenstock, M. and Vergottis, A. (1989), An econometric model of the world market for dry cargo freight and shipping. Applied Economics, 21, 339-356.
Beenstock, M. and Vergottis, A. (1993), Econometric modeling of world shipping. London.

Binkley, J.K. and Bessler, D. (1983), Expectations in bulk ocean shipping: an application of autoregressive modeling. Review of Economics & Statistics, 65(3), 516-520.

Glen, D.R. (2006), The modeling of dry bulk and tanker markets: a survey. Maritime Policy and Management, 33(5), 431-445.

Goulielmos, A. M. and Psifia, M. (2006), Shipping finance: time to follow a new track?. Maritime Policy and Management, 33(3), 301-320.

Judge, G.G., Hill, R.C., Griffiths, W.E., Lutkepohl, H., and Lee, T.C. (1988), Introduction to the theory and practice of econometrics 2nd ed. John Wiley & Sons.

Kaldor, Nicholas (1934), A Classificatory Note on the Determination of Equilibrium. Review of Economic Studies, vol I (February, 1934), 122-36.

SSY Consultancy and Research Ltd (2001), The cost to users of substandard shipping. Prepared for the OECD maritime transport committee.

Stopford, M. (2009), Maritime economics, 3rd Edition, Routledge, London.

Tvedt, J. (2003), Shipping market models and the specification of freight rate processes. Maritime Economics & Logistics, 5, 327-346.

Volk, B. (1984), Shipping Investment in Recession (Bremen: Institute of Shipping Economics and logistics at Bremen)

Port competition using capacity expansion and pricing

Meifeng Luo[1]*, Liming Liu*[1] *and Fei Gao*[2]

[1] Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hong Kong.
[2] The Chinese University of Hong Kong, Hong Kong.

Abstract

This study models two ports, serving the same hinterland, competing strategically using both pricing and capacity investment. Both ports are profit maximizers, and port expansions are lumpy, indivisible and irreversible. The decision making process of the two ports is analyzed using a two-stage game. In the first stage, two ports compete with each other on capacity expansion. In the second stage, they follow Bertrand competition with differentiated products conditional on realized port capacities. Within this formulation, we show the existence of unique Nash equilibrium in the pricing sub game, and the change of the equilibrium price with operation cost, market demand determinants and capacity sizes using comparative statics. In capacity investment game, we identified the pure strategy Nash equilibriums for different scenarios characterized by the incremental benefit of expansion and the annual capital cost of investment. Through both analytical study and numerical simulation, we show that the capacity expansion at any port will decrease the equilibrium prices at both ports, thus beneficial to the port users. Smaller port with more elastic demand and lower operation and investment cost is more likely to expand in an increasing market. Capacity expansion may result in lower total profit of the two ports, which is analogous to a Prisoner's Dilemma.

Keywords: Lumpy Capacity Investment, Pricing, Port Competition, Prisoners' Dilemma

1. Introduction

Container ports serving the common hinterland compete actively for global carriers. Due to the importance of container port activity to the local economy and the huge capital cost involved in port development and operation, the outcome of such competition not only has significant impact on the related private business, but also on the public sectors. Therefore, research on how ports can maintain their competitive edge and how this may affect the public and private sectors is of great importance.

Hong Kong is one of the leading container ports in the world, due to its unique position linking the fast economic development of mainland China to the outside world. However, in the past decade, its market share has been challenged by the container port development in Shenzhen, which not only generated huge capacity in a short period, but also enjoys many advantages, including lower operating costs and proximity to the hinterland. This can clearly be seen from figure 1, the container port throughputs from 1991 to 2008. It demonstrates that the throughput of Shenzhen increased from merely 51,000 TEUs to 21.4 million TEUs. This corresponds to an increase in market share from less than 1% to 47% for Shenzhen port, while that of Hong Kong decreased from 99% to 53%.

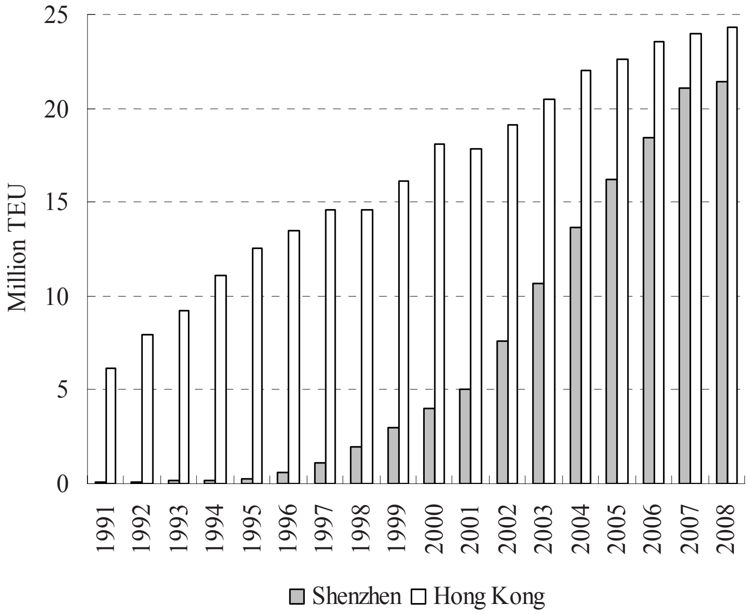


Figure 1: Container port throughputs of Hong Kong and Shenzhen port from 1991 to 2008
Source: Hong Kong Port Development Council, Shenzhen port information center, and UNCTAD

To maintain its competitive edge, Hong Kong Port can take many different strategies, which can be generally divided into two categories, short-term measures and long-term ones. *Short-term measures* include reducing price, increasing service quality, expediting import/export documentation processes, and other strategies that can reduce the user cost incurred when using the port. The *long-term strategies* include capacity expansion, which could improve container loading/unloading efficiency and reduce the vessel turn-around-time. While these measures all contribute to the attractiveness of a port, the competitative outcome is hard to predict when the two ports are competing strategically.

The purpose of this paper is to analyze the possible outcomes of port competition, taking port of Hong Kong and Shenzhen as an example. It models two ports, serving the same hinterland, competing strategically using both pricing and capacity investment. Both ports are profit maximizers, and port expansions are in lumpy, indivisible and irreversible. We use a two-stage model to analyze the decision making process. In the first stage, two ports compete with each other on capacity expansion. In the second stage, they follow Bertrand competition with differentiated products conditional on realized port capacities.

A literature review in capacity expansion and pricing reveals that there are many different research areas on this topic, from specific study in port, to general economics, game theory analysis, and operations research.

- In port study, most of the previous research on optimal port capacity and pricing is for single ports. For example, Devanney and Tan (1975) used dynamic programming to analyze optimal pricing and timing for capacity expansion in a monopoly port. Allahviranloo and Afandizadeh (2008) studied optimal investment on port development through minimizing the net present value of the total transportation cost, facility cost, dredging cost, operation cost, and benefit from the foreign shipping line at the national level. Literature on strategic port capacity investment and pricing is scarce.

- As to the general economics, Chenery (1952) studied the optimal capacity investment for exogenous demand increase over time, with economies of scale in plant size. Manne (1961, 1967) and Bean et al. (1992) analyzed capacity expansion with probabilistic growth, location and time. Abel and Eberly (1996) investigated capacity investment with reversibility issue. Starrett (1978) discussed optimal timing and size of the firm with depreciable capacity and increasing demand, from the perspective of welfare maximization. More recently, Demichelis and Tarola (2006) studied capacity expansion and dynamic monopoly pricing.

- With respect to game theory, Gilbert and Harris (1984) studied the competition between Nash competitors in indivisible and irreversible capacity investment. In their model, output is set equal to the capacity. Therefore, price and marginal cost are not an issue. Besanko and Doraszelski (2004) used dynamic programming method to study the capacity expansion in competitive market, and concluded that the capacity reversibility is a key determinant in firm size distribution in industry. Tabuchi (1994) developed a Hotelling model of spatial duopoly on two-dimensional space using two stage games. Firms select location in the first stage, and compete with each other using price at different location in the second stage. Some other relevant literature includes, but is not limited to, Gilbert and Lieberman (1987), Hay and Liu (1998) and Aguerrevere (2003).

- In operations research, Anupindi and Jiang (2008) discussed a duopoly model for production decision with capacity investment under demand uncertainty, competing in both price and quantity. Hall and Porteus (2000) developed a dynamic model in which firms compete by investing in capacity that affects the customer service level and consequently, the market share of each firm. Price is exogenous, and it has no effect on customer's preference. Liu et al. (2007) extend their work by incorporating a general demand form and further extend the game competition model to an infinite-horizon setting. Acemoglu et al. (2006) studied the capacity investment for service providers of a large-scale communication network and price competition, which is similar to our problem. However, in their research, cost of investment is continuous and proportional to the magnitude of the capacity. While in our problem, we consider lumpy capacity investment for port competition, i.e., each port will decide whether to invest or not, instead of the magnitude of the capacity expansion.

Compared with the existing literature, this paper has two unique properties. *First*, it considers two measures – capacity expansion and pricing – in an integrated framework of port competition. This is increasingly important because, facing an increasing market demand, port with large capacity can enjoy cost advantages from economy of scale, being more competitive in attracting global carriers and more likely to be successful in future competition. *Secondly*, it is the economics in port operation, rather than the port capacity constrain, that creates the needs for port expansion. Capacity is not binding. When port demand is higher than its capacity, a port can handle this with some extra cost, which implies a congestion cost. Therefore, it is the reduced congestion cost and possible revenue from the lower charge that made the expansion beneficial.

Next section presents model basics. Section 3 investigates the competition game between the two ports, with the strategic pricing in Section 3.1 and port expansion decision in Section 3.2. Section 4 provides a numerical analysis of competition between Hong Kong and Shenzhen container ports. A summary of the paper and the findings is in section 5.

2. Model Basics

This section presents the basic framework for the two-stage pricing – capacity expansion game. First, we adopt the Bertrand price competition with differentiated products (Baye and Kovenock, 2009). The base demand ($\bar{x}$) for cargo import and export services to and from the hinterland is assumed inelastic to the port price. The two ports serving this area have different initial market share α_k, where the index k (k=i or j) indicates individual port, and α_i+α_j=1. The two ports are perfect substitutes, and demand for one port decreases with its own price β_k, and increases with the price at the other port β_l,(l=i, j; l≠k) i.e.,

$$x_i(p_i, p_j) = \alpha_i \bar{x} - \beta_i p_i + \beta_j p_j, \quad \text{and} \quad (1)$$

$$x_j(p_i, p_j) = \alpha_j \bar{x} - \beta_j p_j + \beta_i p_i \quad (2)$$

Secondly, we assume each port has its own operational cost function $V_k(x_k, C_k)$, with a positive marginal cost that increases with the throughput and decreases with capacity. This property enables the analysis for the benefits in capacity expansion - the reduced marginal operation cost. To simplify the mathematic derivation, we used a more specific functional form, $V_k(x_k, C_k)=f(C_k)x^2$, assuming $f_k(C)>0$ and $f_k'(C)<0$. The capacity of the port is not binding, reflecting in practice that a port can always handle more than its designed capacity with some additional cost. For numerical examples, we used $f_k(C_k)=\theta_k/C_k$, where θ_k is a positive constant. Larger θ_k means larger average marginal operational costs for a given capacity size.

Finally, we describe the two-stage game for strategic pricing and capacity expansion based on the real world decision-making process, as depicted in Figure 2. At the beginning of stage 1, each port decides whether to expand its capacity, knowing the capacity expansion behavior of the competitor and anticipating the pricing strategy of the two ports after this period. Each port can only add a fixed capacity at a time, and the incremental capacity is the same in two ports. Capacity, once added, is not removable. Then in stage 2, having observed the realized capacities at two ports, each port sets a price to maximize its profit within that period, conditional on the pricing strategy of the competitor.

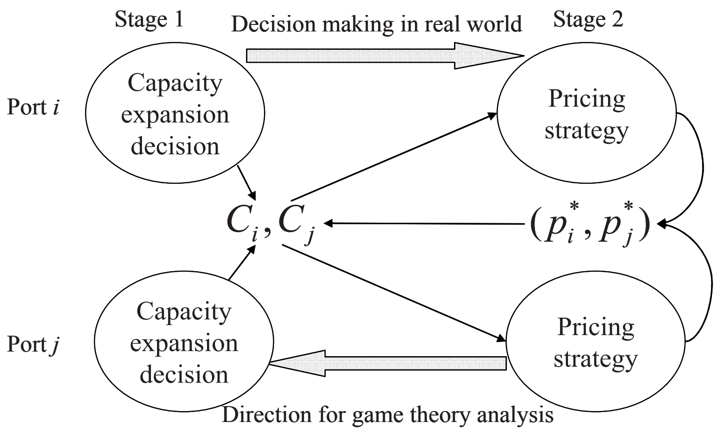


Figure 2: Illustration of the two-stage game

To analyze this decision process, we start from stage 2, where each port set its best price in response to the price of the competitor, and the available capacities at both ports. Then in stage 1, each port determines its best strategy in capacity development. At this stage, if a port chooses to develop, the new capacity will be C_k^1; otherwise, it will be C_k^0. Therefore, in the next section, we first present price competition in stage 2, followed by the capacity investment game in stage 1.

3. Capacity Expansion and Pricing Game

This section presents the game theory analysis for strategic decision making in capacity expansion and pricing. Following backward induction, we show first the pricing subgame in stage two, then the capacity expansion game in stage one. In price subgame, to begin with, we show the Nash equilibrium prices of the two ports for existing capacities. Following that, we present the static analysis of the equilibrium profit and price change with important parameters. In capacity expansion game, we analyze the pure strategy Nash equilibriums for capacity expansion.

3.1. Pricing Subgame

As introduced in the previous section, we adopt the Bertrand competition with differentiated products to model the strategic pricing behavior in the second stage. Under this specification, there exists a unique Bertrand equilibrium (Cheng, 1985), which states that port equilibrium price always exceeds its marginal cost, and two ports can have different prices and positive profits. We first characterize the mutual best response function of each port and show the equilibrium price. To link the pricing strategy in this stage with the capacity expansion outcomes in the first stage and other demand parameters, we apply comparative static analysis on equilibrium price and profit with respect to the concerned parameters. Assuming each port chooses the best price to maximize its profit based on the existing port capacity, i.e.,

$$\max_{p_k} \Pi_k(p_i, p_j, C_k) = p_k x_k(p_i, p_j) - f_k(C_k) x_k^2(p_i, p_j) \tag{3}$$

(the second order condition $\frac{\partial^2 \Pi_k(p_i, p_j, C_k)}{\partial p_k^2} = -2\beta_k - 2\beta_k^2 f(C) < 0$), then the price satisfying the first order condition

$$p_k - \frac{x_k(p_i, p_j)}{\beta_k} = 2 f_k(C_k) x_k(p_i, p_j) \tag{4}$$

maximizes the profit for port k. This is also the best response to the price of the other port. This first order condition can also be expressed in the elasticity term, $-\varepsilon_k^* = 1 + 2f_k(C_k)\beta_k$, where $\varepsilon_k^* < 0$ is the demand elasticity of port k at the optimal point. The right hand side of (4) is the marginal cost of port k, which is increasing in x_k due to congestion.

For each port, the mutual best response function can be obtained by substituting the corresponding demand function into its first order condition, and solving its own price p_k in terms of the price of the other port, p_l :

$$p_k(p_l) = \frac{[1 + 2\beta_k f_k(C_k)]\beta_l p_l + \alpha_k \bar{x}[1 + 2\beta_k f_k(C_k)]}{\beta_k [2 + 2\beta_k f_k(C_k)]}. \tag{5}$$

Using the elasticity term, (5) can also be written as $p_k(p_l) = \frac{\varepsilon_k^*}{\beta_k(\varepsilon_k^* - 1)}(\beta_l p_l + \alpha_k \bar{x})$. Replacing $\frac{\varepsilon_k^*}{\beta_k(\varepsilon_k^* - 1)}$ with φ_k^*, it can be further simplified as

$$p_k(p_l) = \varphi_k^*(\beta_l p_l + \alpha_k \bar{x}) \tag{6}$$

This can be seen in figure 3, which illustrates the best response pricing strategy for each port. According to Cheng (1985), there will be a Nash equilibrium price pair (p_i^*, p_j^*) at the intersection of these two best response curves for the given capacity level of each port.

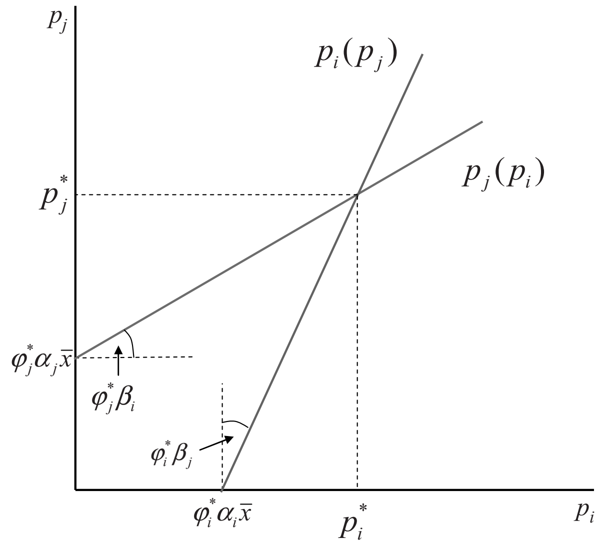


Figure 3: Mutual optimal price response functions for the two ports

3.1.1. Comparative Statics Analysis for Equilibrium Price and Profit

The purpose of comparative statics analysis is to exam the changes in equilibrium price, throughput, and profit with respect to the changes in the important parameters, such as the capacity, demand, and price sensitivity. The results from this analysis can be summarized as follows: Here we give a summary of the results.

(1) The equilibrium price (throughput) of one port increases with all the parameters that increase the demand of the port ($\bar{x}$, α_k, β_l), decreases with the parameters that decreases its demand (α_l, β_k).
(2) Capacity expansion in either port will decrease the equilibrium price. It will increase its own throughput, and decrease the throughput of the competitor.
(3) Port expansion decreases competitor's profit.
(4) Port expansion can have positive gain if $1 + 2f_k(C_k)\beta_k - 2f_l(C_l)\beta_l > 0$. Since the change of marginal cost *w.r.t.* own price is $\frac{\partial MC_k}{\partial p_k} = -2f_k(C_k)\beta_k$, the above condition can be written as $\frac{\partial MC_k}{\partial p_k} - \frac{\partial MC_l}{\partial p_l} < 1$. That is, if the difference in the change of marginal cost with respect to price change between port *k* and *l* is less than one, port *k* can increase its profit through expansion.

The Appendix contains the detailed mathematical derivations for the change of equilibrium price p_k^* and quantity x_k^* w.r.t. the change in demand parameters and capacity, as well as the change in equilibrium profit w.r.t. capacity change.

To summarize, capacity expansion can increase its throughput at the expense of the other port and decrease prices at both ports. This can reduce the profit at the non-expansion port. However, the profit at the expansion port may decrease or increase, depending on the relative sensitivity of marginal cost *w.r.t.* prices at two ports. For one unit decrease in price from capacity expansion, if the increase in marginal cost between the expansion port and the other port is less than one, port expansion will increase its profit.

The properties derived in this section have practical implications for optimal pricing strategy under competition. First, both ports could charge a higher price to cover the congestion cost when the market is good. Second, for larger port with higher existing market share, the optimal price could be higher. If demand is sensitive to this price, it is not optimal to charge higher price, because the user will shift to the other port.

Although capacity expansion can reduce congestion, it may not necessarily increase its profit. However, capacity expansion of one port will definitely reduce the profit of the other. Therefore, to make port expansion decision, it is necessary to consider the behavior in capacity expansion, in addition to the pricing of the competitor.

3.2. *Capacity Expansion Game*

This section explores the strategic investment behavior of the competing ports. In this game, each port decides whether to invest in capacity expansion, knowing that the other port is making the same decision. Denoting the equilibrium price, throughput, profit and annual capital cost in stage 2 for port k as $p_k^*(C_i, C_j)$, $x_k^*(C_i, C_j)$, $\Pi_k^*(C_i, C_j)$, and I_k, where C_i, C_j are the capacities for port i and port j respectively, for port i, facing the capacity of the other port C_j, its decision problem is:

$$\max_{C_i \in \{C_i^0, C_i^1\}} \Pi_i^*(C_i, C_j) - I_i \cdot 1_{C_i = C_i^1} \tag{7}$$
$$= \max_{C_i \in \{C_i^0, C_i^1\}} p_i^*(C_i, C_j) x_i^*(C_i, C_j) - V_i(x_i^*(C_i, C_j), C_i) - I_i \cdot 1_{C_i = C_i^1}.$$

If $I_i < \Pi_i^*(C_i^1, C_j) - \Pi_i^*(C_i^0, C_j)$, i.e., for the given capacity at port j, the gain from capacity expansion can offset the capital cost in that period, then the port should expand. Let

$$\Delta\Pi_i^*(C_j) = \Pi_i^*(C_i^1, C_j) - \Pi_i^*(C_i^0, C_j) \tag{8}$$
$$\Delta\Pi_j^*(C_i) = \Pi_j^*(C_i, C_j^1) - \Pi_j^*(C_i, C_j^0) \tag{9}$$

(8) and (9) are the gain in port expansion for given capacity of the competitor. Ports decide whether to expand by comparing these with the annual capital cost. The decisions are analyzed using a normal form game, such as the one in Table 1, which contains the corresponding net profits for the two ports. The letter *Y* and *N* stand for two possible development strategies – *Y* for expand, *N* for not expand. To see the decision-making process, we first investigate a special simpler case where the two ports are identical.

Table 1: Net Profits at Equilibrium price for Different Capacity Investment Decisions

	j:N	*j:Y*
i:N	$[\Pi_i^*(C_i^0, C_j^0); \Pi_j^*(C_i^0, C_j^0)]$	$[\Pi_i^*(C_i^0, C_j^1); \Pi_j^*(C_i^0, C_j^1) - I_j]$
i:Y	$[\Pi_i^*(C_i^1, C_j^0) - I_i; \Pi_j^*(C_i^1, C_j^0)]$	$[\Pi_i^*(C_i^1, C_j^1) - I_i; \Pi_j^*(C_i^1, C_j^1) - I_j]$

3.2.1. *Identical Competitors*

Here we assume the two ports have same demand and cost functions, and the same initial capacity before and after expansion, i.e., $C_i^0 = C_j^0 := C^0$ and $C_i^1 = C_j^1 := C^1$, and the costs are the same, i.e., $I_i = I_j := I$. Since the two ports are identical, we have $\Pi_i^*(C, C) = \Pi_j^*(C, C)$ ($C = C^0$ or C^1) and $\Pi_k^*(C^1, C^0) = \Pi_l^*(C^0, C^1)$. Using C^0 for other port not expand and C^1 for other expand, we have $\Delta\Pi_i^*(C) = \Delta\Pi_j^*(C) := \Delta\Pi^*(C)$, ($C \in \{C^0, C^1\}$). Under these assumptions, the investment decision rules are as follows:

(1) $I < \min(\Delta\Pi^*(C^0), \Delta\Pi^*(C^1))$: The annual capital cost is less than the gain, regardless of expansion decision of the other ports. In this case, both ports will choose to expand and (*Y, Y*) is the unique Nash equilibrium. For example, if both ports have serious congestion problems facing a rapid

increase in demand, they will both expand, irrespective of the decision of the other.

(2) $\Delta\Pi^*(C^1)<I<\Delta\Pi^*(C^0)$: The annual capital cost is less than the gain if the other port does not expand, and more than the gain if it does. For each port, if the other port expands, it will not expand since $I>\Delta\Pi^*(C^1)$. If the other port does not expand, it will expand because $I<\Delta\Pi^*(C^0)$. Therefore, it is optimal for the port to choose the opposite strategy as the competitor. (*N*, *Y*) and (*Y, N*) are two Nash equilibriums. This is similar to the situation where the demand is just enough for one port to develop. If both develop, they will end up over capacity, which is bad for both ports.

(3) $\Delta\Pi^*(C^0)<I<\Delta\Pi^*(C^1)$: The annual capital cost is more than the gain if the other port does not expand and less than the gain if it does. This happens when the expansion of one port exerts a detrimental impact on the other. It is better for the other to follow the strategy, to counteract the impact of the expanding port. Hence, one port better expands, if the other port expands, since $I<\Delta\Pi^*(C^1)$; if the other port does not expand, it will not expand since $I>\Delta\Pi^*(C^0)$. Thus (*N*, *N*) and (*Y*, *Y*) are two Nash equilibriums.

(4) $I>\max(\Delta\Pi^*(C^0), \Delta\Pi^*(C^1))$: The annual capital cost for expansion is larger than the gain regardless of the number of expanding ports. In this case, both ports will not expand. (*N*, *N*) is the unique equilibrium of the investment decision for the whole competition game.

3.2.2. General Competitors

Having explored the decision rules for the identical competitors, this section begins to investigate the capacity investment game where ports have different operational cost and market demand functions. First, for each port, we define a symbol for possible gains from expansion, with possible strategies of the other port.

Let :

$L_i=\Delta\Pi_i^*(C_j^0)$: The gain for port *i* when port *j* does not expand.

$M_i=\Delta\Pi_i^*(C_j^1)$: The gain for port *i* when port *j* expands.

$L_j=\Delta\Pi_j^*(C_i^0)$: The gain for port *j* when port *i* does not expand.

$M_j=\Delta\Pi_j^*(C_i^1)$: The gain for port *j* when port *i* expands.

Following the discussions for identical ports, for each port, we have the following scenarios and their corresponding decision rules:

- $I_k<\min(L_k, M_k)$: it is optimal to expand;
- $M_k<I_k<L_k$: it is optimal to make a different decision from its competitor;
- $M_k>I_k>L_k$: it is optimal to make the same decision as its competitor;
- $I_k>\max(L_k, M_k)$: it is optimal not to expand.

When the competitors are not identical, $L_i\neq L_j$, $M_i\neq M_j$, there will be 16 scenario combinations, as for each decision rule in one port, there will be four responses from the other port. For each combination, two ports determine their respective best expansion strategy, using the normal form game presented in Table 1. Table 2 lists all the possible equilibrium strategies for each of the 16 combinations. Unlike the pricing subgame that has a unique equilibrium, the capacity investment game may have multiple equilibriums or no equilibrium in some scenarios. When there is no equilibrium, we cannot predict the strategy for the player with certainty. We analyze such scenarios in the following.

Table 2: Nash Equilibrium of the Capacity Investment Game

	$I_j<\min(L_j, M_j)$	$M_j<I_j<L_j$	$M_j>I_j>L_j$	$I_j>\max(L_j, M_j)$
$I_i<\min(L_i, M_i)$	(*Y, Y*)	(*Y, N*)	(*Y, Y*)	(*Y, N*)
$M_i<I_i<L_i$	(*N, Y*)	(*Y, N*)(*N, Y*)	No Equilibrium	(*Y, N*)
$M_i>I_i>L_i$	(*Y, Y*)	No Equilibrium	(*Y, Y*)(*N, N*)	(*N, N*)
$I_i>\max(L_i,M_i)$	(*N, Y*)	(*N, Y*)	(*N, N*)	(*N, N*)

When $M_k>I_k>L_k$, i.e., port expansion is only feasible when the other port also expands, there are two equilibriums (Y, Y) and (N, N). This situation occurs when the investment cost of a port is larger than the gain from the port expansion ($I_k>L_k$). However, if the competitor expands, it will exert serious impact on the target port. Thus, the best response for the target port is also to expand, to counter-balance the impact from the other port. However, if ports know the action of the other, no port will expand first, as $I_k>L_k$. Therefore, as a response strategy, (Y, Y) is impossible.

For the scenario $M_k<I_k<L_k$, there are two equilibriums (Y, N) and (N, Y). For each port k, $M_k<L_k$ means that the gain from expansion will be weakened by the expansion of the competitor, thus the port will be reluctant to make expansion decision if the other port expands. By comparing the net profits of each port under the two equilibriums, we find there is no dominating equilibrium.

For the scenarios $M_k < I_k < L_k$ and $L_l < I_l < M_l$, for first port k, it is optimal to make a different expansion decision than the second port l; while for port l, it is optimal to make the same decision as port k. Thus there is no equilibrium.

From Table 2, it is obvious that if one port has a clear indication on its strategy [$I_k<\min(L_k, M_k)$ or $I_k>\max(L_k, M_k)$], there will be a unique pure strategy Nash equilibrium. All the border cells in the table contain at least one port that has a very clear choice for expansion strategy. $I_k<\min(L_k, M_k)$ happens when the port expansion cost is low, and/or the gain from expansion is high. These are often associated with small scale ports and increasing demand in the hinterland, such as the port in Shenzhen in the past. On the other hand, $I_k>\max(L_k, M_k)$ often happens when the port is in an area with high construction costs, large scale operation, and the stable demand, which is much like the case for Hong Kong Port. If one port is in either one of the above conditions, the strategy of the other is easy to formulate, especially when it is depend on the strategy of the first one.

However, the investment cost – gain relation will change, and at certain stage, neither port can indicate a clear direction. In this case, the competition strategy will be more interactive and inter-dependent, such as the strategies in the middle cells of Table 2. At present, the scale of port capacity and port throughput in Shenzhen and Hong Kong is very close. However, there are still big differences in the port construction cost and operation cost between these two ports. Thus, port of Shenzhen can still give a clear indication on expansion strategy. If these costs increase in the future, then the port development strategies of the two ports will be more interdependent.

Next, we analyze how profit changes with capacity expansion. As the capacity expansion is discrete, we cannot analyze the gain using comparative static method. Therefore, we check the sign of the gain by subtracting the profit of non-expansion from that with expansion. Assuming the likelihood of expansion is proportional to the gain from expansion, we can determine which port is more prone to expand.

The proof of the nature of incremental profit with respect to its own capacity expansion is in Appendix. Note that the port expansion will increase its own profit only when $2\beta_i f_i(C^u)-2\beta_j f_j(C^l)+1>0$, where C^u is the new capacity that port i is considering reaching, and C^l is the existing capacity at the other port. In marginal cost, the condition is:

$$\frac{\partial MC_i^u}{\partial p_i}-\frac{\partial MC_j^l}{\partial p_j}<1 \qquad (10)$$

This condition states that for an expanding port to increase its profit, the marginal cost increasing rate should not be higher than that of the competitor by one. This points out that if the demand sensitivity is high at one port, expansion is not a good strategy to use as it cannot increase the profit.

3.2.3. Numerical Examples

In this section, we show the application of capacity expansion game using numerical examples, to illustrate pure Nash equilibriums in table 2.

Example 1 *In this example, we set* $\bar{x}=1$, α_i=0.6, α_j=0.4, β_i=0.1, β_j=0.2, θ_i=2, θ_j=1, C_i^0=3, C_j^0=2, ΔC=1. I_i=0.014, I_j=0.01, *where* ΔC *is the capacity increment due to the lumpy investment.*

The profits with different investment decisions under equilibrium prices are in Table 3. In this example, L_i=3.1240-3.1104=0.0136, M_i=3.0175-3.0032=0.0143, L_j=1.1751- 1.1638=0.0113, M_j=1.1478-1.1364=0.0114. This corresponds to the scenario $L_i<I_i<M_i$ and $I_j<\min(L_j, M_j)$. Since port *j* always has an incentive to expand, port *i* will expand, too. Therefore, the pure strategy Nash equilibrium is (Y, Y). Compared with the (N, N) decision pair, the (Y, Y) decision pair result in lower profits for both ports since 3.0175-0.014=3.0035<3.1104, 1.1478-0.01=1.1378 <1.1638. This is a Prisoners' Dilemma.

Table 3: Example 1 when investment cost not subtracted

	$j:N$	$j:Y$
$i:N$	3.1104; 1.1638	3.0032;1.1751-0.01
$i:Y$	3.1240-0.014;1.1364	3.0175-0.014;1.1478-0.01

The next four numerical examples are to show the impacts of base demand, market share, price sensitivities, and cost parameters on the investment decisions.

Example 2 *(Base demand and market share) In this example, we set* $\beta_i=\beta_j$=0.1, $\theta_i=\theta_j$=1, $C^0_i=C^0_j$=2, ΔC=1 *and* $I_i=I_j$=0.05. *Table 4 provides the equilibrium investment decisions for different base demand, and market shares of each port.*

Table 4: Impact of market demand on capacity expansion decision

	$\bar{x}=1$	$\bar{x}=2$	$\bar{x}=3$	$\bar{x}=4$
α_i=1;α_j=0	(*N, N*)	(*Y, N*)	(*Y, Y*)	(*Y, Y*)
α_i=1/2;α_j=1/2	(*N, N*)	(*Y, Y*)	(*Y, Y*)	(*Y, Y*)
α_i=1/3;α_j=2/3	(*N, N*)	(*N, Y*)	(*Y, Y*)	(*Y, Y*)

The results of example 2, given in Table 4, show that with the increase of the base demand $\bar{x}$, the two ports will be more likely to expand. When the total base demand is fixed at $\bar{x}=2$, the port will be more likely to expand when its market share increases, and the other port tends not to expand when its market share decrease.

Example 3 *(Price Sensitivity) In this example, we set* $\bar{x}$ =1, $\alpha_i=\alpha_j$=0.5, $\theta_i=\theta_j$=1, $C^0_i=C^0_j$=2, ΔC=1 *and* $I_i=I_j$=0.0125. *Table 5 shows the equilibrium investment decisions for different values of* β_i *and* β_j.

Table 5: Impact of price sensitivities on the expansion decisions

	β_j=0.1	β_j=0.2	β_j=0.3	β_j=0.4
β_i=0.1	(*Y, Y*)	(*Y, Y*)	(*N, Y*)	(*N, Y*)
β_i=0.2	(*Y, Y*)	(*Y, Y*)(*N, N*)	(*N, N*)	(*N, N*)
β_i=0.3	(*Y, N*)	(*N, N*)	(*N, N*)	(*N, N*)
β_i=0.4	(*Y, N*)	(*N, N*)	(*N, N*)	(*N, N*)

Example 3 shows the impact of price sensitivities on expansion decisions. When β_i=0.1, the increase in the value of β_j will make port *i* reluctant to expand. This is because the gain in port expansion decreases with the increase in cross price sensitivity. Table 5 also shows that the increase in the values of both β_i and β_j will make both ports not expand. When demand is insensitive to prices (small β_k), expansion can generate positive gain from the reduction in congestion. Otherwise, the gains from expansion cannot offset the increase in congestion cost due to the increased throughput.

Example 4 *(Operational Cost) Following example 3, set* $\bar{x}=1$, $\alpha_i=\alpha_j=0.5$, $\beta_i=\beta_j=0.1$, $C_i^0=C_j^0=2$, $\Delta C=1$ *and* $I_i=I_j=0.015$. *The investment decisions at equilibrium are in Table 6 for different values of* θ_i *and* θ_j.

Table 6: Impact of operational cost on the expansion decisions

	θ_j=0.5	θ_j=1.0	θ_j=1.5	θ_j=2.0
θ_i=0.5	(*N, N*)	(*N, N*)	(*N, Y*)	(*N, Y*)
θ_i=1.0	(*N, N*)	(*N, N*)	(*N, Y*)	(*N, Y*)
θ_i=1.5	(*Y, N*)	(*Y, N*)	(*Y, Y*)	(*Y, Y*)
θ_i=2.0	(*Y, N*)	(*Y, N*)	(*Y, Y*)	(*Y, Y*)

θ_k is proportional to the average marginal cost in the specific functional form for a given throughput level. This example shows that if a port has a large average marginal cost, its expansion can effectively reduce congestion cost.

Example 5 *(Investment Cost) In this example, we set* $\bar{x}=1$, $\alpha_i=\alpha_j=0.5$, $\beta_i=\beta_j=0.1$, $\theta_i=\theta_j=1$, $C_i^0=C_j^0=2$, *and* $\Delta C=1$. *Table 7 provides equilibrium investment decisions for different values of the investment costs* I_i *and* I_j.

Table 7: Impacts of investment cost on the expansion decision

	I_j=0.010	I_j=0.012	I_j=0.014	I_j=0.016
I_i=0.010	(*Y, Y*)	(*Y, Y*)	(*Y, N*)	(*Y, N*)
I_i=0.012	(*Y, Y*)	(*Y, Y*)	(*Y, N*)	(*Y, N*)
I_i=0.014	(*N, Y*)	(*N, Y*)	(*N, N*)	(*N, N*)
I_i=0.016	(*N, Y*)	(*N, Y*)	(*N, N*)	(*N, N*)

Example 5 shows that the increase in investment cost can reduce the possibility for expansion. When two ports have different investment costs, the one with lower investment cost is more likely to expand.

4. A Case Study

In this section, we use numerical analysis to demonstrate the application of the model in analyzing the competition between Hong Kong and Shenzhen container ports, the two ports sharing the same hinterland. Compared with Shenzhen, Hong Kong port has a relatively larger market share, a larger initial capacity size, and a higher operation cost for the same throughput. To compete, both ports can consider capacity expansion and adjustment in price. For equal size capacity expansion, the container port in Hong Kong needs a higher investment cost. Therefore, in this numerical analysis, we use a smaller port with lower market share, lower operational and investment costs to represent Shenzhen, and the other port for Hong Kong, i.e., $\alpha_H>\alpha_S$, $\theta_H>\theta_S$, $I_H>I_S$, $C_H^0>C_S^0$, where the index *H* and *S* refer to Hong Kong and Shenzhen respectively. Let $\alpha_H=0.6$, α_S=0.4, $\theta_H=1.5$, θ_S=1.0, $I_H=0.05$, I_S=0.03, $C_H^0=2.0$, C_S^0,=1.5, ΔC=1. In Table 8 and 9, we calculate the Nash equilibrium in expansion strategy, equilibrium prices (p_H^*, p_S^*), and equilibrium profits ($\tilde{\Pi}_H^*$, $\tilde{\Pi}_S^*$) for different values of β_H and β_S when both ports compete using pricing and investment. The letters in the parenthesis stand for the strategies for Hong Kong port and Shenzhen port respectively.

Table 8: β_H=β_S=0.1

	$\bar{x}$ =1.2	$\bar{x}$ =1.4	$\bar{x}$ =1.6
Both ports do not invest	p_H^* =7.29 p_S^* =6.42 $\tilde{\Pi}_H^*$ =4.3142 $\tilde{\Pi}_S^*$ =3.4231	p_H^* =8.50 p_S^* =7.49 $\tilde{\Pi}_H^*$ =5.8722 $\tilde{\Pi}_S^*$ =4.6593	p_H^* =9.71 p_S^* =8.56 $\tilde{\Pi}_H^*$ =7.6698 $\tilde{\Pi}_S^*$ =6.0856
At investment equilibrium	(N, N) p_H^* =7.29 p_S^* =6.42 $\tilde{\Pi}_H^*$ =4.3142 $\tilde{\Pi}_S^*$ =3.4231	(N, Y) p_H^* =8.37 p_S^* =7.26 $\tilde{\Pi}_H^*$ =5.7000 $\tilde{\Pi}_S^*$ =4.6941	(Y, Y) p_H^* =9.30 p_S^* =8.15 $\tilde{\Pi}_H^*$ =7.5023 $\tilde{\Pi}_S^*$ =5.9238

Table 9: β_H=0.1, β_S=0.3

	$\bar{x}$ =1.2	$\bar{x}$ =1.4	$\bar{x}$ =1.6
Both ports do not invest	p_H^* =7.77 p_S^* =2.45 $\tilde{\Pi}_H^*$ =4.9133 $\tilde{\Pi}_S^*$ =1.0981	p_H^* =9.07 p_S^* =2.85 $\tilde{\Pi}_H^*$ =6.6876 $\tilde{\Pi}_S^*$ =1.4946	p_H^* =10.37 p_S^* =3.26 $\tilde{\Pi}_H^*$ =8.7348 $\tilde{\Pi}_S^*$ =1.9521
At investment equilibrium	(N, N) p_H^* =7.77 p_S^* =2.45 $\tilde{\Pi}_H^*$ =4.9133 $\tilde{\Pi}_S^*$ =1.0981	(N, Y) p_H^* =8.74 p_S^* =2.65 $\tilde{\Pi}_H^*$ =6.2073 $\tilde{\Pi}_S^*$ =1.5297	(Y, Y) p_H^* =9.70 p_S^* =2.97 $\tilde{\Pi}_H^*$ =8.1578 $\tilde{\Pi}_S^*$ =1.9276

There are several interesting results from these two tables.

1. Both ports will expand if and only if the base demand is high. This is illustrated in the last column ($\bar{x}$ =1.6) in both tables.
2. Compared with Shenzhen port, Hong Kong requires a larger base demand to justify the expansion decision. When $\bar{x}$ =1.4, the optimal strategies for Shenzhen (Hong Kong) in both table 8 and 9 are to expand (not to expand). This implies that Shenzhen port enjoys an advantageous position in capacity expansion.
3. Any port expansion will decrease the equilibrium prices at both ports. From both tables, whenever the equilibrium investment strategies include expansion at any port, the prices at both ports are lower than their corresponding prices with no expansion.
4. Shenzhen port can gain from expansion. For example, in Table 9, when $\bar{x}$=1.4, expansion in Shenzhen can increase its equilibrium profit from $\tilde{\Pi}_S^*$=1.4946 to $\tilde{\Pi}_S^*$=1.5297, a gain of 0.0351. On other hand, Hong Kong will always suffer a profit loss from expansion.
5. Furthermore, port expansion will not increase total profit of the two ports. For example, while Shenzhen port gains from expansion, it is less than the losses in Hong Kong. The total profit of the two ports is lower comparing with the non-expansion case. In Table 8, when $\bar{x} = 1.4$, Shenzhen can gain profit from expansion (4.6941-4.6593=0.0348). However, the total profit of the two ports decreased by (5.8722+4.6593)-(5.7000+ 4.6941)=0.1374. When the pure strategy Nash equilibrium is both expands, they all have a lower profit than that with no expansion, which resemble the Prisoner dilemma situation.

5. Summary

This paper studied the competition between two ports, serving the same hinterland, using both capacity expansion and pricing. Quantity demanded at each port is a function of its own and the competitor's prices. In addition to the short-run market competition measures, such as pricing, the increasing market demand from the same hinterland also provides opportunity for capacity expansion. For given throughput, port expansion can reduce the marginal operation cost, which can lead to lower user cost and higher market share. To counter balance this impact, the other port also needs to reduce its price(s), to maximize its overall profit.

Based on the decision making process for port expansion and pricing, this paper constructed a two-stage game theory model to analyze the possible outcomes in this duopoly market. Using backward induction, we first analyzed the pricing strategy of the two ports for given capacity at both ports. We showed the unique Nash equilibrium for the pricing subgame following the Bertrand competition with differentiated products. For each port, its equilibrium price increases with its marginal cost, base market demand, and market share. Port expansion can reduce prices at both ports, which is beneficial to the users. We also analyzed the impact on equilibrium throughput and profit with the capacity change.

In capacity investment game, we identified the pure strategy Nash equilibriums between two ports, for different scenarios characterized by the possible gains from port expansion, and the investment costs. Using numerical examples, we show that each port will be more inclined to expand when the total market demand is high, or it has a large market share. A port will be more likely to expand if it has high operational cost, low investment cost and low own price sensitivity. In considering expansion, for one unit decrease in price, if the difference in marginal cost increases between the expansion port and competitor is less than one, then the expansion can bring positive gain to the expanding port.

Our case study, based on numerical examples, demonstrates possible outcomes from port competition between Hong Kong and Shenzhen. If expansion is constrained to be one at a time, the numerical results show that both ports can expand only when the market demand is sufficiently high. Shenzhen is more likely to expand when the market is increasing, but not sufficient for both to expand. Port expansion can bring benefit to users, but not necessarily to ports. Shenzhen can benefit from its expansion if Hong Kong does not expand. However, the gain from expansion at Shenzhen port cannot compensate the losses at Hong Kong Port. If both expand, then each port will have a lower profit than if both do not expand.

Finally, this study focuses on the payoffs for the two ports from private business operator's perspective. The social outcomes of port competition with pricing and capacity development will be a direction for future study.

6. Acknowledgements

The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. B-Q08K).

References

Abel, A.B. and Eberly, J.C. (1996) Optimal investment with costly reversibility. Review of Economic Studies 63, 581–593.

Acemoglu, D., Bimpikis, K. and Ozdaglar, A. (2006), Price and Capacity Competition. Working Paper.

Aguerrevere, F.L. (2003), Equilibrium Investment Strategies and Output Price Behavior: A Real-Options Approach. The Review of Financial Studies 16, 1239-1272.

Allahviranloo, M. and Afandizadeh, S. (2008), Investment Optimization on Port's Development by Fuzzy Integer Programming. European Journal of Operational Research 186 423-434.

Anupindi, R. and Jiang, L. (2008), Capacity Investment Under Postponement Strategies, Market Competition, and Demand Uncertainty. Management Science 54:11, pp. 1876-1890.

Baye, M. and Kovenock, D. (2009), Bertrand Competition Forthcoming in New Palgrave Dictionary of Economics, 2nd Edition. http://www.nash-equilibrium.com/ baye/Bertrand_Palgrave2e.pdf, accessed March 18, 2009.

Bean, J.C., Higle, J., Smith, R.L. (1992), Capacity expansion under stochastic demands. Operations Research 40, 210–216.

Besanko, David and Ulrich Doraszelski (2004), Capacity dynamics and endogenous asymmetries in firm size. RAND Journal of Economics. Vol. 35, No. 1., pp. 23-49

Chenery, H. (1952), Overcapacity and the acceleration principle. Econometrica.

Cheng, L. (1985), Comparing Bertrand and Cournot equilibria: a geometric approach. RAND Journal of Economics 16, 146-52.

Devanney, J.W.III and Tan, L.H. (1975), The Relationship between Short-run Pricing and Investment Timing: The Port Pricing and Expansion Example. Transpn Res. 9, 329-337.

Demichelis, S. and Tarola, O. (2006), Capacity expansion and dynamic monopoly pricing. Research in Economics, 60, 167-178.

Gilbert, R.J. and Harris, R.G. (1984), Competition with Lumpy Investment. Rand Journal of Economics 15, 197-212.

Gilbert, R.J. and Lieberman, M. (1987), Investment and Coordination in Oligopolistic Industries. The RAND Journal of Economics 18, 17-33.

Hall, J. and Porteus, E. (2000), Customer Service Competition in Capacitated Systems. Manufacturing & Service Operations Management 2, 144-165.

Hay, D.A. and Liu, G.S. (1998), The Investment Behavior of Firms in an Oligopolistic Setting. The Journal of Industrial Economics 46. 79-99.

Manne, A.S. (1961), Capacity expansion and probabilistic growth. Econometrica.

Manne, A.S. (Ed.) (1967), Investments for Capacity Expansion Size Location and Time Phasing. The MIT Press, London, Cambridge, MA.

Liu, L., Shang, W. and Wu, S. (2007), Dynamic Competitive Newsvendors with Service-Sensitive Demands. Manufacturing & Service Operations Management 9, 84-93.

Starrett, D.A. (1978). Marginal Cost Pricing of Recursive Lumpy Investments. The Review of Economic Studies 45, 215-227.

Tabuchi, Takatoshi (1994), Two-stage two-dimensional spatial competition between two firms. Regional Science and Urban Economics. 24, 207-227

Appendices

1. Comparative statics for the equilibrium price and profit.

Solving the two best response functions contained in (5) for the equilibrium price:
$p_k^* = \bar{x}(1+2f_k(C_k)\beta_k)\mu_k/(\lambda\beta_k)$,
where $\lambda = 3+2f_k(C_k)\beta_k + 2f_l(C_l)\beta_l$, and $\mu_k = 1+\alpha_k + 2f_l(C_l)\beta_l$

Differentiate the equilibrium price with respect to base market share, own price sensitivity and cross price sensitivity, we can obtain:

$$\frac{\partial p_k^*}{\partial \bar{x}} = (1+2f_k(C_k)\beta_k)\mu_k/(\lambda\beta_k) > 0$$

$$\frac{\partial p_k^*}{\partial \beta_k} = -\frac{\bar{x}\{3+4f_k(C_k)\beta_k[1+f_k(C_k)\beta_k]+2f_l(C_l)\beta_l\}\mu_k}{\lambda^2\beta_k^2} < 0$$

$$\frac{\partial p_k^*}{\partial \beta_l} = \frac{2\bar{x}f_l(C_l)[1+2f_k(C_k)\beta_k]\mu_l}{\beta_k\lambda^2} > 0$$

$$\frac{\partial p_k^*}{\partial \alpha_k} = \frac{2\bar{x}[1+2f_k(C_k)\beta_k][1+f_l(C_l)\beta_l]}{\beta_k\lambda} > 0$$

$$\frac{\partial p_k^*}{\partial \alpha_j} = \frac{\bar{x}[1+2f_k(C_k)\beta_k][1+2f_l(C_l)\beta_l]}{\beta_k\lambda} > 0$$

$$\frac{\partial p_k^*}{\partial C_k} = \frac{4\bar{x}[1+f_l(C_l)\beta_l]\mu_k f_k^{'}(C_k)}{\lambda^2} < 0$$

$$\frac{\partial p_k^*}{\partial C_l} = \frac{2\bar{x}[1+2f_k(C_k)\beta_k]\mu_l\beta_l f_l^{'}(C_l)}{\lambda^2\beta_k} < 0$$

From the demand function, we can obtain quantity demanded at the equilibrium by substituting the price with the equilibrium prices. The equilibrium quantity demanded at the equilibrium price is:

$$x_k^* = \frac{\bar{x}(1+\alpha_k+2f_l(C_l)\beta_l)}{3+2f_k(C_k)\beta_k+2f_l(C_l)\beta_l} = \frac{\bar{x}\mu_k}{\lambda}$$

The properties of the equilibrium throughput w.r.t. the demand parameters and capacity change are:

$$\frac{\partial x_k^*}{\partial \bar{x}} = \frac{\mu_k}{\lambda} > 0$$

$$\frac{\partial x_k^*}{\partial \alpha_k} = \frac{\bar{x}}{\lambda} > 0$$

$$\frac{\partial x_k^*}{\partial \alpha_l} = -\frac{\bar{x}}{\lambda} < 0$$

$$\frac{\partial x_k^*}{\partial \beta_k} = -\frac{2\bar{x}f_k(C_k)\mu_k}{\lambda^2} < 0$$

$$\frac{\partial x_k^*}{\partial \beta_l} = \frac{2\bar{x}f_l(C_l)\mu_l}{\lambda^2} > 0$$

$$\frac{\partial x_k^*}{\partial C_k} = -\frac{2\bar{x}\beta_k\mu_k f_k^{'}(C_k)}{\lambda^2} > 0$$

$$\frac{\partial x_k^*}{\partial C_l} = \frac{2\bar{x}\beta_l[1+\alpha_k+2f_k(C_k)\beta_k]f_l^{'}(C_l)}{\lambda^2} < 0$$

Using the equilibrium price and quantity, we can calculate the equilibrium profit. Differentiate the equilibrium profit with respect to its own capacity and the capacity of the other port, we obtain:

$$\frac{\partial \pi_k^*}{\partial C_k} = -\frac{\bar{x}^2[1+2f_k(C_k)\beta_k - 2f_l(C_l)\beta_l]\mu_k f_k^{'}(C_k)}{\lambda^3}$$

$$\frac{\partial \pi_k^*}{\partial C_l} = \frac{4\bar{x}^2[1+2f_k(C_k)\beta_k]\mu_l\beta_l\{2\alpha_k[1+f_l(C_l)\beta_l]\}f_l^{'}(C_l)}{\beta_k\lambda^3}$$

Since $f_k^{'}(C_k)<0$, we can see that if $1+2f_k(C_k)\beta_k - 2f_l(C_l)\beta_l > 0$, $\frac{\partial \pi_k^*}{\partial C_k} > 0$; $\frac{\partial \pi_k^*}{\partial C_l}$ is always negative.

2. Proof of $\Delta\Pi_i^*(C_j) \ge 0$

$$\Delta\Pi_i^*(C_j) = \Pi_i^*(C_i^1, C_j) - \Pi_i^*(C_i^0, C_j)$$

$$\Pi_i^*(C_i^1, C_j) = \frac{\bar{x}[1+f_i(C_i^1)\beta_i][1+\alpha_i+2\beta_j f_j(C_j)]^2}{\beta_i[3+2f_i(C_i^1)\beta_i+2f_j(C_j)\beta_j]^2} \quad \text{(A-1)}$$

$$\Pi_i^*(C_i^0, C_j) = \frac{\bar{x}[1+f_i(C_i^0)\beta_i][1+\alpha_i+2\beta_j f_j(C_j)]^2}{\beta_i[3+2f_i(C_i^0)\beta_i+2f_j(C_j)\beta_j]^2} \quad \text{(A-2)}$$

$$\Delta\Pi_i^*(C_j) = \left\{\bar{x}[1+\alpha_i+2\beta_j f_j(C_j)]^2\left[\frac{1+f_i(C_i^1)\beta_i}{[3+2f_i(C_i^1)\beta_i+2f_j(C_j)\beta_j]^2} - \frac{1+f_i(C_i^0)\beta_i}{[3+2f_i(C_i^0)\beta_i+2f_j(C_j)\beta_j]^2}\right]\right\} \Big/ \beta_i$$

The sign of above term is determined by the sign of the term in square bracket

$$\frac{1+f_i(C_i^1)\beta_i}{[3+2f_i(C_i^1)\beta_i+2f_j(C_j)\beta_j]^2} - \frac{1+f_i(C_i^0)\beta_i}{[3+2f_i(C_i^0)\beta_i+2f_j(C_j)\beta_j]^2} =$$

$$= \frac{1}{[3+2f_i(C_i^1)\beta_i+2f_j(C_j)\beta_j]^2} - \frac{1}{[3+2f_i(C_i^0)\beta_i+2f_j(C_j)\beta_j]^2} \quad \text{(Part A)}$$

$$+ \frac{f_i(C_i^1)\beta_i}{[3+2f_i(C_i^1)\beta_i+2f_j(C_j)\beta_j]^2} - \frac{f_i(C_i^0)\beta_i}{[3+2f_i(C_i^0)\beta_i+2f_j(C_j)\beta_j]^2} \quad \text{(Part B)}$$

$$\text{Part A} = \frac{[3+2f_i(C_i^0)\beta_i+2f_j(C_j)\beta_j]^2 - [3+2f_i(C_i^1)\beta_i+2f_j(C_j)\beta_j]^2}{[3+2f_i(C_i^1)\beta_i+2f_j(C_j)\beta_j]^2[3+2f_i(C_i^0)\beta_i+2f_j(C_j)\beta_j]^2}$$

$$= \frac{4\beta_i[f_i(C_i^0)-f_i(C_i^1)]\{3+\beta_i[f_i(C_i^0)+f_i(C_i^1)]+2f_j(C_j)\beta_j\}}{[3+2f_i(C_i^1)\beta_i+2f_j(C_j)\beta_j]^2[3+2f_i(C_i^0)\beta_i+2f_j(C_j)\beta_j]^2}$$

$$\text{Part B} = \frac{4\beta_i^3 f_i(C_i^0)f_i(C_i^1)[f_i(C_i^0)-f_i(C_i^1)] - \beta_i[f_i(C_i^0)-f_i(C_i^1)][3+2f_j(C_j)\beta_j]^2}{[3+2f_i(C_i^1)\beta_i+2f_j(C_j)\beta_j]^2[3+2f_i(C_i^0)\beta_i+2f_j(C_j)\beta_j]^2}$$

Add two parts together, and factor out the common part,
Part A + Part B=

$$\frac{\beta_i[f_i(C_i^0)-f_i(C_i^1)]}{[3+2f_i(C_i^1)\beta_i+2f_j(C_j)\beta_j]^2[3+2f_i(C_i^0)\beta_i+2f_j(C_j)\beta_j]^2}$$

$\times\left(4\{3+\beta_i[f(C_i^0)+f(C_i^1)]+2f(C_j)\beta_j\}+4f(C_i^0)f(C_i^1)\beta_i^2-[3+2f(C_j)\beta_j]^2\right)$ --- part C

Part C= $\left(4[3+2f(C_j)\beta_j]+4\beta_i[f(C_i^0)+f(C_i^1)]+4f(C_i^0)f(C_i^1)\beta_i^2-[3+2f(C_j)\beta_j]^2\right)$

$=\left([3+2f(C_j)\beta_j][1-2f(C_j)\beta_j]+4\beta_i[f(C_i^0)+f(C_i^1)]+4f(C_i^0)f(C_i^1)\beta_i^2\right)$

$=\left(3-4f(C_j)\beta_j-4[f(C_j)\beta_j]^2+4\beta_i[f(C_i^0)+f(C_i^1)]+4f(C_i^0)f(C_i^1)\beta_i^2\right)$

The assumption ensures that $2\beta_i f_i(C^u)-2\beta_j f_j(C^l)+1>0$. Use this inequality and note $f_k(\cdot), k=i,j$ are decreasing functions, it can be proved that

$$
\begin{aligned}
&3+4\beta_i f_i(C_i^0)+4\beta_i f_i(C_i^1)+4\beta_i^2 f_i(C_i^1)f_i(C_i^0)-4\beta_j f_j(C_j)-4\beta_j^2 f_j^2(C_j)\\
&>3+4\beta_i f_i(C^u)+4\beta_i f_i(C^u)+4\beta_i^2 f_i^2(C^u)-4\beta_j f_j(C^l)-4\beta_j^2 f_j^2(C_j)\\
&>1+4\beta_i f_i(C^u)+[2\beta_i f_i(C^u)-2\beta_j f_j(C^l)][2\beta_i f_i(C^u)+2\beta_j f_j(C^l)]\\
&>1+4\beta_i f_i(C^u)-[2\beta_i f_i(C^u)+2\beta_j f_j(C^l)]\\
&=1+2\beta_i f_i(C^u)-2\beta_j f_j(C^l)\\
&>0.
\end{aligned}
$$

Therefore, $\Delta\Pi_i^*(C_j)\geq 0$.

The change process in a state-owned airline: the implementation of new electronic services

Kostas Malagas[1], Nikitas Nikitakos[2] and Maria Ampartzaki[3]

[1] Olympic Airways, Athens, Greece.
Tel: +30 210 9267 127
Fax: +30 210 9267 177
E-mail: kmalagas@olympic-airways.gr
[2] Department of Shipping Trade and Transport, University of the Aegean, Chios 82100, Greece.
Tel: +302271035201
Fax: +302271035299
E-mail: nnik@aegean.gr
[3] School of Education, University of Crete, Rethymno 74100, Greece.
Tel: +30 28310 77704
Fax: +30 28310 77654
E-mail: mabarjaki@edc.uoc.gr

Abstract

This paper examines how new electronic services were implemented in a state-owned airline. Olympic Airlines and the Olympic Airways Services decided to implement the e-ticketing later than the other airlines following the instructions given by the International Air Transport Association (IATA). This study examines whether these innovations followed the phases of change that John P. Kotter (1995) suggested. According to Kotter (1995), changes and innovations should follow eight (8) phases in order to succeed. If one or more phases do not take place there is a great risk to the expected results. Kotter's theory will be questioned with reference to the experience of a number of employees, those who were involved more in the change process. CEOs, General Directors and members of the team who was accountable for the implementation of the above project participated in a semi-structured interview. The data enable us to have an insight and draw useful conclusions about the process and the phases of change.

1. Introduction

Aviation is one of the most competitive industries, operating at the frontiers of technological innovation. Technological innovation is the commonest change the aviation industry makes in an effort to meet the demands of a competitive market which includes low-cost, no frills carriers, and the e-commerce (Doganis, 2001). Companies prepared to adopt and innovate can take advantage of a rapidly changing technology which offers new services to meet every demand (Goffee and Jones, 2003). The present paper examines the phases of change that the introduction of electronic services went through in the Olympic Airlines and Olympic Airways organizations. These electronic services contained e-ticketing, a new Reservation system, a new Departure Control system, new Internet services, a new Revenue Management System and a new Frequent Flyer Program. By documenting the process of change in a state-owned airline, it is anticipated that the conclusions may be useful to other airlines and organizations from other industries which are in a similar situation.

2. Electronic Services as a Feature of Information Technology Development in the Contemporary Airline Industry

Information Technology (IT) assists aviation industry to improve its customer services and respond to contemporary challenges. Thus, IT reinforces the development of new relationships between all the participants in the aviation industry such as airlines, passengers, airport authorities, operators, ground handlers and aircraft (SITA, 2008). All the participants should collaborate and their activities are interrelated, however here we focus on the IT impact on the airline operation. Thus, IT has helped airlines to reduce costs, respond efficiently to the demands, create good value for money, generate new revenues and, improve customer services (Ibid.). The past decade was affected by three major IT innovations: 1) the introduction of Internet; 2) the simplification with the introduction of e-ticketing; and 3) the rise of e-commerce (Ibid.). The introduction of e-ticketing, as this was dictated by IATA, was important to the aviation industry and to passengers. Since aviation industry started using the program of 'Simplifying the Business' with key point the implementation of e-ticketing, it achieved annual savings of US$6.5 billion (IATA, n.d.).

IT services require changes within industry and airlines which make the necessary changes in order to respond to the new requirements and support the new services. In most of the cases airlines outsource these services from the few providers that operate in the global market. Although the e-ticket played a pivotal role for the adoption of new technologies, e-ticket alone is not an issue of product differentiation and a competitive tool (Coleman in McDonald, 2009a). However, airlines do not operate autonomously and they must interact with distributors and service providers. They should follow the industry standards and trends in order to differentiate and outshine faster, as Moore (in McDonald, 2009b) stressed. In addition, airlines are benefited from the synergies that providers can achieve hosting a large number of airlines (for example the SabreSonic customer service system hosts 100 airlines) which means they can offer advanced and cost-efficient services to customers.

3. Olympic Airways and Olympic Airlines – In Need of Innovation

Olympic Airways is the Greek state-owned national airline which suffered a large debt that developed into a spiral of annual losses (Doganis, 2001). In 1994 and 1998 two restructuring plans, imposed by the European Union were proposed in order for the airline to recover. However the 1998 restructuring plan had not been put in place, none of the financial objectives had been met and the viability of the company was jeopardized. Olympic Airways had lacked equity since 1999 and remained very much in debt (European Union reports, 2002). This led to the Commission's decision (11 Dec. 2002) to penalize Olympic Airways for not achieving the objectives of the second restructuring plan. In 2003, Olympic Airlines was created, getting from Olympic Airways the operations part and the related staff. The airline implemented the new Passenger Service System and took up the related costs in order to become more competitive and more attractive to private investors.

Olympic Airlines and Olympic Airways have decided to introduce new electronic services later than the other airlines and just a year before the deadline for the e-ticket implementation imposed by IATA (31 May, 2008). The two companies' (i.e. Olympic Airlines and Olympic Airways) management decided to change the reservation, airports departure control, internet sales, frequent flyer, and the revenue management systems. This project was extremely crucial for the companies' viability. It was also highly risky due to the introduction of major changes in a short period and a year before the IATA's deadline for the paper tickets. In addition, the Olympic Airlines and Olympic Airways managements decided to outsource these services. In the previous system, the majority of the work was done by the airline's IT Division. Most of the biggest providers responded to this invitation and submitted their proposals. Olympic Airlines and Olympic Airways appointed technical and financial committees in order to select the most suitable provider. The migration project started on 13 February 2007 and was completed on 17 June 2007. Since then, the airline uses the new services with significant implications on their operations and in extent, on the services the carrier offers to passengers.

Thus, Olympic Airways and Olympic Airlines did launch the introduction of a new electronic services system of several components with a migration project which lasted only five months. According to industry officials Olympic achieved the quickest migration from the old system to the new one (IATA, 2008). However, as mentioned before, the company did face some serious problems at that time. Having too many different aircraft types, bureaucratic and over-centralised management and poor service quality both in the air and on the ground are the main symptoms of the "distressed state airline syndrome", which Olympic Airways clearly suffered from (Doganis, 2001). But the major problem for Olympic was the continuous government intervention. In addition, some particular employee resistance was expected from the IT department, as the company's management decided to outsource the new electronic services from a well-established provider. Previous systems were developed and maintained internally but after careful consideration and consultation of the IT department the Olympic's management decided to go for the external provider.

4. The Phases of Organizational Change

The implementation of this innovation and the organizational change which occurred in Olympic seemed to be quite difficult and complex as it is usually the case. Organizational change influences people's lives and keeps people away from a feeling of unwillingness to do anything. Success or failure of these efforts may have significant impact to an organization's viability. Researchers who focused on the study of transition process and the implementation of innovation in companies suggested that change goes through certain phases (Lewin, 1946; Marzalek-Gaucher and Coffey, 1991; Bair and Gray, 1992; Conner, 1992; Steinburg, 1992; Jackson, 2000). Browsing the relevant literature, we came across Kotter's (1995) theory about the eight phases of a change process. After examining a large number of organizations which implemented important projects, Kotter found that the change process goes through different phases. Each of them needs its own time to be completed. Skipping some phases might give the sense of a greater speed but this is not expected to lead to success. Kotter's eight phases are particularly helpful when examining a company into the process of change because they encompass all the factors and parameters mentioned by other researchers. It is an all-inclusive theory which takes into account the time, the necessary preparation, the strategies, and the human factors (including people's motivation and the change of behaviors).

The eight phases an organizational change and innovation needs to necessarily go through according to Kotter are:

1. the development of a sense of urgency among the staff and the innovation developers
2. the development of a strong project team to act as a guide for the rest of the company
3. the development of a strong and vivid vision
4. the transmission of this vision to the staff
5. the assignment of extra responsibilities to a growing team of participants in the implementation of change
6. planning for quick short-term successes and targets to increase motivation
7. stabilization of the improvements and planning for further changes
8. stabilization of new methods (including the change of company culture and behaviors)

Olympic Airways and Olympic Airlines experienced the urgency to change and went through a series of changes and developments in order to survive. The question, therefore, is what contributed to this successful migration? Was it the result of some good planning and strategic management? Did this process of change go through the phases proposed by Kotter? Did it take into account all the different factors that affect people and processes? Or, were there any other elements and risks which we need to be aware of when speaking about innovation and change within a state-owned company such as Olympic?

5. Sampling and Methodology

The sample of this study was 13 employees, those who were most involved in the implementation of the new services and the change process that took place in Olympic Airlines and Olympic Airways. These included CEOs, General Directors and members of the implementation/project team. This "convenience sampling" method was applied because we purposefully wanted to choose the staff who experienced directly and intimately the process of migration to the new system of electronic services. These were the people who prepared the way, took risks and made decisions, demonstrated new behaviors and educated the staff so that the new knowledge and information was transmitted to all.

We wanted to analyze the experience of the key-persons and find out the phases this process went through. The study did not purpose to study the behaviors of the entire staff involved in the process, thus, we cannot be certain that the rest of the employees shared the same view as the interviewees. This is a qualitative study in which external validity is replaced by the concept of transferability. Transferability is the ability of research results to transfer to situations with similar parameters, populations and characteristics. The study has a significant degree of transferability which is supported by careful transcription, thorough and systematic analysis of the data and information about the theoretical context which was used to generate and make sense of the data.

The present study is based on the analysis of the employees' narrative. This was because we wanted to find out how the individuals personally experienced the different phases of the migration project and how they understood their role and contribution within this particular context and situations. As narrative is experiential, it can provide a description of "*tacitly held personal knowledge without abandoning the particular, the contextual, and the complex*" (Conle, 2000: 51). The narrative was guided by a semi-structured interview comprised by twenty-two questions. These questions were designed and developed having Kotter's phases in mind. Furthermore, some additional questions were added to investigate the employee's feelings about the outsourcing decision and the opinion they developed about the external provider. The interview questions were piloted with five participants first and some minor changes and improvements had been made. The participants had been made aware of the purpose and the content of our research. During the interview the interviewer remained mainly silent, to avoid making any comments which would influence the participants answers.

We must acknowledge that the truth the narrator believes s/he is telling is very fluid and depends on the time and the perspective adopted. The same narrative might be influenced as the story is re-narrated or by the involvement of other actors. As life itself is never tidy, it might appear that a piece of narrative highlights a particular time in the participant's life, thinking and understanding. The answers interviewees gave to the questions of the research can only be considered as *snapshots* of their understanding at that particular place and time. Conle (Ibid.) points out that the narrators may not recognise how much their story is influenced by their socio-cultural milieux or that it has grown organically within a certain socio-cultural environment, such as that of the Greek state-owned companies. They also do not recognise that their interpretation of the events is also culturally bound and, therefore, their understanding might not be the same with that of people from a different group (such as the other company departments). However, narrative can reveal how people experience change and which factors influence situations and make particular events possible.

This study faced the dilemma of anonymity as there are some particular details that identify the participants of this research. The participants, for example were people who had a significant contribution to the aforementioned project. This group of people was relatively small (13 participants) and the majority of them had key managerial positions and roles within the company and within the project. They could, therefore be identified since their work, stance towards the project and their opinions were well known. However, all names had been eliminated to provide a basic degree of anonymity and special effort was made so that all the information entrusted was handled with care.

6. Data Analysis

The analysis of the data went through three phases as those identified by Wolcott (1994): the transcription, the analysis and the interpretation phase. The first phase included the transcription of the interviews. In the analysis phase we tried to identify key words, essential features, experiences and the interrelationships between them (Ibid.). Particular attention had been paid so as to identify whether the participant answered positively or negatively to the question because this was not always clear. The analysis also was employed to identify how things worked (Ibid.), that is what were the particular circumstances that defined each participant's experience and influenced the way s/he experienced the different phases and issues of the project.

In the interpretation phase we tried to "*make sense of meaning in context*" (Miles and Huberman, 1994: 14), in other words to link the case of each participant with larger issues and the theory

(Wolcott, 1994). In our case, we linked the participant's answers with the theory of the eight phases of change that John P. Kotter (1995) suggested. According to Kotter, changes and innovations should follow eight (8) phases in order to succeed. The main question was: can we identify the occurrence of the eight phases in what the participants had experienced? Is Kotter's theory to be confirmed or do we see some kind of special variation due to the particular circumstances? As we understand, this kind of interpretation included some hermeneutic work which aimed to introduce new aspects to the original narrative that exceeded the mere description (Bennetts, 2002).

7. Results and Discussion

The results of the present study show that in general the introduction of new e-services in Olympic did follow the eight phases Kotter (1995) suggests an implementation of innovation goes through. However, there were some particular features which make this process a special case. Let's look at the features of each phase separately, make a comparison with Kotter's theory and discuss the results.

In the first phase Kotter (1995) emphasizes that there is a real need for a company to create a sense of urgency to her employees. The introduction of change and innovation starts when some individuals or groups in a company start looking at the competitiveness of a company, her place in the market, new technological developments and the company's financial performance. Once people realize that the company is about to fall behind they can also alert the others or even the management of the company so that a resolution is made. There are not very many chances, according to Kotter for a company to succeed in change unless people feel the urgency to act upon issues or problems. As we discussed above, other researchers do seem to agree that this sense of urgency is of pivotal importance to the process of change (Conner, 1992; Simpson and Cacioppe, 2001). Although the best approach to change is to move in smaller rather than large steps (Kiefer, 1992) usually changes have to be implemented within strict time and budget limits (Conner, 1992; Steinburg, 1992). It is likely, therefore, to cause high levels of stress and speculation to an organization's management demanding multilevel handling and skillful treatment. In our case, the participants of the study can confirm the existence of the above mentioned sense of urgency and the dissemination of all the necessary information to the staff so that every individual understood how significant the project of e-services was for the company's viability. However, as a significant number of the respondents highlight, this sense of urgency came from an external alert which in fact, was IATA's resolution about the introduction of the e-ticket. The interviewees also commented on the company's hesitation to follow current industry practices and the slow developmental rhythms caused by the bureaucratic nature of the state management. The government who is the owner of Olympic and the company's bureaucratic management left the decision for change up to the last minute which was a very serious risk for the company. Kotter (1995) estimates that this first phase is the most difficult and the company's dull situation apart from worrying is also beneficial since it *fuels* this big sense of emergency. However, it does not allow sufficient *space for maneuvering* which might make the situation critical. In the case of Olympic there was no space and time for alternative plans, failed trials or mistakes and the whole project became critical for the company's viability. This had a series of knock-on effects which influenced the other phases of the change process as we are going to see at a later point.

The second phase of the innovation process is the development of a project team with strong managerial skills, excellent subject knowledge and commitment to change. This includes an authorization so that they can function and surpass the official hierarchy, the norms, the expectations and the protocol of the company in order to make the implementation of decisions quickly and efficiently. It is also vital for this team to have strong cooperative skills and be able to inspire the rest of the staff (Ibid.). Kiefer (1992) is a researcher who agrees, among others (e.g. Marzalek – Gaucher and Coffey, 1991) that any kind of change implementation demands the building of a leadership team, which is going to create a strategy for guiding the rest of organization through change. Part of the team's power derives from the provision of the necessary capacity and resources by the company's top management (Simpson and Cacioppe, 2001). The participants in this study did also confirm the development of a highly skilled project team and they could also tell that this team was relatively independent so as to be fast, flexible and efficient. However, the team followed the company's protocol in general and as it was reported by the interviewees, there were times when higher order decisions had to obtain the general director's consent (who was not part of the project team).

What comes next is the development of a vision which acts as a guide to the effort for change. Consequently, strategies for the realization of this vision need to be planned and implemented (Kotter, 1995). Simpson and Cacioppe (2001) agree that a commitment to a clear shared vision is an important element which contributes not only to the implementation of change, but also to the transformation of a company's culture. In our case all participants were clear about the vision which was the introduction of new e-services within the strict time limits set by IATA (with the consequent survival of the company) and the improvement of the company's profile which was quite low in the market. However, the interviewees could not provide clear answers regarding the strategies developed to lead to the successful realization of the vision. This causes doubt and questions regarding the communication between the management, the project team and the rest of the staff regarding the whole issue of the implementation of change.

It brings us to the next phase identified by Kotter (1995) which is the transmission of the vision of change to the staff. As Conner (1992) argues communication is vital in making change work. This communication should clarify not only the general aims and the specific targets of the change project but also the expectations from staff, the impact of this change, how they are going to achieve what the vision includes and some assessment of the effort made up to the moment. It is also of particular importance that the project team demonstrates behaviors and models the way things should work in order to succeed (Kotter, 1995). Simpson and Cacioppe (2001) agree that modelling of new behaviours and the use of new knowledge is a "ground rule" important to every transformation, and Steinburg (1992) stresses that if radical changes should take place, old hierarchies, traditional jobs and traditional ways of making decisions should be replaced. As we have seen in the results, the participants in this interview referred to a variety of ways which were used by the company to disseminate information to the staff. However, from the collected answers we came to realize that there was no systematic effort to keep the staff up to date with the upcoming developments. We realized that the majority of the staff, even the individuals who were heavily involved in the project had no clear idea about the strategies the company wanted to implement. The participants did recognize that new behaviors were transmitted and modeled by the project team for the rest of the staff.

In the fifth phase of the change process Kotter (1995) identifies the need for the involvement of more people and their encouragement to take extra responsibilities and risks in order to help the implementation of strategies. In this phase the *seeding* of other departments and groups is attempted which aims to spread the learning an innovation demands (Kiefer, 1992). This phase also includes the removal of obstacles whether these are related to the company structure and working protocol or to the particular attitudes developed by the employees. One good strategy to achieve this is to find groups of people who want the change and respond positively to it. Again, it is necessary for the management to give those people authority to model and enforce the change behaviors to other members of staff (Kotter, 1995). The interviewees agreed that with time more people were involved in the project and they were given permission to run mini projects in order to support the process of innovation. However, the issue of employee resistance came up. The fear towards the new and the new work conditions was the obstacle participants mentioned most. Other obstacles included financial issues and delays caused by the company's bureaucracy. The interviewees also reported that risks were taken with their personal responsibility which was the most dangerous of all. It shows that in a few cases the project team did not receive the appropriate support by the company's top management.

The sixth phase Kotter (Ibid.) identifies within the change process is the planning and completion of short-term successes and successful steps, what Simpson and Cacioppe (2001) call *actionable first steps*. Apart from strategic planning, this includes the rewarding process for the staff that is mostly responsible for these successes and the improvement of the staff's morale (Kotter, 1995, see also Simpson and Cacioppe, 2001). The answers collected from the interviewees in this study were contradictory. This shows that there was no strategic planning of short-term successes as such to boost the staff's morale neither was there any reward system to be used for the support of the change process.

In the seventh phase there is a need for the stabilization of improvements and the development of further changes. The management of the project is to use the example of successful implementation in order to persuade the staff about the reliability and success of the changes (Kotter, 1995, see also Simpson and Cacioppe, 2001). As part of this process Kiefer (1992) suggests the development of learning communities in which different groups can learn from each other. Rather than changing everything in once, the company might use this idea for gradual spread and the learning of new knowledge and behaviors. The best learning in organizations happens when organizations learn from their own successes and failures (Ibid.). In the case of Olympic the implementation of the new e-services did help the staff to see the project's reliability and the need for further improvements and changes, but there was no coordinated effort to bring groups, teams and departments together in order to spread, reinforce and enhance change.

Finally, in the eighth phase Kotter (1995) identifies the need for a change in the company culture and behavior so that consolidation and further development continues. The development of an organization culture which complies with and supports the stabilization of improvements and a continuous evolving progress is the ultimate target of this phase. Research shows that organizational culture directly affects performance and the quality of customer services a company can offer (Kotter and Heskett, 1992; Cummings and Worley, 1997; Lee and Yu, 2004). Schurink (1990) and Burkey (1993) claim that the foundation of change is based on a personal process of change, which involves the acquisition and assimilation of knowledge and information. Referring to Lewin's 'unfreezing phase' Peters and Tseng (1983) pointed out that the unfreezing process is often quite difficult because it involves moving away from the old comfortable way of doing things to a new unfamiliar way. Organizations are learning that they have to take into account human resources. First and foremost, managements should change not only the way they run the business but also the way people think and interact (Steinburg, 1992). This is because humans develop problematic behaviors when they are not able to adjust to changes. In this case the rate of resilience among employees should increase so that higher levels of change are accepted without people being distracted (Conner, 1992). The literature suggests that the active participation of all personnel involved in the process of change is necessary in order for this process to become successful and lasting (Cope, 1981; Jackson, 2000; Kennedy, 2000). Therefore, it is important for all the staff to understand the required change and align with the adopted innovations. Alexander (1985) and Kotter (1995) stated that the majority of change projects are not successfully implemented due to difficulties and obstacles they come across during the change process. The right motivation system therefore is of crucial importance. Providing the necessary awards to employees and giving the required information about the benefits yielded by the introduction of change, management can achieve the employees' high level of enthusiasm which is vital to the change process (Steinburg, 1992). However, the participants of the study can argue that a change in the organizational culture of the Olympic is not a straightforward issue. They can see the deeper change needs further initiatives and a closer look at the company's fundamental problems. The participants can see some change initiated by the work requirements of the new services; however, this does not seem to be enough to generate a whole company improvement. Here, a variety of reasons which were spotted by the interviewees seem to have an important effect. These are identified in several other research studies (Alexander, 1985; Overholt et al*.*, 1994; Carnall, 1995; Strabel, 1996) and include the financial difficulties, priority of other businesses (as, in our case, the expected privatization of Olympic), technical difficulties (such as the shortage of aircrafts in our example), fear of insecurity, lack of skills and resources, commitment to the current practices, strong organizational culture, powerful trade unions and, finally, government regulations.

In addition, we have explored the interviewees' feelings about the solution of outsourcing which was chosen instead of the in-house development. Firms use outsource solutions in order to introduce advanced services particularly for IT projects because they may not have sufficient knowledge or expertise in-house, or they may be under time or budget pressures (Bathelemy, 2001; Gantz, 1990; Sengupta and Zviran, 1997). The success of outsourced IT projects depends, to a large degree, on the commitment of the client to the project, the skills of the consultants, the collaboration between the client and the vendor, and the level of information shared between client and the vendor (Bettencourt, et.al. 2002; George and Chattopadhyay, 2005; Slaughter and Ang, 1996). Nevo, Wade and Cook (2006) point out that frequent and honest communication, the existence of in-house dedicated staff, and contract management are also necessary for the successful completion of a project. The cultural compatibility between the two organizations and the understanding of the client's business goals are necessary. They also stressed that firms should use external assistance after evaluating their internal IT knowledge and should this is weak. The above situation seems to apply to the Olympic case. The participants of the study realize that the IT provider was selected due to the similarity with the carrier's previous system and the lower cost. They can also recognize that carrier's IT department was unable to provide these services in-house. The lack of the appropriate software, sufficient or specialized staff and the strict time limits were also critical factors which influenced the decision of outsourcing. Interviewees believe that although there was some resistance by the carrier's IT staff, the majority of employees had understood the importance of these services to the carrier's viability. The similarity of the new system with the old one was the basic reason that the resistance eventually lowered. The high ranked managers supported outsourcing in most cases and the results look very good for the company for the time being (Malagas and Nikitakos, 2008).

8. Conclusions and Implications for Further Research

The aim of the present study was to examine whether the introduction of new electronic services to Olympic Airlines and Olympic Airways has followed the phases of change that Kotter (1995) suggested. Taking into account the experience of a number of employees, who were most involved in the change process the study searched for evidence which suggests that the transition to change followed the eight important phases Kotter suggested and whether this explains the success of the given migration project.

In general, the results showed that the eight phases suggested by Kotter (1995) have occurred in the case of Olympic also. However, a second, closer look causes some wonder and further speculation. The company's management did make the decision to migrate to new electronic services and made a careful selection of an external provider. The management then moved on with the careful construction of a highly skilled and committed project team which undertook the development and implementation of the migration project. However, as interviews show, it was hereabout the management's effort slowed down. As interviewees revealed it was mainly their commitment and willingness to take risks which, actually, took them through this project. Project leaders had to fight against the odds straight from the beginning since the decision for the introduction of new electronic services was already late. It seems that a considerable degree of this success was very much due to people's improvisation rather than the company's strategy. In addition, there was not a systematic and efficient reward system or further effort on the company's behalf to boost the staff's morale and keep the impetus alive. Although some overtime had been paid and project leaders received some kind of bonus the interviewees feel that the company had not done enough about this as she never actually does. We have here the particular phenomenon, which frequently appears in the Greek society: Greeks are moved and initiated to act by *philotimo*. *Philotimo* is a sense of honor, empathy, and friendship the Greeks might develop when confronted by a difficult situation, or a common threat. They will, then, unite and act with selflessness or even self-sacrifice (Potamianos, 1999). It is clear that the Olympic employees who were involved in the project most, acted primarily on the basis of *philotimo* and a sense of responsibility to the Greek people.

Although everyone became clear about the criticality of the situation the management did not direct or coordinate the further updating and education of all the staff so that project leaders get everyone's support. In addition, leaders had to take on the responsibility of educating the staff and modeling new behaviors and a new organizational culture. Education is very important for a successful implementation of change as most of the times resistance to change comes from the employees' smatter. It is expected that every company that goes through change gets into a transition period of smatter until the staff becomes accustomed with the new products and acquire new knowledge from using them (Carnevale, 1992). Therefore, it is not an easy task to unlearn a specific way of doing business and to implement a new one and top managements have a lot to do towards this. A trained and educated staff that knows more can do more when management enables them with the right resources and strategies (Conner, 1992). According to Mezirow (1981) transformative learning is a process of critical reflection on the current knowledge, attitudes and beliefs that puts new and old practices into a new perspective. As the interview showed, the participants changed their perspective about customer services and the competitiveness of the Olympic. This enabled them not only to change their own viewpoints but also to assess the general performance of the company (and so they identified that a more inclusive and holistic innovation plan was necessary for the total make over of the company). However, these reflection skills do not seem to be widespread among the rest of the staff – the particular issues interviewees highlight regarding each phase of the change process, show that the company as a whole (staff and managers included) was never actually involved in any kind of critical reflection in the light of the new developments. The new electronic services had already had good commercial and financial effects (Malagas and Nikitakos, 2008). However, without a significant and overall cultural change the positive impact is going to be minimized or undermined as in the case of Olympic, which is now on the way to privatization as the big losses continue and the state is rendered incapable to run the business. According to Kotter and Heskett (1992) the organisation culture influences the company's long-term economic performance even when the company is comprised by a team of highly skilled staff.

An important factor which proved to be pivotal for the success of the migration project was the fact that the newly introduced system of electronic services had many similarities with the old one, fitted adequately the company's organisation and caused less stress for staff training. The selection of the electronic solution the company chose seem to have the strategic and cultural fit and the technology readiness that McLaughin (2009) identifies as necessary for the successful implementation of an e-product. However, despite the cultural fit, the modified organizational culture that emerged changed employees' life after the project implementation. The new automatic services created new jobs on one hand and eliminated others. The result was that at least 80 people lost the subject of their work. As the company functions within the framework of a state bureaucratic regime she cannot terminate those job contracts neither create new job descriptions for these people easily. This means that the Olympic could not flexibly relocate people and responsibilities and that certainly had an impact on the overall culture the developers of the new product were trying to create.

Summarizing, we can see that the introduction of new electronic services (and the change process that occurred) did follow Kotter's phase of change, however, this was not due to a well-developed strategic planning. What actually contributed to the successful migration into the new technology was the project leaders' commitment and personal interest. The critical factor of bureaucracy which is a defining element of every Greek state-owned organization was very influential to the outcome. It had an impact on functions at all levels of organizational structure and decision making. In combination with all the other operational problems that company faced, the introduction of new electronic services did not manage to take Olympic out of the critical financial situation. We can therefore tell that the phases an implementation of change might go through will vary and develop different characteristics according to the specific situations a company might face at the time of innovation (see for example, Okumus and Hemmington, 1998). Organizational and national cultures as well as the conditions of employment are significant factors which need to be taken into account before a theory about change is drawn.

Based on the above evidence we can now see the need for some further study into the field of Olympic's organization culture. An insight into the working lives of Olympic's employees would enable us to *triangulate* (Miles and Huberman, 1994) the results of the current research and find out how Olympic's particular cultural features influenced the implementation of the new electronic services and the development of a new customer charter. An ethnographic study is highly recommended for any future work on this issue. Besides, a detailed study of the employee's working lives, attitudes and skills would also enable us to understand better the role the project leaders played during the transition process and help us verify the results and conclusions of the present paper.

Adopting a "census approach" would also be beneficial since it would allow us to detect the different levels of behavior, understanding and attitudes towards innovation among the entire staff.

Closing this discussion we have to acknowledge the possible influence the time factor on the participants' responses. These interviews were taken about a year after the implementation of the migration project and a few months before the initiation of the company's privatization process. We cannot be sure about the impact of this development on the interviewees' answers. It is important however, to bear in mind that the final verdict that Olympic can no longer survive as a state-owned business must have caused several worrying feelings and some degree of disappointment, since the leaders of the given project had worked hard for the survival of the company.

References

Alexander, L.D. (1985), Successfully implementing strategic decisions. Long Range Planning, 18(3): 91-97.

Bair, J. and Gray, M. (1992), The occupational therapy manager, Maryland: American Occupational Therapy Association.

Bathelemy, (2001), The hidden costs of IT outsourcing, MIT Sloan Management Review, 42(3): 60-69.

Bennetts, C. (2002), Traditional mentor relationships, intimacy and emotional intelligence, Qualitative Studies in Education, 15(2): 155-170.

Bettencourt, L.A., Ostrom, A.L, Brown, S.W. and Roundtree, R.I. (2002), Client co-production in knowledge-intensive business services, Management Review, 44(4): 100-128.

Burkey, S. (1993), People first: A guide to self-reliant participatory rural development, ZED Books: London.

Carnall, C.A. (1995), Managing Change in Organizations, Prentice-Hall: Hemel Hempstead.

Carnevale, A. (1992), The Path Is the Thing, Training and Development, March 1992: 36-37.

European Union reports (2002), The Olympic Airways Case. Commissions Press Room, 11/12/2002

Conle, C. (2000), Narrative Inquiry: research tool and medium for professional development, European Journal of Teacher Education, 23(1): 49-63.

Conner, D. (1992), Five views of change - the Difference Between Winners and Losers, Training and Development, 46(3): 34.

Cope, D.E. (1981), Organisation development and action research in hospitals, Aldershot, Hants: Gower.

Cummings, T. and Worley, C. (1997), Organization Development and Change, 6th ed., South-Western College Publishing: Ohio, USA.

Doganis, R. (2001), The airline business in the 21st century, Routledge: London.

Gantz, J. (1990), Outsourcing: threat or salvation?, Networking Management, 10(90): 25-40.

George A. and Chattopadhyay, P. (2005), One foot in each camp: the dual identification of contaract workers, Administrative Science Quarterly', 50(1): 68-99.

Goffee R. and Jones, G. (2003), The character of a corporation – How Your Company's Culture Can Make or Break Your Business, Second Edition, Profile Books Ltd: New York, USA.

International Air Transport Association (IATA)'s website: www.iata.org, (n.d.), Simplifying the Business' initiative, last accessed in July, 13, 2008.
International Air Transport Association (IATA)'s website: www.iata.org (2008), IT Symposium, last accessed in December, 18, 2008.

Jackson, D. (2000), Becoming dynamic. Creating and sustaining the dynamic organisation, Macmillan: London.

Kennedy, D. (2000), Acceptance. Developing support for change, Berrett-Koehler Communications: San Francisco.

Kiefer, C. (1992), Changing Organisms Instead of Mechanisms, Training and Development, March 1992: 36.

Kotter, J.P. (1995), Leading change: why transformation efforts fail, Harvard Business Review, Vol. 73 (2), March-April: 59-67.

Kotter, J.P. and Heskett, J.L. (1992), Corporate Culture and Performance, Free Press: New York, N.Y.

Lee, S.K.J. and Yu, K. (2004), Corporate culture and organizational performance, Journal of Managerial Psychology, Vol. 19(4): 340-359.

Lewin, K. (1946), Action research and minority problems, Journal of Social Issues, Vol. 2: 34-36.

Malagas, K. and Nikitakos, N. (2008), Introduction of electronic services in an airline: Change management issues. Paper presented at the12th Air Transport Research Society (ATRS) World Conference: Athens July, 6-10, 2008.

Marzalek-Gaucher, E. and Coffey, R.J. (1991), Transforming healthcare organisations, Jossey-Bass: San Francisco.

McDonald, M. (2009a), Do technology standards make sense for airlines' attempts to differentiate? Travel Technology Update.

McDonald, M. (2009b), Technology standards don't stifle differentiation, proponents insist. Travel Technology Update

McLaughin, S. (2009), The imperatives of e-business: case study of a failed project, Journal of Business Strategy, 30(1): 40-49.

Miles, M.M. and Huberman, A.M. (1994), Qualitative Data Analysis: An Expanded Sourcebook, Sage: London.

Mezirow, J. (1981), A critical theory of adult learning and education, Adult Education, Vol. 32: 3-27.

Nevo, S., Wade, M.R. and Cook, W.D. (2006), An examination of the trade-off between internal and external IT capabilities, Schulich School of Business York University, Operations Management and Information Systems: Keele Street, Toronto, Ont., Canada M3J 1P3.

Okumus, F. and Hemmington, N. (1998), Barriers and resistance to change in hotel firms: an investigation at unit level, International Journal of Contemporary Hospitality Management, 10(7): 283-288.

Overholt, M., Kroegor, L. and Prager, K. (1994), Changing people, changing organizations, Information Strategy, The Executive's Journal, 11(1): 38-48.

Peters, J.P. and Tseng, S. (1983), Managing strategic change in hospitals, American Hospital Publishing: Chicago, USA.

Potamianos, J. (1999), While turning around Olympic Airways, how can disgruntled employees be brought along?, Marketing Assignment submitted in partial fulfilment of the part-time degree MBA, University of Lancaster, U. K.

Schurink, E.M. (1990), Participatory action research as a tool for sustainable social development and reconstruction, In: A.S. de Vos (Ed.), Research at grass roots, J. L. van Schaik.: Pretoria, pp. 405-418

Sengupta, K. and Zviran, M. (1997), Measuring user satisfaction in the outsourcing environment, IEEE Transactions on Engineering Management, 44(4): 414-421.

Simpson, S. and Cacioppe, R. (2001), Unwritten ground rules: transforming organization culture to achieve key business objectives and outstanding customer service, Leadership and Organization Development Journal, 22(8): 394-401.

SITA, (2008), Speech: IT Trends in Air Transport", by Francesco Violante, CEO, Brussels – 19 June, 2008.

Slaughter, S. and Ang, S. (1996), Employment outsourcing of information systems, Communications the ACM, 39(7): 47-54.

Steinburg, C. (1992), Taking charge of change, Training and Development, March 1992: 26-32.

Strabel, P. (1996), Why do employees resist change?, Harvard Business Review, 74(3), May-June: 86-92.

Wolcott, H.F. (1994), Transforming qualitative data: Description, analysis, interpretation. Sage: Thousand Oaks, CA.

Maritime security instruments in practice: a critical review of the implementation of ISPS code in the port of Hong Kong

Koi Yu Adolf Ng

Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hong Kong.
Tel: (852) 3400 3625
Fax: (852) 2330 2704
Email: lgtan@polyu.edu.hk

Abstract

The 9/11 terrorist attack has exposed the brittleness of the transportation system which can lead to unprecedented disruption of the global trade system. In responding to such challenge, various security enhancement instruments have been introduced by the international community, notably the ISPS Code. Although various works on maritime security had been undertaken, works dedicated to port security outside developed, western economies, like Asia, remained very scarce, where comprehensive review on how such international guidelines can be applied in a local perspective was found wanting. Hence, focusing the port of Hong Kong, this paper critically reviews how the ISPS Code has been promulgated and implemented in a local perspective. This paper argues that the port of Hong Kong is largely a follower rather than an innovator in complying with the ISPS Code and that port security is perceived as more a problem to solve rather than an opportunity to innovate. This paper can provide valuable insight on the problems, obstacles and solutions when promulgating and implementing maritime security instruments to different global regions.

Keywords: maritime, port security, ISPS Code, Hong Kong

1. Introduction

The outbreak of 9/11 terrorist attack has exposed the potential brittleness of the transportation system. A terrorist event involving the system could lead to unprecedented disruption of the global trade system (Flynn, 2006) which would not only involve human casualties, but also economic, political and social impacts, notably the breakdown of global supply chains and potentially global economic recessions (Greenberg et al., 2006), and it becomes clear that further and, perhaps radical, changes are needed to maximize maritime and supply chain securities in the 21st century (Mensah, 2003).

Being nodal points, port security is arguably pivotal in ensuring the smoothness and efficiency of an increasingly complex intermodal logistical supply chains (Robinson, 2002; Ng, 2007). As defined by Ng and Gujar (2008), port security includes all security and counterterrorism activities which fall within the port domain, including the protection of port facilities, as well as the protection and coordination of security activities when ship and port interact[1]. Although a number of works on maritime security had been undertaken, both academic (for instance, see: Mensah, 2003; Bichou, 2004; King, 2005; Zhu, 2006; Bichou et al., 2007; Talley, 2008) and industrial (for instance, see: OECD, 2003; Greenberg et al., 2006), works dedicated to port security had, so far, remained scarce, or rather being technical in nature (for instance, see: Bichou, 2004; Kumar and Vellenga, 2004), where comprehensive review on how such maritime security instruments can be applied in a local perspective, including obstacles and solutions, is clearly lacking. Even within these few works, attention has often focused on developed, western economies (for instance, see: Ng, 2007; Pallis and Vaggelas; 2007 and 2008; Pinto et al., 2008) where comprehensive analysis on other regions,

[1] Despite the broad definition of port security, since 9/11, much attention had been paid on fighting the threats from terrorist attacks, of which this is also the focus of this paper.

including various globally important economic regions in Asia, remains largely unjustifiably understudied, or simply descriptive rather than analytical in nature (for instance, see: Huxley, 2005; Tan, 2005), possibly with the works of Ng and Gujar (2008) being the only notable exception, thus leaving significant research gaps yet to be filled.

To address such deficiency, through investigating the port of Hong Kong, this paper provides a critical analysis on how the international requirements on port security, as decided by the International Maritime Organisation (IMO), have been imposed in a local perspective. The remaining of this paper is as follows. Sections 2 and 3 will briefly discuss the major international mechanisms initiated by International Maritime Organisation (IMO) related to port security, i.e., the ISPS Code, as well as the research methodology. After then, section 4 will discuss how such guidelines have been implemented in the port of Hong Kong. Before the conclusion in section 6, section 5 will discuss the major challenges that ports are currently facing in addressing port security issues, and how they should improve this situation.

2. The ISPS Code and Port Security

At international level, port security is governed by rules issued by IMO based on the amendments made in December 2002 to the *International Convention for the Safety of Life at Sea* (SOLAS) 1974 as amended, as well as the addition of *Special Measures in Enhancing Maritime Security* (Chapter XI-2) to SOLAS, resulting in the introduction of the *International Ship and Port Facility Security (ISPS) Code*, adopted by IMO on December 2002 and fully implemented on 1 July 2004[2] [3]. Being labelled as the 'comprehensive security regime for international shipping' (Mensah, 2003), deliberate guidance on maritime security, including ports, has been included in the ISPS Code. It is important to note that, however, the code primarily addresses how terrorist attacks can be deterred and minimized, whereas the detailed procedures in addressing the aftermath of a significant security incident, i.e. crisis management, are not mentioned. Indeed, by the time when this paper is written, significant security incidents in ports have yet to take place.

In compliance with the ISPS Code, all ships over 500 gt and port facilities are required to conduct vulnerability assessments and develop security plans to deter potential terrorist attacks e.g. passenger, vehicle and baggage screening procedures, security patrol, the establishment of restricted areas and its execution, procedures for personnel identification, access control, installation of surveillance equipment, etc. The main objectives of the ISPS Code include: (i) detecting security threats; (ii) implementing security measures; (iii) collating and promulgating information related to maritime security; (iv) providing a reliable methodology in assessing maritime security risks; (v) developing detailed security plans and procedures in reacting to changing security levels; and (vi) establishing security-related roles and responsibilities for contracting governments (and their administrations), ship companies and port operators at national and international levels, including the provision of

[2] Despite the international nature of the ISPS Code, however, it was very much an American initiative led by the US Coast Guard, being part of the US government's response to the tragic events of 9/11 with the target of creating a consistent security programme for ships and ports (and their operators and governments) to identify and deter threats from terrorists more effectively.

[3] Apart from international initiatives as mentioned above a number of US-initiated programmes had also been promulgated, many of which have de facto become global port security programmes due to the US's global influences, notably the Container Security Initiative (CSI) and the Custom-Trade Partnership Against Terrorism (C-TPAT), which have been formally codified into law in the US through the Security and Accountability for Every Port Act (SAFE Port Act), adopted in 2006. Apart from the codification of law, the SAFE Port Act also provides further guidance in enhancing port security which is perceived to pose significant global implications, e.g. additional requirements for maritime facilities, transport worker identification credential, port security grants, foreign port assessments, establishment of interagency operational centres for port security, etc. SAFE Port Act was adopted largely in response to the political chaos due to the sale of P&O Ports, including its US port assets, to Dubai Ports World (DPW). The ensuing controversy had led to charges that such purchase could pose a significant national threat. Facing such dilemma, in December 2006, DPW sold its US port assets to AIG. Given the paper's objective, however, this section only reviews the ISPS Code and its impacts on port security.

professional training. Given the fact that ISPS Code was largely a US initiative, it was not surprising to found that the objectives and contents of the ISPS Code are largely equivalent to the US Maritime Transportation Security Act (MTSA) adopted in 2002.

The ISPS Code consists of two main components. Part A provides the minimum mandatory requirements that ships (and their respective companies) and ports (and the Contracting Government) must follow, while Part B provides more detailed, but not compulsory, guidelines and recommendations in the implementation of security assessments and plans. The section outlines of the two parts are largely equivalent, of which Part A mainly illustrates the principles that maritime stakeholders need to follow, while Part B mainly discusses how such principles should/can be put into practice. Within the ISPS Code, three aspects are directly related to port security, namely: (i) changing security levels; (ii) responsibilities of the contracting governments; (iii) port facility security, including the procedures of undertaking Port Facility Security Assessment (PFSA), preparing Port Facility Security Plans (PRSP) and appointing Port Facility Security Officer (PFSO). While the details of the ISPS Code can be found in IMO's website (IMO, 2007), the port security-related aspects are briefly discussed in the following:

- *Changing security levels* - One of the most significant requirements is the introduction of changing security levels. At all times, a security level system (L1, L2 and L3) must be introduced at all ports within the territory of the contracting government, with higher security levels indicate a greater likelihood of occurrence of a security incident, based on an assessment on the degree of credibility, collaboration, specific and imminent nature of the threat information, as well as the potential consequence of such an incident. Similar security level system also exists in ships.

- *Responsibilities of the contracting governments* - Contracting governments should appoint a designated authority (DA) dedicated for port security affairs, while at the same time establish an administrative structure in supporting the DA in carrying out its duties, including local legal backup. In turn, the DA should set security levels in accordance to Part A of the ISPS Code and provide guidance from security incidents taken place in ports, especially necessary and appropriate instructions to affected ships and port facilities in the case of higher security levels (L2 and L3). They are also responsible to approve PFSA reports and PFSP, as well as testing their effectiveness. Finally, contracting governments should also establish the requirements when a Declaration of Security (DoS) is required when ship and port facilities interact.

- *Port Facility Security* - Under the ISPS Code, ports (and their facilities) are required to act in accordance to security levels set by their respective contracting governments, of which the degree of protective measures should be increased with changing security levels in the following security-related issues: performance of security duties, access and monitoring of port facility and restricted areas, supervision of the handling of cargoes and ship's stores and the availability of security communication.

  Apart from daily routine operation, contracting governments (or its designated authorities) must periodically assess port facilities, namely the Port Facility Security Assessment (PFSA), and report the outcomes (or approve the report if done by a separate designated authority). Through an appropriate risk-based methodology, the assessment must at least address the following issues: (i) identification and evaluation of important assets and infrastructure it is important to protect; (ii) identification of possible threats to the assets and infrastructure and the likelihood of their occurrence; (iii) identification, selection and prioritisation of counter measures and procedural changes and their level of effectiveness in reducing vulnerability; and (iv) identification of weaknesses, including human factors in the infrastructure, policies and procedures (IMO, 2002b).

  Finally, based on the assessment outcomes, Port Facility Security Plans (PFSP) should be developed for each port facility, with provisions in addressing the three security levels in the issues including: measures to prevent weapons/dangerous devices from being introduced in the port, authorised access to restricted areas, effective security of cargo and cargo-handling equipment and security of security information; procedures in responding to security threats, new/amended security instructions, evacuation, interfacing with ship security activities, periodic review and updating of PFSP, reporting security incidents, audition of the plan and facilitation of shore leave for ship's personnel or personnel change; as well as identification of port security officer and the duties of security-related personnel. The ISPS Code has noted that one single PFSP in covering more than one port facility is possible, provided that the operator, location, operation, equipment and design of these facilities are highly similar to each other.

To execute the above plans, a Port Facility Security Officer (PFSO) should be appointed for each designated port facility (or one PFSO for multi-facilities if they are largely similar to each other). The PFSO is usually selected by the port facility management, subject to the approval of the contracting government before formally appointed. A PFSO is responsible to ensure that the PRSA exercises and PFSP are well-prepared and being carried out effectively. Apart from routine duties, PFSO also needs to make sure that the facilities concerned are secure through inspection and supervision of facilities, the distribution of responsibilities to his/her subordinates, security-related information gathering, as well as managing the training, drilling and exercises on port facility security. Finally, PFSO also acts as the liaison between the contracting government and the shipping companies, often through the Ship Security Officers (SSO) and Company Security Officers (CSO).

Despite the general consensus that port's security is essential in safeguarding maritime security (Mensah, 2003), the IMO has assumed that the responsibility of port security virtually lied within the hands of the public sector, as reflected in its emphasis on the roles of the contracting governments, where they had the final say in virtually all decisions, e.g., the approval of PFSA and PFSP, the endorsement of PFSO appointment, the right to request DoS, reviewing (parts of) the ship security plan in outstanding circumstances, etc. This implies that non-governmental port stakeholders, including terminal operators, would be largely expected to be followers to international standards and government policies, and would play peripheral roles in the development of port (and its facilities) security issues. In some countries, such shortcomings have been addressed through the formation of committees and working groups in port security. In the US, for example, MTSA required the establishment of Area Maritime Security Committee (AMSC) in throughout all US ports to coordinate the activities of all port stakeholders, including public agencies of different levels, as well as the industry, with specific tasks in collaborating on port security plans, so that resources dedicated for security can be more efficiently utilised. Such emphasis on contracting governments also implies the criticality of training capable manpower in dealing with such new requirements effectively, as pointed out by O'Neil (2003) and Zhu (2006).

3. Research Methodology

Given the study nature, apart from documental review, the author had also conducted various semi-structured, in-depth interviews with key stakeholders who play pivotal roles in carrying out port security measures in the port of Hong Kong, including the Marine Department of the HKSAR Government and port facility operators (hereinafter called 'interviewees'). The objective of such interviews was to identify and obtain information which was otherwise unavailable through published sources. Interviews were mainly conducted at interviewees' respective offices between November 2007 and January 2008.

4. Promulgation and Implementation: Port of Hong Kong

The remainder of this paper will focus on how the ISPS Code has been imposed within the port of Hong Kong. This section is divided into three sub-sections, namely: (i) legal document; (ii) administrative structure; (iii) security levels; (iv) control of ships within/intending to enter the port; and (v) port facility security.

4.1. The Legal Document

In Hong Kong, the main legal document in addressing port security issues is entitled as *An Ordinance to implement the December 2002 amendments to the International Convention for the Safety of Life at Sea (SOLAS), 1974 and the International Ship and Port Facility Security (ISPS) Code and related provisions in the Convention to enhance security of ships and port facilities, and to provide for incidental or related matters*, or the *Merchant Shipping (Security of Ships and Port Facilities) Ordinance* for short title (CAP582, Ordinance No.: 13 of 2004, thereafter termed as 'Ordinance'). Section 6 of the Ordinance is complemented by an empowering document, entitled *Merchant Shipping (Security of Ships and Port Facilities) Rules* (CAP582A, thereafter termed as 'Rules'). Both documents were signed by the then Chief Executive, Mr. Tung Chee-hwa, and enacted by the Legislative Council (Hong Kong's de facto Parliament) in June 2004.

To fulfil its objective of implementing maritime security issues in accordance to SOLAS Chapter XI-2 and the ISPS Code, in explaining its rules, the Ordinance and Rules often make reference to these two documents. For example, in port facility security, the Rules clearly state that a designated port facility shall comply with regulation 10.1 of Chapter XI-2 of the Convention (Section 23), while references to Part A of the ISPS Code in issues related to port facility security had been made four times (Sections 24, 25, 28 and 29). The major difference, however, lies in the fact that the Ordinance and Rules provide much more detailed information and guidance on how SOLAS Chapter XI-2 and the ISPS Code should be put into practice in Hong Kong with, for instance, procedures on how an approval of a PFSP amendment can be withdrawn by the Director of HKMD (e.g. CAP582A, Section 27), the power and limitations of inspections by HKMD personnel (CAP582, Sections 11, 12 and 13), the possible fines and/or penalties if the port facility management fails to comply with the set standards (e.g. CAP582, Section 13; CAP582A, Sections 28 and 30), as well as the port facility management's appeal procedures against any decisions made by the Director of HKMD (e.g. CA582A, Section 31)

4.2. Administrative Structure

The security administration structure of the port of Hong Kong can be found in Figure 1.

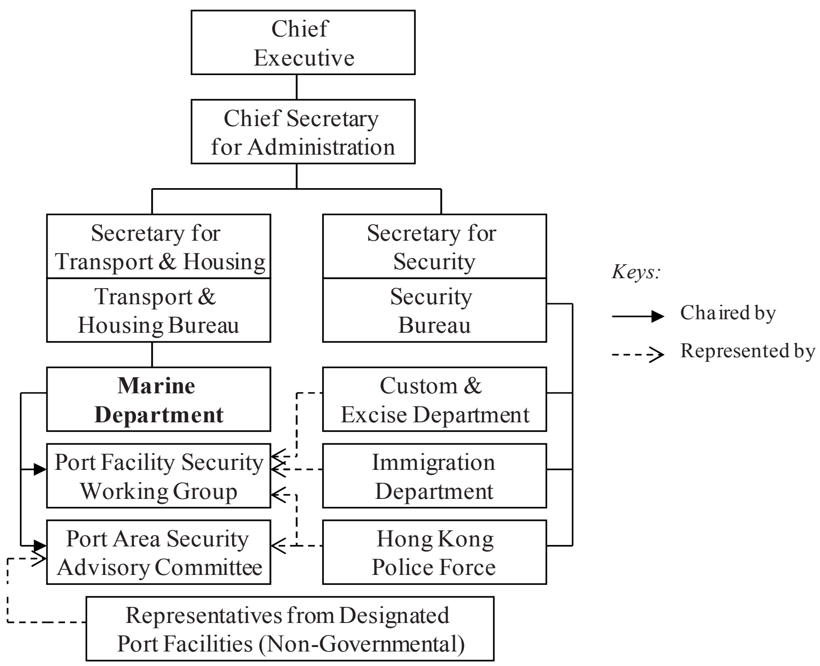


Figure 1: Port security administration structure in Hong Kong in 2008

The Hong Kong Marine Department (HKMD), subordinated to the Transport and Housing Bureau, is the DA for the contracting government, i.e. Government of the Hong Kong Special Administrative Region (HKSAR), in discharging port security duties in accordance to the mandatory requirements of the ISPS Code. According to the Ordinance and Rules, HKMD's Director (hereafter termed as 'Director') may specify the extent of application of SOLAS Chapter XI-2 and the ISPS Code in relation to any designated port facility (Section 5, CAP582), designating security organisations in executing certain port security duties, as long as such organisations possess the appropriate expertise knowledge and not in non-compliance of the Section 4.3, Part A of the ISPS Code, authorisation of officers (Section 9, CAP582) and granting exemptions from the provision of the Ordinance (Section 14, CAP582).

Under HKMD, an advisory, non-statutory committee had been established in June 2003, namely the Port Area Security Advisory Committee (PASAC). The function of PASAC is to advise to the HKSAR government and its designated authority, i.e. HKMD on all matters in connection with the implementation of SOLAS Chapter XI-2 and the ISPS Code in Hong Kong (PASAC, 2003a), as well as to monitor its application in Hong Kong (PASAC, 2004a). In terms of membership, the committee consists of about 20 members chaired by HKMD's Deputy Director, comprised of governmental representatives from HKMD and the Hong Kong Police Force, as well as non-governmental representatives from the designated port facilities Each facility group, e.g. container terminal, bulk terminal, dockyards, etc., will nominate one representative to sit in the committee (PASAC, 2003a). Until November 2007, nine PASAC committee meetings had been conducted. The primary focus of PASAC was on port security (not ship security) while, in some cases, matters related to ship-port interface would also be covered (PASAC, 2003a). The composition of PASAC can be found in Table 1.

Table 1: The composition of PASAC in 2008

Position or Representative	Number
Chairman (Deputy Director of Marine)	1
Secretary (Marine Officer/Port Security Administration)	1
Hong Kong Marine Department (HKMD)	5
Hong Kong Police Force	2
Container Terminal Operators	2
Oil Terminal Operators	2
River Trade Terminal Operators	1
Ship Repairs Industry	2
Cruise Industry	1
Bulk Industry	1
Hong Kong Liner Shipping Association	1
Total	**19**

Source: HKMD website

On July 2003, the Port Facility Security Working Group (PFSWG) was established, chaired by HKMD and represented by the Custom and Excise Department, Immigration Department and the Hong Kong Police Force. PFSWG acts as the executive arm in discharging the obliged duties in sustaining the security of the Port of Hong Kong. The working group is also responsible to evaluate of PFSAs and PFSPs undertaken and prepared by facility operators, before submitting them to HKMD for final approval.

Any port security issues, like new requests from IMO, will be discussed within PFSWG concerning its implications and practicality in Hong Kong and, if found necessary, will be bring up to the agenda of the next PASAC meeting. In most cases, any new amendments, including the Ordinance, the Rules and the details of implementing the articles in IMO's Maritime Safety Committee (MSC) Circulars, will be firstly discussed and compromised within PASAC. According to internal information, although non-statutory in nature, HKMD will always ensure that any new policies would have obtained the endorsement of PASAC before implementation.

In terms of finance, neither the HKMD nor the HKSAR government prepare a budget related to port security, and the income received from the issuance of security certificates and audit exercises are too trivial to cover the administration costs[4]. Also, the companies which own and operate their respective designated port facilities are responsible for all the financial costs in the execution of their respective PFSA, the preparation of PFSP and the actions. During the second PASAC meeting, the chairman had made clear to the facility operators that the government would not subsidise, or provide any loans, to any port security-related projects (PASAC, 2003b).

In general, HKMD is responsible to execute its security obligations in three major categories, namely: (i) setting the security levels; (ii) control of ships within/intending to enter the port; and (iii) port facility security.

4.3. Security Levels

A security level system, L1, L2 and L3, has been introduced, of which the updated status is live on internet, accessible at the HKMD official website 24 hours per day. As illustrated in Figure 2, the definitions of different security levels are equivalent to the guidelines found in Section 1.8, Part B of the ISPS Code.

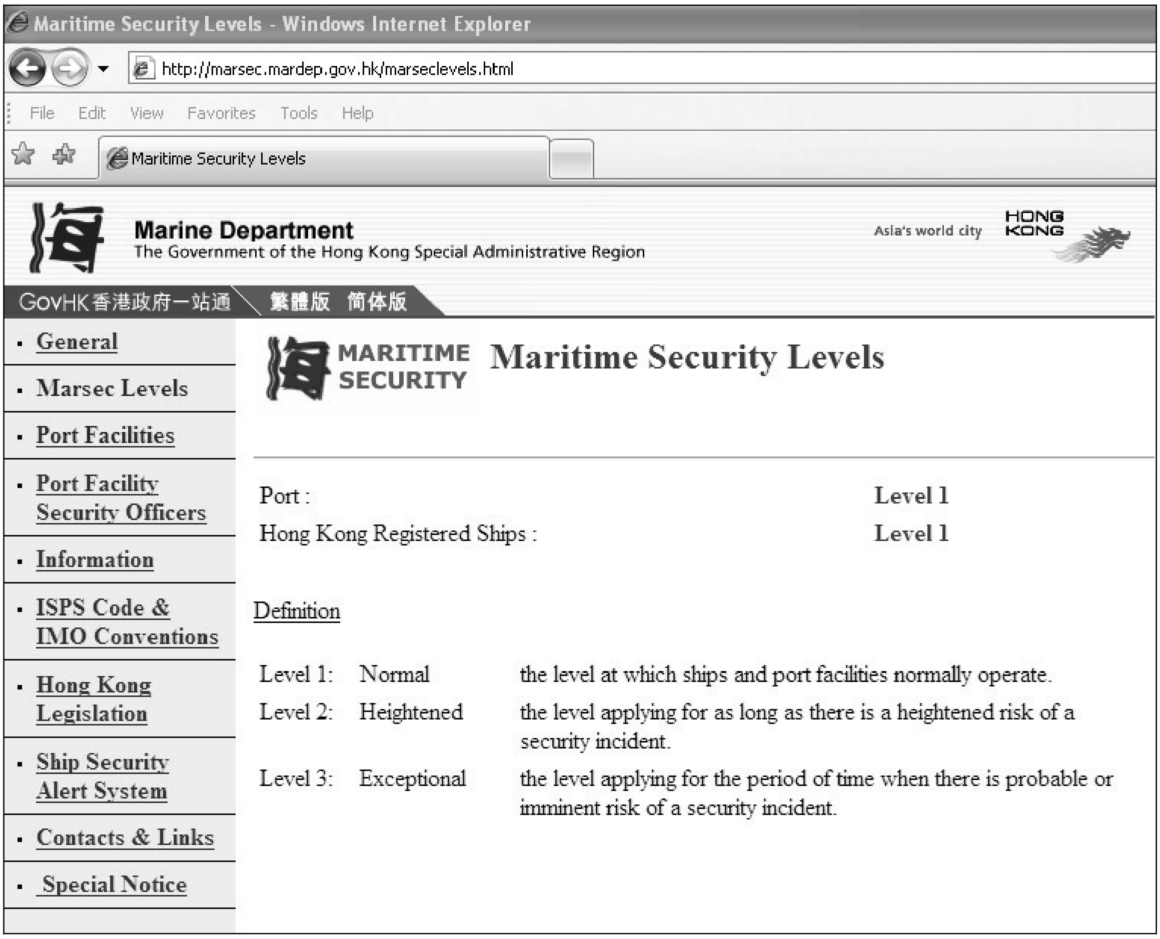


Figure 2: An illustration on different security levels in the port of Hong Kong
Source: HKMD website

[4] According to Section 33 of the Rules, HKMD can charge an hourly rate of HKD 1,115-3,270 for services including issuing/endorsing (interim) security certificate, approving PFSP, designated port facility inspections. However, during the fifth PASAC meeting, the chairman had already indicated to committee members that the government had no intention to shift the financial burden of regular security audit exercises to facility operators (PASAC, 2004b).

All information and intelligence related to port security, which can possibly lead to changes in security levels, are provided by the Intelligence Unit of the Hong Kong Police Force. The Police Force will first assess the credibility and potential consequences of the intelligence, before advising HKMD on the necessity to change the security level, of which the website will be updated if a change is confirmed by HKMD[5]. There is a general understanding that the government would instruct a facility operator to close down its facility only when the security level changes to L3, although the ultimate sanction should lie with the DA (PASAC, 2003b).

4.4. Control of Ships within/intending to Enter the Port

The HKSAR government strictly follows the mandatory requirements of international documents in controlling ships within or intending to enter the port of Hong Kong. For example, in the Rules, all the Sections which are related to this issue (Sections 11 and 12, CAP582A) have made full reference to the requirements as indicated in SOLAS Chapter XI-2 (Regulation 9). Any additional regulations on this issue are virtually non-existent. On the other hand, all necessary information and guidelines for ships within/intending to enter the port, including HKMD notices and information notes, pre-arrival security information, DoS and security advice to Hong Kong-registered ships, are easily accessible and downloadable from the internet, through HKMD's website.

4.5. Port Facility Security

Given the traditional port policy of Hong Kong which emphasised on active non-intervention by the public sector (the so-called *laissez-faire* policy), it is not surprising to found that all but three of the designated port facilities are privately owned and operated (the exception being China and Macau Ferry Terminals and Buoys and Anchorage Services, which are operated by HKMD) and the PFSAs and PFSPs are also carried out by these companies (or any recognised security organisation (RSO) chosen by them) themselves, while HKMD takes up the responsibility in undertaking and preparing PFSAs and PFSPs respectively for China and Macau Ferry Terminals as well as Buoys and Anchorage Services. PFSAs and PFSPs will be submitted to PFSWG for evaluation and vetting, before being recommended to HKMD for final approval (Figure 3). Until 2008, HKMD has reviewed and approved 33 PFSPs, consisting of container and ferry terminals, wharfs and dockyards, oil jetty and terminals, power stations, fuel receipt facilities and mooring buoys and anchorages. See Appendix.

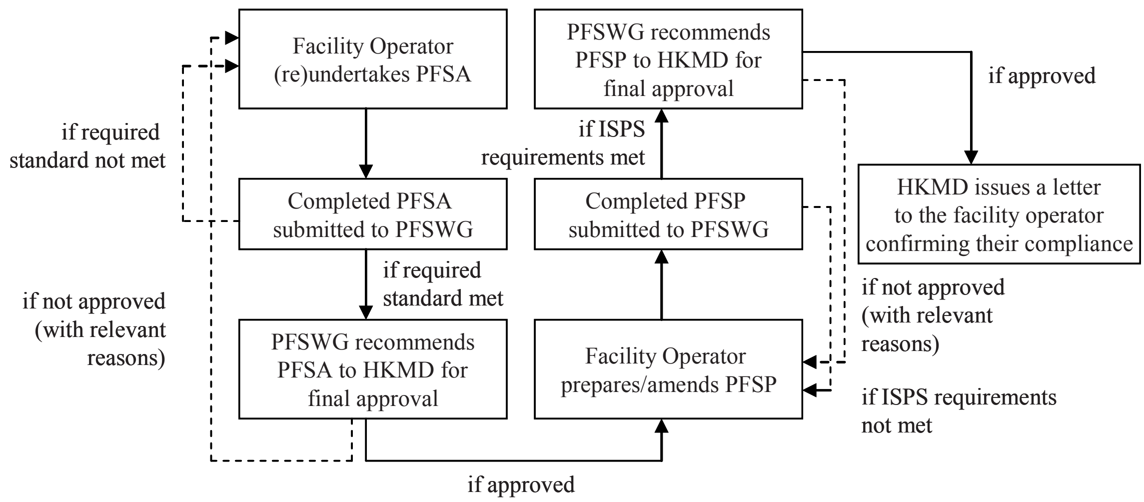


Figure 3: Procedures of approving PFSA and PFSP in the port of Hong Kong
Source: Derived from PASAC (2003a)

[5] During the third PASAC meeting, an issue was raised concerning the transmission of threat assessment since the ISPS security levels did not match the conventional security levels used by the Police Force. An *ad hoc* meeting for this issue was conducted at February 2004 between the parties concerned, and had been resolved before the ISPS Code was fully implemented at 1 July 2004.

In compliance to Section 16.8, Part A of the ISPS Code, some companies, notably Hongkong International Terminals (HIT), Modern Terminals Ltd. (MTL) and ExxonMobil (Hong Kong) Ltd., have chosen to prepare a single PFSP for all the terminals that they own and operate[6]. Subsequently, these companies have also chosen to make single PFSO appointment for their respective terminals, as in compliance to Section 17.1, Part A of the ISPS Code. The selection and appointment of PFSOs is decided by the port facility management subject to the formal approval by HKMD. By 2007, 24 PFSOs have been appointed, either as a dedicated position, or undertaken by safety/operation-related managers. HKMD has also ensured that their names and contacts are easily accessible from the internet.

All PFSOs must have received training and certification from a HKMD-accredited local port security programme (or has attended a similar programme overseas, of which verification will be decided on a case-by-case basis) (HKMD, 2007a). In security personnel training and certification, in accordance to Section 4.3, Part A of the ISPS Code, such responsibilities have been fully outsourced to recognised security organisations (RSO), as long as the institution concerned has submitted a proposal to HKMD outlining a programme which fulfils the prerequisites laid down in IMO's Maritime Safety Committee (MSC) Circular No. 1188 (IMO, 2007) and the *Guidelines for Approving PFSO Training Course* (HKMD, 2007b) and a formal accreditation process undertaken by designated HKMD officials[7]. By September 2007, HKMD has approved two maritime institutions in offering security training and certification programmes, while verification of certificates issued by overseas institutions is decided on a case-by-case basis. An example of the contents of a PFSO programme accredited by HKMD can be found in Table 2. The validity of the PFSO qualification is five years, and is renewable subject to the criteria that the personnel concerned had served at least 12 months as (Deputy) PFSO within the validity period (HKMD, 2007a).

Table 2: The modules of a HKMD-accredited PFSO programme in Hong Kong

Module 1	Introduction
Module 2	Maritime Security Policy
Module 3	Security Responsibilities
Module 4	Port Facility Security Assessment
Module 5	Security Equipment
Module 6	Port Facility Security Plan
Module 7	Threat Identification
Module 8	Port Facility Security Actions
Module 9	Emergency Preparedness
Module 10	Security Administration
Module 11	Security Training

Source: Institute of Seatransport (2007)

The major mechanism in auditing the PFSP is through (notified in advance) site visits to the designated port facilities[8], which would tie in with the validity of the Statement of Compliance (SoC) issued (PASAC, 2004b). Auditing is divided into 'full' and 'partial' audits, of which they have to be undertaken at a five- and one-year interval respectively. The auditing schedule and arrangement with the designated security facility management team are arranged by a designated officer from HKMD,

[6] In practice, however, a single security certificate has been issued to each designated facility, so as to ensure that other facilities can still operate normally even when one or more facilities have to close down due to security threats and/or incidents. For example, three security certificates have been issued to MTL's container terminals (DA01, DA02 and DA03, see Appendix), but they are covered and managed by one PFSP and PFSO respectively.

[7] According to the HKMD's *Guidelines for Approving Port Facility Security Officer Training Course*, formal approval to the institution concerned in providing SFSO training and certification would be granted only after the first course of the programme concerned has been monitored and assessed by HKMD officials and the officer(s) concerned, with positive feedbacks.

[8] According to internal information, un-notified inspections cannot be undertaken due to the shortage of financial support from the HKSAR government.

of which the audit team also consists of representatives from the Police Force, Custom and Excise and Immigration Departments[9].

Audit categories are divided into seven areas, namely: (i) documentation; (ii) access control; (iii) handling of cargoes; (iv) port-ship interface; (v) control of restricted area; (vi) awareness; and (vii) security infrastructure. All areas will be examined during a full audit, while four of them will be selected by the audit team leader (usually HKMD's designated marine officer) for each partial audit. According to HKMD, investigating the physical condition of the designated facilities is the most important function during the site visit, of which evaluation results, recommendations and mandatory actions will be laid down in a confidential audit report. In auditing the 'soft' aspects, like personnel arrangement and documentation, the designated facility management needs to fill in a dedicated questionnaire prepared by HKMD.

Minor deficiencies which are unlikely to seriously threaten the designated facilities in complying with SOLAS Chapter XI-2 (like worn-out fencing and non-adequate lighting), Part A of the ISPS Code, the Ordinance and the Rules, will not affect the endorsement of the validity of the security certificate. However, during the site visit, the facility management needs to provide a binding promise to the audit team on when they can rectify the problem(s), and defected facilities will be re-inspected during the next auditing exercise, of which HKMD would void the validity of the facility's security certificate if the non-compliance persists (PASAC, 2004b).

5. Discussions

From the above analysis, it is recognised that several characteristics existed. On a positive note, the port of Hong Kong has mostly, if not fully, fulfilled the mandatory requirements as laid down in SOLAS Chapter XI-2 and Part A of the ISPS Code. Virtually all the core elements of the international mandatory requirements have been addressed, while necessary security information to maritime stakeholders and the public are easily accessible. The HKSAR government is also able to provide a well-supported legal and structural backup in facilitating the implementation of international requirements in Hong Kong and the basic mechanisms in complying with the international requirements are in place. Facility operators are also, in general, quite cooperative with the designated authority in complying with the mandatory requirements from SOLAS Chapter XI-2 and the ISPS Code.

Despite such effectiveness, however, constructive innovation in the implementation of port security issues is rather limited. The local legal documents, i.e. the Ordinance and Rules, had been made as simple as possible, with all the core sections actually confirming that the necessity of implementing the international prerequisites in the port of Hong Kong. Additional security requirements and measures are virtually non-existent, not helped by some local situations which have practically disabled Hong Kong to carry out security-related innovative activities. For example, until now, the port of Hong Kong is still unable to introduce biometric identity system on port workers (which had been carried out in many US and some European ports) because Hong Kong does not have any significant labour unions, while it is not compulsory for workers to join/register for any labour unions. Another example lies in the difference in legal system between Hong Kong and the US. Under Hong Kong's legislation, the designated authority, i.e. HKMD, is not empowered to shut down any port facilities directly, but through providing directions to the non-complying facility to rectify the deficiencies, and even if this is not followed, HKMD could only shut down the facility through withdrawing security certificate and report it to IMO (PASAC, 2006). This implies that the port of Hong Kong is potentially less immediate in reacting to extraordinary, and requiring immediate actions,

[9] According to interviewees, while the Police Force will always send representatives to the audit team, Custom and Excise and Immigration Departments will only send representatives to selected designated facilities of which they are interested. Generally speaking, Custom and Excise Department is only interested in cargo-related facilities, while Immigration Department is only interested in passenger-related facilities.

security incidents like L3, of which in this situation the occurrence of security incident is likely to be imminent.

Furthermore, while fully acknowledging the public nature of port security, at the same time, it is quite clear that the HKSAR government is trying hard to keep port security issues parallel to the city's traditional *laissez-faire* policies in port operation and governance. Apart from the compulsory obligations as laid down in Section 4.3, Part A of the ISPS Code, nearly all the optional responsibilities, including the execution of PFSA (Section 15.2), preparation of PFSP (Section 16.1.1), appointment of PFSO (Section 17.1) and training and certification (Section 18), have been outsourced to RSOs through legislations (like Sections 25 and 26 of CAP582A) and practical means, while the government also resists to recommend any RSOs and thus operators are completely free to choose their own RSOs (PASAC, 2003a). Moreover, as mentioned, the government insists on its non-subsidising policy to any security related projects for any designated facilities (PASAC, 2003b) and does not even allocate any significant financial resources in carrying out port security. Thus, the extent of which port security measures can be implemented in the designated facilities very much depends on the attitude of facility operators.

Given the fact that Hong Kong has yet experienced any changes in its security level from L1 or experiences any significant security incidents (until the end of 2007), Hong Kong is largely considered as a low-risk port with little chance from terrorist attack (PASAC, 2003b). It is therefore not surprising that both the government and designated facility operators are also not very enthusiastic in the idea of investing heavily in security-related projects other than fulfilling the basic mandatory requirements[10]. For example, according to a HKMD's senior marine officer who is actively involved in port security, during the discussion of implementing any new security requirements (either from IMO or the HKSAR government), the core discussion point between PASAC often lies in the financial obligations that facility operators need to devote, where significant gap often exists between public and private expectations. The unwillingness of the government to significantly finance the issue has also further added to the difficulty in becoming an innovator in port security, as exemplified by the fact that HKMD does not even possess the necessary resources in carrying out any additional (un-notified in advance) facility inspections other than routine annual audits, not to mention any potential opportunities for research and development. Indeed, the experience of Hong Kong in complying with security instruments is not completely dissimilar to many other global regions, like Asia and even the European Union, of which stakeholders often feel discontent with the imposition of further rules based on security issues (for instance, see: Ng and Gujar, 2008; Pallis and Vaggelas, 2008).

6. Conclusions

The outbreak of 9/11 terrorist attack has exposed the potential brittleness of the transportation system which can lead to unprecedented disruption of the global trade system. In responding to such challenge, various security enhancement instruments have been introduced by the international community, notably the ISPS Code. Although a number of works on maritime security had been carried out, works dedicated to port security, including globally important economic regions in Asia, remained scarce, where comprehensive review on how such international guidelines can be applied in a local perspective is clearly lacking. Understanding this, through studying the port of Hong Kong, this paper critically reviews on how the international requirements, i.e., the ISPS Code, have been implemented on a local region.

Based on this paper's analysis, it is found that Hong Kong is largely a 'follower', rather than an 'innovator', in dealing with port security issues. The port security administrative structure is

[10] According to anecdotal information from interviewees, the fact that Hong Kong is part of China, of which China, in general, maintains rather friendly relationship with most Middle Eastern countries/regions, has also strengthened the 'safe image' of the port of Hong Kong, which has further discouraged any additional financial incentives to enhance security in designated port facilities.

fundamentally a designated authority purely for the implementation of SOLAS Chapter XI-2 and the ISPS Code with virtually no innovation at all. Also, from the above analysis, it seems that security issue is not widely regarded as an important issue in port operation in Hong Kong. Indeed, such perception is reflected by the fact that, as confirmed by various interviewees, in many port facility operating companies, the PFSO (or security manager) is often a rather junior position within the company, while the government, as mentioned, is quite reluctant to input resources of any significance into addressing the issue. Contrary to the major American and some Western European ports (notably Rotterdam), a 'security culture' has yet to establish in the port of Hong Kong. Indeed, from author's self-observation, the core rationale of compliance by Hong Kong maritime stakeholders seems to be avoiding potential economic loss due to non-compliance (like losing the American market), rather than appreciating the concept of 'more secured port'. In other words, port security is more a problem to solve rather than an opportunity to innovate. Indeed, the port of Hong Kong can partly reflect the situation of Asian ports, where the approach of carrying out port security measures is rather half-hearted, pro-trade and economically driven (Ng and Gujar, 2008).

Conclusively speaking, the case studied in this paper illustrates that rules and standards may not be completely effective, where local circumstances and other software aspects (like attitudes and governance system) should not be overlooked if port (and indeed maritime and supply chain) security can be carried out effectively. Further research is required to investigate how such obstacles can be effectively overcome. Last but not least, by undertaking a detailed investigation on the imposition of port security measures in a local perspective, this paper has provided valuable insight on the problems, obstacles and solutions in applying maritime security measures to different global regions, as well as a decent platform for further research on this increasingly important, but understudied, topic.

7. Acknowledgements

The author would like to thank all the interviewees and anonymous reviewers who have provided useful information and comments to this paper. All errors and omissions remain the sole responsibility of the author.

References

APEC (2002), APEC 2002 Leader's Declaration, The Tenth APEC Economic Leaders' Meeting, Los Cabos, Mexico, 27 October, accessible at: http://www.apec.org/apec/leaders__declarations/2002.html.

APEC's official website: http://www.apec.org, last accessed on February 2008.

Asia Pacific Foundation of Canada (2004), New security measures challenge competitiveness of Asian ports, Asia Pacific Bulletin, 30 April, accessible at: http://www.asiapacificbusiness.ca.

Bichou, K. (2004), The ISPS Code and the cost of port compliance: an initial logistics and supply chain framework for port security assessment and management, Maritime Economics and Logistics 6: 322 – 348.

Bichou, K., Bell, M.G.H., and Evans, A. (eds.) (2007), Risk Management in Port Operations, Logistics and Supply Chain Security, London: Informa LLP.

Flynn, S.E. (2006), Port security is still a house of cards. Far East Economic Review, Jan/Feb 2006, accessible at: http://www.feer.com/articles1/2006/0601/free/p005.html.

Greenberg, M.D., Chalk, P., Willis, H.H., Khiko, I. and Ortiz, D.S. (2006), Maritime Terrorism: Risk and Liability. Santa Monica: RAND Corporation.

HKMD website: http://www.mardep.gov.hk, last accessed on March 2009.

HKMD (2007a): Guidelines for Application of Qualification Recognition as Port Facility Security Officer. Accessible at: http://marsec.mardep.gov.hk/pfso_training.html.

HKMD (2007b): Guidelines for Approving Port Facility Security Officer Training Course. Accessible at: http://marsec.mardep.gov.hk/pfso_training.html.

Hong Kong Institute of Seatransport (2007): Port Facility Security Officer Course. Hong Kong: Institute of Seatransport.

Huxley, T. (2005), Southeast Asia 2004 stable, but facing major security challenges, In: K.W. Chin and D. Singh (eds.): Southeast Asian Affairs 2005, Singapore: Institute of Southeast Asian Studies, pp. 3-23.

IDSS (2004), Maritime Security in the Asia-Pacific: Report of a Conference Organized by the Institute of Defence and Strategic Studies (IDSS), held in Singapore, 20-21 May.

IMO (2002a), Amendments to the Annex to the International Convention for the Safety of Life at Sea, 1974 as Amended. London: International Maritime Organisation, SOLAS/CONF.5/32, December 2002.

IMO (2002b), International Code for the Security of Ships and Port Facilities (ISPS Code). London: International Maritime Organisation, SOLAS/CONF.5/34, December.

King, J. (2005), The security of merchant shipping, Marine Policy 29: 235-245.

Kumar, S.H. and Vellenga, D. (2004), Port security costs in the US: a public policy dilemma, Proceedings of the Annual Conference of the International Association of Maritime Economists, Izmir, Turkey, July.

Mensah, T.A. (2003), The place of the ISPS Code in the legal international regime for the security of international shipping, WMU Journal of Maritime Affairs, 3(1): 17-30.

Ng, K.Y.A. (2007), Port security and the competitiveness of short sea shipping in Europe: implications and challenges. In: Bichou, K., Bell, M. and Evans, A. (eds.): Risk Management in Port Operations, Logistics and Supply Chain Security. London: Lloyd's of London Press and Informa, pp. 347-366.

Ng, K.Y.A. and Gujar, G. (2008), Port security in Asia, In: W.K. Talley (ed.): Maritime Safety, Security and Piracy. Informa LLP, London, pp. 257-278.

O'Neil, W.A. (2003), The human element in shipping, WMU Journal of Maritime Affairs, 2(2): 95-97.

OECD (2003), Security in Maritime Transport: Risk Factors and Economic Impact, Paris: OECD Maritime Transport Committee.

Pallis, A.A. and Vaggelas, G.K. (2007), Enhancing port security via the enactment of EU policies, In: Bichou, K., Bell, M. and Evans, A. (eds.): Risk Management in Port Operations, Logistics and Supply Chain Security. London: Lloyd's of London Press and Informa, pp. 303-334.

Pallis, A.A. and Vaggelas, G.K. (2008), EU port and shipping security, In: Talley, W.K. (ed.): Maritime Safety, Security and Piracy. Informa LLP, London, pp. 235-255.

PASAC (2003a), Minutes of the First Meeting of the Port Area Security Advisory Committee. Hong Kong: Hong Kong Marine Department (July 2003).

PASAC (2003b), Minutes of the Second Meeting of the Port Area Security Advisory Committee. Hong Kong: Hong Kong Marine Department (September 2003).

PASAC (2004a), Minutes of the Fourth Meeting of the Port Area Security Advisory Committee. Hong Kong: Hong Kong Marine Department (May 2004).

PASAC (2004b), Minutes of the Fifth Meeting of the Port Area Security Advisory Committee. Hong Kong: Hong Kong Marine Department (September 2004).

PASAC (2006), Minutes of the Eighth Meeting of the Port Area Security Advisory Committee. Hong Kong: Hong Kong Marine Department (October 2006).

Pilling, D. and Mitchell, T. (2006), US official urges Asia to improve port security, Financial Times, 28 March.

Pinto, C.A., Rabadi, G. and Talley, W.K. (2008), US port security, In: Talley, W.K. (ed.): Maritime Safety, Security and Piracy. Informa LLP, London, pp. 217-233.

Robinson, R. (2002), Ports as elements in value-driven chain systems: the new paradigm, Maritime Policy and Management, 29(3): 241-255.

Tan, A.T.H. (2005), Singapore's approach to homeland security, In: K.W. Chin and D. Singh (eds.): Southeast Asian Affairs 2005. Singapore: ISAS, pp. 329-362.

W.K. Talley (ed.) (2008), Maritime Safety, Security and Piracy, London: Informa LLP.

Zhu, J. (2006), Asia and IMO technical cooperation. Ocean and Coastal Management 49: 627-636.

Appendix: Designated port facilities of which PFSPs have been reviewed and approved by Marine Department, HKSAR Government

PFSP Approved:*

DA01	Container Terminal 1, Modern Terminals Ltd.
DA02	Container Terminal 2, Modern Terminals Ltd.
DA03	Container Terminal 5, Modern Terminals Ltd.
DA05	Container Terminal 4, Hongkong International Terminals
DA06	Container Terminal 6, Hongkong International Terminals
DA07	Container Terminal 7, Hongkong International Terminals
DA08	Container Terminal 9 (North), Hongkong International Terminals
DA09	Container Terminal 8 (East), COSCO-HIT Terminals (Hong Kong) Ltd.
DA10	Container Terminal 3, CSX World Terminals Hong Kong Ltd.
DA11	Container Terminal 8 (West), Asia Container Terminals Ltd.
DA12	Ocean Terminal, Harbour City Estates Ltd.
DA13	Hongkong United Dockyards, Hongkong United Dockyards Ltd.
DA14	Lok On Pai Oil Jetty, Hong Kong Petrochemical Company Ltd.
DA15	Shiu Wing Steel Wharf, Shiu Wing Steel Ltd.
DA16	Castle Peak Power Station Coal Unloading Jetty, Castle Peak Power Co. Ltd.
DA17	Green Island Cement Wharf, Green Island Cement Company Ltd.
DA18	ExxonMobil Oil Terminal East, ExxonMobil Hong Kong Ltd.
DA19	ExxonMobil Oil Terminal West, ExxonMobil Hong Kong Ltd.
DA20	Aviation Fuel Receipt Facility, AFSC Operations Ltd.
DA21	Lamma Power Station Coal Unloading Jetty, Hongkong Electric Co. Ltd.
DA22	China Ferry Terminal, Marine Department, HKSAR Government
DA23	Macau Ferry Terminal, Marine Department, HKSAR Government
DA24	Container Terminal 9 (South), Modern Terminals Ltd.
DA25	Government Mooring Buoys & Anchorages, Marine Department, HKSAR Government
DA26	River Trade Terminal, River Trade Terminal Company Ltd.
DA27	Towngas Wharf - Tolo Harbour, The Hong Kong and China Gas Co. Ltd.
DA28	Chevron Oil Terminal, Chevron Companies (Greater China) Ltd.
DA29	Shell Oil Terminal, Shell Hong Kong Ltd.
DA30	CRC Oil Terminal, China Resources Petrochems (Group) Co. Ltd.
DA31	Euroasia Dockyard, China Merchants Container Services Ltd.
DA32	Yiu Lian Dockyards, Yiu Lian Dockyards Ltd.
DA33	China Merchant-Wharf, China Merchant Godown, Wharf & Transportation Co. Ltd

* DA04 (MTL Terminal 8 (West)) was cancelled because MTL had transferred the terminal's ownership to ACT and thus had been inscribed into ACT Terminal 8 (West), i.e., DA11

Source: HKMD website

Maritime policy using multilayer CBR

Prof. Nikitas Nikitakos and Georgios Fikaris

Department of Shipping and Transport, University of the Aegean, Chios, 82100, Greece.

Abstract

Today's maritime industry can be described as an extremely important and competitive sector of the transport industry which in turn is a vital function of the world's international commerce. Major players in the maritime arena have to cope with a variety of problems that require effective action in usually strict deadlines. Formation of maritime policy is a complicated and difficult task due to a number of factors such as the global character of shipping, the multiple bodies that produce policy, the tremendous amount of conventions, laws, regulations, directives, circulars, etc, that already exist, plus the number of stakeholders affected by the policy whose influence must be taken into account by any wise policy maker.
These observations served as a motivation to use CBR (Case Based Reasoning) for the development of a hybrid system in policy making. CBR tools are used in a variety of domains such as defense, economy, education, medicine, etc, but little has been done for the maritime industry and especially in policy making. Since policies have a top – down influence and differ significantly in respect of the implementation level, CBR was selectively used in layers in order to have a better vertical penetration in the problem's structure.

Keywords: Maritime policy, Case Based Reasoning, Recruitment

1. Introduction

The adoption of integrated maritime policies on behalf of the EU made policy formation throughout the domain much harder, since policies concerning entirely segregated or mostly unrelated maritime sectors with conflicting interests must be proven compatible before implementation. Domains like transport, energy, environment, fisheries, employment, etc, must find common grounds and develop policies that minimize trade – offs and maximize the continent's competitive advantage. The focal points of this new approach are the development of a thriving maritime economy in an environmentally sustainable manner aided by excellence in maritime research, technology and innovation (Blue Book, *Vision for a European Maritime Policy*, 2007).

A successful maritime policy must incorporate a number of features in order to overcome the difficulties, satisfy the demands and being able to achieve the objectives set by an emerging integrated framework (Psaraftis, 2002). Such features are a) the ability to set proper, feasible and measurable goals taking into account the necessary stakeholders, b) identify the decisions to be taken and the available means associated with them in order to achieve the goals, c) examine the decision situation and estimate the possibility of a successful implementation of the policy at hand, d) properly measure the resources available and the interfering potential of the policy maker, e) correctly assess the expected results and the consequences of those decisions to neighboring domains and finally f) investigate for any conflicts with pre existing policies (see also UK government, Ministry of Health, *14 criteria for excellence in policy making*).

These observations served as a motivation for the development of a hybrid CBR system able to recommend solutions and courses of action for problems of maritime interest and especially policies. CBR tools are used in a variety of domains such as defense, economy, education, medicine, etc (Kolodner, 1993), but little has been done for the maritime industry and especially in policy making.

Since policies have a top – down influence and differ significantly in respect of the implementation level (Alter, 1980), CBR was used in layers in order to have a better penetration to the problem's structure. For simplicity reasons only the first policy layer (strategic) will be presented hereby, since lower layers are applied similarly.

It must be stressed that the system is still under development so this essay is mainly conceptual aiming to describe the system's basics and possible approaches about procedures not yet fully developed i.e. adaptation, storage, implementation and monitoring. The rest of the paper is organized as follows: Section 2 examines the compatibility of applying CBR for policy making and reviews some relevant work while section 3 presents the stakeholder identification and goal selecting procedure. Section 4 discusses the case representation and indexing while section 5 is dedicated to the retrieval, adaptation and storage. Section 6 deals with the policy's rationale. Finally, section 7 is about conclusions and suggestions for future work.

2. Literature Review and Related Work

A widely accepted definition for policy states that policy is "a set of interrelated decisions taken by a political actor or group of actors concerning the selection of goals and the means of achieving them within a specified situation where those decisions should, in principle, be within the power of those actors to achieve" (Jenkins, 1978). By this definition, it is easily derived that a policy intervention occurs when a political actor wishes to solve a problem taking place in a given present or future environment with limited resources as well as own influencing power. The political actor's goals express the problem since they represent the difference between the desired and current status of the world (Locke, 2001) while the available and usable means affecting those goals can lead towards a possible solution path provided they are transformed to a set of decisions for actions. This set of decision can eventually evolve to a solution – policy if this set of actions is properly defined and proven optimum, compared to other sets leading to a similarly effective result (Sheffrin, 2003).

It is also fair to assume that if this political actor had dealt successfully before with the same problem in a similar situation he would probably have used the same solution – policy to solve the current problem provided he was certain for the optimality and outcome of his choice. This practice is closely related to human problem solving and reasoning which is considered the principal concept of CBR. Case based reasoning systems solve new problems by adapting solutions to older problems (Kolodner, 1993). A central part of CBR is the "case" defined as a contextualized piece of knowledge representing an experience that teaches a lesson fundamental to achieving the goals of the reasoner - system (Leake, 1996).

The system described here is related to the work of Legna (2000), Kljajic et al (2002), who developed a CBR system for recommendations concerning strategic decisions in multiple domains, geographically located in the Canary Islands (Legna and Gonzalez, 2005). The Canary System had its roots to Simon's paper about management decision making published in 1977 that described a computer model – based system able to assist strategic decisions consisting of 3 crucial phases labelled "Intelligence", "Design" and "Choice". Legna and Kljajic added the phases of "Implementation" and "Monitoring" and used CBR in order to overcome quantitative difficulties and uncertainty over relations between variables.

Objective selection, understanding of regional structure, future prediction and problem identification are the four sub phases comprising the intelligence phase of the Canary system. A point to remember is the "Understanding of regional structure" sub phase because the authors used an initial model describing the overall societal structure in broad lines, while each separate activity was also analyzed through a more detailed sub model that simulated its behaviour, identified the role of the variables and detected the direct and indirect relations between them (Kljajic et al., 2003, 2003b). The application of system dynamics to the model led to the identification of enhancing and balancing loops in the structure as well as the determination of the principal trends and forces that guide it. Another important issue is the construction of scenarios in order to estimate the future situation. Those scenarios are based on estimations about the variables and their relations identified previously. The time horizon and the predictability potential for each variable are considered important for the credibility of each scenario.

There are several differences between this approach and the work of Legna and Kljajic. The stakeholder and goal identification processes were done using formal tools instead of brainstorming sessions. A model is used for the identification of the problem and solution variables, but those variables are examined in a dynamic way (time dimension) by exploiting the particularity of the maritime industry to reduce the computation load. This means that the situation and user similarity are used as primary indexes since the maritime historical data can be clustered in time periods with almost specific features given their periodicity while maritime players can also being classified within groups with increased similarity. Thus, in order to include the time dimension, the system first searches for similar users and situations and then performs the conventional case retrieval and ranking. Moreover, the future situation is not based on scenarios but on forecasting the most important variables. Finally, tools for the explanation of the policy rationale are included in the system.

3. The Recruitment Problem – Stakeholders and objectives

The problem of officer's shortage in the maritime industry is a well documented one and the need for the development of a policy dealing with it is well established. The Institute of Employment research at Warwick University in cooperation with BIMCO, ISF publish a 5 year report based on maritime data which includes global and regional features regarding the availability of seafarers for the world's merchant fleet. Their latest release showed that the global supply of seafarers was 466,000 officers and 721,000 ratings, while demand was estimated at 476,000 officers and 586,000 ratings (BIMCO, ISF, Manpower Update 2005). This means that there is a continuing shortage of qualified officers (2%) globally and a significant surplus of ratings. Secondary findings were the continuing east bound supply shift, the increasing demand for seafarers which is more intense on certain grades and types of ships, and the significant ageing of senior officers, especially in Western countries. This case examines the Recruitment problem in Greece while a regional or global approach is possible.

Any policy has an overall objective or vision which is served through the satisfaction of the goals selected. Moreover every policy requires the allocation of scarce resources in order to achieve the selected goals (Dye, 1976). Two fundamental questions arise before starting to formulate a policy and concern the suitability of the selected way to achieve those goals and the verification of the optimum usage of the allocated resources. The complexity of an AI system able to explore the possibilities of resource allocation significantly increases and is considered beyond the philosophy of this essay. Therefore, it is assumed that some kind of intervention (policy) has been proven necessary and that any opportunity cost scenario will not be examined. This system recommends the optimum way to achieve prespecified objectives and the associated goals taking into account any given constraints and limitations. The entire procedure is generally based on ROAMEF cycle (HM Treasury, *Green Book: Appraisal and Evaluation in central government*, 2003) and includes tasks like explaining the rationale behind the intervention, setting objectives, conducting the appraisal, monitoring and evaluation of the policy and providing feedback about the actual results. The sequence may be altered and steps may be omitted according to the problem's nature and the method's limitations. The "Recruitment problem" is used as a theoretical base since more detailed essays concerning its procedures are available and under way for publication.

Since the system faces prespecified problems, the setting of an overall objective takes place when the policy maker selects the particular problem to deal with. For example, the "Recruitment" problem's objective is to ensure the proper number of sea going personnel while the "Autopilot's" (Nikitakos and Fikaris, 2008) objective is the optimum automatic steering of a ship. An overall objective can be broken down to specific goals set directly by the policy maker without any assistance from a knowledge base. On the other hand these goals may not be so obvious in certain problems and the policy maker may omit some of them, critical for the achievement of the overall objective or include some of negligible importance. The development of a representation model (Figure 1) for better understanding the problem may be helpful towards this task. Moreover, setting a particular –not adequately analyzed- goal may lead to unexpected results for the target group or other parts of the society. In order to deal with these problems the system incorporates a routine for the selection of critical stakeholders and identification of their objectives and goals.

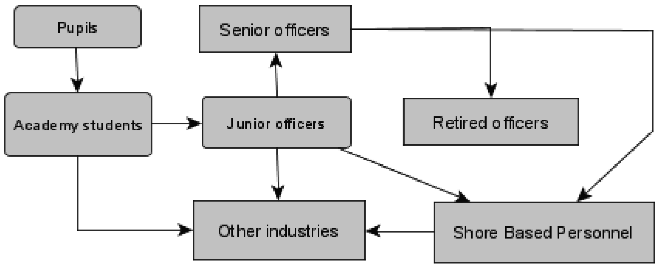


Figure 1: Manpower flow for the maritime industry – The case of Greece

Stakeholder selection is heavily based on the policy maker's judgement and personal preferences. The system uses a semi autonomous stakeholder identification tool based on a method of separation, role specification and ranking according to importance, strength and influence of each stakeholder (Vos and Achterkamp, 2006) in order to assist and simplify this procedure by interacting with the policy maker, detect his beliefs and preferences and guide him towards an objective method of selecting the proper stakeholders for the specific policy. This tool is slightly modified according to the view that when it comes to policy making the existence of a relation is not a sufficient condition for a societal group being characterized as a stakeholder while ethics imply it would (Freeman, 1984). The tool includes an importance weighting component –with default setting but subject to user modification- so the identified stakeholders are listed and presented weighted (Table 1):

Table 1: Weighted stakeholder list – The case of Greece

Stake holder	Importance	Power	Interest	Predictability	Support	Sum
Government						1,0
M M Marine						1,0
M. Education						1,0
Policy maker						1,0
Shipowners	10	10	10	10	-5	0,35
Pupils	10	10	10	0	0	0,30
Mar. academies	10	5	10	0	-5	0,20
Society	10	10	0	0	0	0,20
Maritime Unions	5	5	5	0	0	0,15
M R Companies	0	0	0	10	-5	0,05

After the selection and weighting of stakeholders the routine continues with a need's assessment for each selected one. A needs assessment is a systematic set of procedures undertaken for the purpose of setting priorities and making decisions about program or organizational improvement and allocation of resources. The priorities are based on identified needs (Witkin and Altschuld, 1995). A needs assessment is a participatory process that takes place in three phases: the preassessment where exploratory work takes place, the assessment where a combination of means[1] is used to collect data[2] about the problem and the post assessment where data is analyzed and priorities are set. This procedure will also serve as a basis for the indexing phase described later and is assisted by a developed model representing the crucial variables that contribute to the overall objective:

[1] Those means include single step or cross sectional surveys, multi step syrveys (Delphi technique), community forums (or public hearings), focus groups, nominal group processes, interviews, observation, etc (Burke J, Needs Assessment: The key to why things change is the key to everything, University of Memphis, USA)

[2] Data can be primary when the person who conducts the NA asks specific questions to the target group and secondary when data is collected by someone else and is available for use.

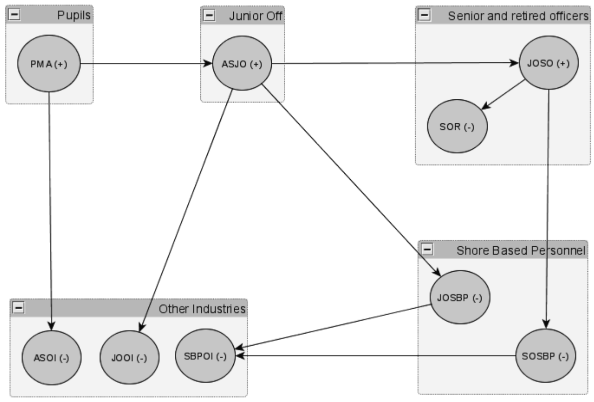


Figure 2: Goal variables and associated sub models – The case of Greece

Key goal variables identified were the rate of pupils selecting the maritime academies / total pupils (PMA), the analogy of academy graduates becoming junior officers (ASJO) towards those leaking to other industries (ASOI), the proportion of junior officers becoming senior ones (JOSO) compared to those leaking to other industries (JOOI) or those taking over shore based jobs (JOSBP), the number of senior officers retiring (SOR) or moving to shore (SOSBP) and finally the trade off between shore based professionals and other industries (SBPOI) which affects the overall demand. All these represent the policy maker's possible goals which he / she must modify according to his preferences. This model is further analyzed to sub models in order to identify the variables affecting the goal variables. An example of the PMA sub model is shown in Figure 3:

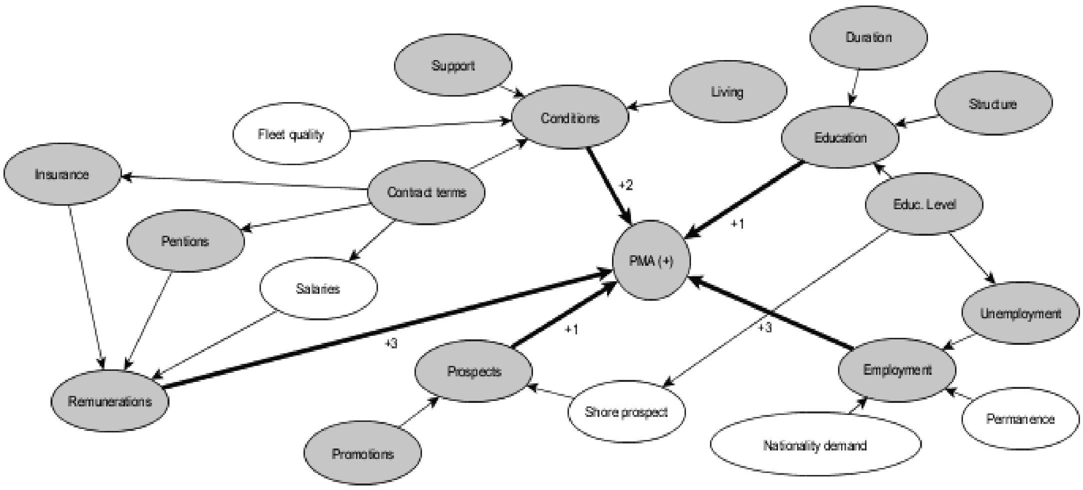


Figure 3: PMA Sub model: Factors affecting the PMA variable – The case of Greece

The PMA sub model consists of five basic dimensions representing conditions on board, educational system, remunerations, career paths and employment prospects with their corresponding affecting variables. The grey variables are the ones affected by the policy maker[3] while the white ones are considered non – controlled. Each model variable (including the goal variables) is inserted to the knowledge base with the associated limitations and constraints derived from the need assessment of the selected stakeholders in the form of rules for use during the retrieval procedure.

4. The Recruitment Problem - Representation and Indexing

The compulsory part for a CBR case is the description of the problem and solution, (Gebhardt F., Vob A., Grather W., and Schmidt – Beltz B., 1997) while this description can use any Knowledge

[3] For this Recruitment problem's session, it is the Ministry of merchant Marine.

Representation formalism, like frames, objects, predicates, semantic nets, time dependent data and rules. Additionally a case may include a description of the situation (environment) where the problem takes place and a result demonstrating the state of the world after the application of the solution suggested (Kolodner, 1993). The dominating approaches for case representation include the relational databases whereas each column represents a feature and each row constitutes a case, the object oriented representation which is mostly suitable with CAD, multimedia and GIS applications and the predicate representation where cases are expressed as a set of IF … THEN … rules that give flexibility to the system but cannot manage a large case base (Bergmann R., Kolodner J., Plaza E., 2005).

Regardless of the representation technique a case base must be organized properly in order to facilitate the retrieval process and ensure its efficiency and completeness. The serial ways of organizing a case base include the flat memory – serial search approach where cases are searched sequentially, the shared - featured networks where cases are organized hierarchically and the search space is reduced to a small subset with members sharing common attributes and discrimination networks where the case base is subdivided to smaller subsets after a sequence of questions. Parallel approaches include redundant discrimination networks which use different orders of questions, the flat library – parallel search approach that uses parallel processors and the hierarchical memory – parallel search that combines hierarchical memory with parallel retrieval algorithms (Kolodner, 1993).

Indexing is the assignment of indexes to cases for future retrieval and comparison. They must reflect the important aspects of the case and especially the attributes that influence the outcome, simultaneously being adequately abstract in order to retrieve a reasonable number of cases (Birnbaum L., Collins G. 1989, Hammond K. J., 1989). Even though the indexing procedure is usually manual, automated indexing procedures have been also developed. Some of them include inductive techniques for learning local weights by comparing similar cases (Bonzano A., Cunningham P., and Smyth B., 1997), indexing methods by features and dimensions across the domain (Acorn T., and Walden S., 1992), computation of differences between cases and adaptation guided indexing and retrieval (Smyth B., Keane M. T., 1998).

The system uses the relational database approach and a typical case includes the problem, its solution, and the situation where the policy takes place as well as its outcome. The problem's indexes are the variables identified in the model with the stakeholder's goals and constraints associated with any variable inside the model. These indexes guide the retrieval process since their dynamic evolution is measured over the time horizon set by the user and is compared with the desired goal setting of the new case. The solution consists of the model's variables values affecting each goal provided that they do not violate goals or constraints set by important stakeholders. The solution is also tracked and presented dynamically over the same time horizon. The situation is represented by model internal or external variables describing the maritime environment at the time period of the policy. This description is dynamic and includes the situation behind (what was the situation) as well as the situation ahead (what might happen) and for that reason external forecasts for key variables are used. Additionally, the situation module includes indexes about the user –especially if it is a country- since a successful policy usually works again in similar users. Finally, the case includes outcome indexes valued after the policy implementation in order to measure the success of the policy – case.

5. The Recruitment Problem – Retrieval, Adaptation and Storage

Common retrieval approaches can be divided in two categories (Liao, Zhang and Mount, 1998), the first based on distance computation between cases and the second on the representation and indexing structure of the cases followed by a comparison with the structure of other cases. Distance based approaches include the weighted Euclidean distance (Sankar K. P., Shiu S. C. K., 2004) which calculates the square root of the sum of squares of the differences of the corresponding coordinates of two objects, the Hamming distance (Hamming, 1950) which calculates the number of bits that are different between two bit sectors and the Levenshtein distances (Soukoreff and MacKenzie, 2001) which estimates the number of deletions, insertions or substitutions required to transform a string to the target string. Text similarity is dealt using Dice, Jaccard and Cosine coefficients which are based on the frequency of occurrence of an object i.e. word within a text. Index comparing approaches include the works of Tversky (1977), Sebag and Schoenauer (1993) and Weber (1994) that are based on the number of attributes that are common or similar between the new and the stored case and their manipulation using several algorithms. Other approaches are using the number of consecutive matches (Hunt et al., 1995), the relevance of certain attributes (Cain et al., 1991), structural similarities (Burke et al., 2001), etc.

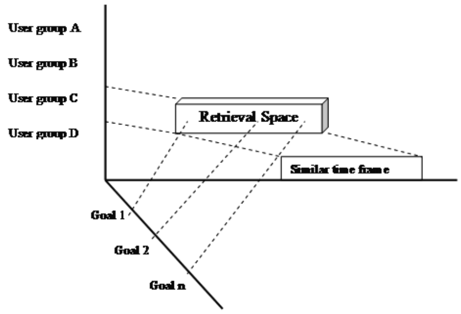


Figure 4: Narrowing the retrieval space using situation and user similarity

The system detects cases having a similar rate of change as the new case as far as the goal variables are concerned given the time horizon. Since this is a very time and resource consuming procedure the system narrows its search space by matching the user and maritime environment attribute values to similar time frames and users of the past. As for the environment the system uses the maritime market circles and the associated forecasts to find a similar situation dealt before. The case base is already clustered to historical time frames characterized as growth, recession, stagnant, etc, while the system estimates the critical situation variables of the new case and classifies it to the corresponding cluster. Similarly, the user i.e. a country is classified to the corresponding cluster according to its maritime similarity with its members (Figure 5).

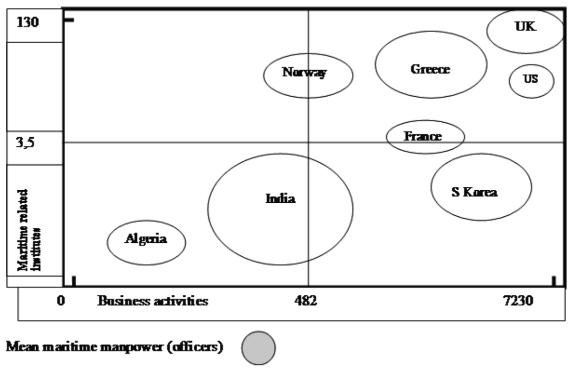


Figure 5: User clustering matrix – Maritime Countries

After the user and situation classification the system uses the classical nearest neighbour approach (Kolodner, 1993) comparing the dynamic goal evolution in the case base with the desired change in the problem specifications (Montani and Portinale, 2005). Then, the cases are ranked according to their utility based on the goal relative importance, the constraint satisfaction and the policy maker's preferences. The system detects the constraint – free variables that significantly affect the goal variables and presents them dynamically to the user. Additionally, the expected results are presented as well as a chain of reasoning.

The most common adaptation procedures include structural adaptation in which adaptation rules are applied directly to the solution, derivational adaptation that reuses the algorithms that generated the original solution, null adaptation that applies the retrieved solution as it is, parameter adjustment that compares specific parameters of the retrieved and current case for modification, critic based adaptation that uses a critic to look for combinations that can cause a problem, model guided repair that uses a causal model, case based substitution that uses cases to suggest solution adaptation (Watson and Marir, 1994).

The system uses specific adaptation rules associated with the goals and constraints set by the policy maker and selected stakeholders for each variable contained in the model. These rules explain the solution's problem and suggest a repair strategy for the problematic variable by altering or tuning its affecting variables while reconstructing the case. The policy maker may choose to skip the adaptation procedure and proceed to the second best retrieved solution.
Finally, the successful cases which are significantly different from their associated class are stored in memory for future use.

6. The Recruitment Problem – Explanation of Rationale

Explaining the rationale means proving that a policy is needed and showing that the benefits from this intervention exceed the possible costs. Successful public policies must ensure economic efficiency, equity, additionality and regeneration in order to be rationale (HM Treasury, *Green Book* Annex I, 2003). Economic efficiency requires that any resources will be used in the most productive manner possible (Barr, 2004) while equity means that the costs and benefits of an intervention will be distributed equally among societal groups (Messick and Cook, 1983). Additionality is the increase in output or employment in a particular area after the policy implementation while regeneration requires that the policy contributes to the overall development.

Even though a policy's rationale is often proved by simple observation, common sense or a short literature review, complex problems require hard evidence in order to persuade the public and key stakeholders for the necessity of an action. When the system faces not so well documented problems, a justification procedure based on the four aforementioned requirements for rationale is activated. Most active applications are using Pareto[4] (Barr, 2004; Tevfik, 1996) or Kaldor – Hicks efficiency rule (Kaldor 1939; Hicks 1939) to ensure economic efficiency and this is also the case for the Recruitment problem. After the deduction of the resources required by the suggested policy to monetary values (for each stakeholder) a compensation potential on behalf of the "winners" is estimated in order to give a clearance for implementation.

Equity is closely related to this, since those monetary values accrued from the policy decisions are particularized for every stakeholder affected by the policy and compared according to prespecified equity thresholds set by the policy maker, executing a per stakeholder cost benefit analysis (Baily et al., 1993). If qualitative variables are included a similar method[5] capable of handling those is used. The main concept here is to ensure that no stakeholder burdens beyond a preset limit and simultaneously verify that there is a reasonable benefit per output unit exceeding the corresponding cost.

Additionality is calculated using a combination of forecasts and expectations about manpower supply and demand for a preset time horizon after the policy implementation (UK, The National Regeneration Agency, *Additionality: A full guide*, English partnerships, 2001). Finally, regeneration is examined through the comparison of the policy effects towards the society needs analyzed below (UK, HM Treasury, EGRUP, *A framework for the Evaluation of regeneration projects and programmes*, 1995).

7. Conclusion – Future research

As mentioned in the beginning of this essay this paper is mostly conceptual since much of the work described here is currently under evaluation. A more detailed version including the entire case study

---

[4] A WPO for less structured problems and a SPO for better supported ones
[5] Multiattribute utility theory or AHP for example

will follow shortly. In this paper, the assumption that CBR is suitable for maritime policy making applications and possibly competitive with other approaches is gaining some ground. The nature of the maritime industry permits experiments as the retrieval facilitation through user and situation clustering or multi layer CBR implementation. The use of formal tools for the stakeholder and goal identification enhances the robustness of CBR and ensures completeness.
The employment of fuzzy logic or other appropriate soft techniques will produce better results and help significantly in handling insufficient or vague data in certain applications. The dynamic character of CBR data must be the rule and not the exception in policy making applications and further study should be done towards this area. Finally, combining CBR with rule based or model based systems in the maritime sector can produce significantly better results than stand alone methods.

References

Acorn, T. and Walden, S. (1992), SMART: support management automated reasoning technology for Compaq customer service, Proceedings of the 4th Innovative Applications of Artificial Intelligence Conference.

Alter, S.L. (1980), Decision support systems: current practice and continuing challenges. Reading, Mass., Addison-Wesley Pub, 1980.

Baily, N., Burtless, G. and Litan, R. (1993), Growth with Equity: Economic Policy making for the next century, Brookings Institution, Washington, 1993.

Barr, N. (2004), Economics of the welfare state, N.Y, Oxford University Press, USA, 2004.

Bergmann, R., Kolodner, J. and Plaza, E. (2005), Representation in case based reasoning, Knowledge Engineering Review, 20(3), 209 – 213.

BIMCO ISF (2005), Manpower Update, Warwick Institute of Employment Research, 4th edition.

Birnbaum, L. and Collins, G., (1989), Reminding and engineering design themes: A case study in indexing vocabulary, Pensacola, Florida, Morgan Kaufman.

Bonzano, A., Cunningham, P. and Smyth, B., (1997), Using introspective learning to improve retrieval in CBR: A case study in air traffic control, in Proceedings of International Conference on Case-Based Reasoning, Springer Verlag, pp413–424.

Burke, E.K., MacCarthy, B., Petrovic, S. and Qu, R. (2001), Case Based Reasoning in Course timetabling: an attribute graph approach, Proceedings of the Fourth International Conference on Case Based Reasoning, Vancouver, Canada, Springer Verlag, Berlin pp. 90 – 104.

Burke, J., Needs Assessment: The key to why things change is the key to everything, University of Memphis, USA.

Cain, T.P., Pazzani, M.J., and Silverstein, G. (1991), Using domain knowledge to influence similarity judgement, Proceedings of the 1991 DARPA Workshop on Case Based Reasoning, San Francisco, Morgan Kaufmann, pp. 191 – 199.

Dye, T.R. (1976), Policy Analysis. University of Alabama Press, 1976.

EU, Blue Book, Vision for a European Maritime Policy, 2007.

Freeman, R.E. (1984), Strategic Management, a Stakeholder Approach, Pitman, Boston.

Gebhardt, F., VoB, A., Gräther, W. and Schmidt-Belz, B. (1997), Reasoning with Complex Cases, Kluwer Academic Publishers.

Hamming, R.W. (1950), Error detecting and error connecting codes, Bell system technical journal vol. 9, pp. 147 – 160.

Hammond, K. (1989), Case Based Planning: Viewing planning as a memory task, Boston, Academic Press.

Hicks, J. (1939), The Foundations of Welfare Economics", Economic Journal 49: 696–712.

HM Treasury, Green Book: Appraisal and Evaluation in central government, 2003.

Hunt, J.E., Cooke, D.E., and Holstein, H. (1995), Case memory and retrieval based on the immune system, Proccedings of the First International Conference on Case Based Reasoning, Sesimbra, Portugal, Springer Verlag, Berlin, pp. 205 – 216.

Jenkins, W. (1978), Policy Analysis: A Political and Organizational Perspective, 1978.

Kaldor, N. (1939), Welfare Propositions in Economics and Interpersonal Comparisons of Utility. Economic Journal 49 (195): 549–552, 1939.

Kljajic, M. et al (2002), System Dynamics Model Development of The Canary Islands for Supporting Strategic Public Decisions, Proceedings of the 20th International Conference of the System Dynamics Society, The System Dynamics Society, Albany, NY.

Kljajic, M., Legna, C., and Skraba, A. (2003), Simulation model of Canary Islands, Preliminary Results. 21st System Dynamics Conference, July 20-22, New York.

Kljajic, M., Legna, C. and Skraba, A. (2003b), Simulation Model of Canary Islands for Public Decision Support, International Conference on Applied Simulation and Modelling, (Hamza, M., Ed.), Marbella, 3-5 sep. 2003, pp. 644-649,ACTA Press, Zürich.

Kolodner, J. (1993), Case Based Reasoning, Morgan Kaufman, San Francisco, 1993.

Leake, D.B. (1996), Case Based Reasoning Experiences, Lessons and Future Directions, AAAI Press / MIT Press, Cambridge, MA, 1996.

Legna, C. and Gonzalez, C., (2005), An Intelligent Decision Support System (IDSS) for public decisions using System Dynamics and Case Based Reasoning (CBR), Scientific papers Vol 38.

Legna, C. (2000), Knowledge and knowledge management for the improvement of strategic public decisions, IFIP TC8/WG8.3, International Conference on Decision Support through Knowledge Management, (Carlsson S., Humphreys P., Lundberg B. G., Cosh A. & Rajkovi V., Editions) Stockholm 9 – 11 / 07 / 2000 pp. 175 – 190, Stockholm University: Royal Institute of Technology, Stockholm.

Liao, T.W., Zhang, Z., and Mount, C.R. (1998), Similarity measures for retrieval in case based reasoning systems, Applied Artificial Intelligence, vol. 12, pp. 267 – 288.

Locke, E.A. (2001), Motivation by Goal Setting, Handbook of Organizational Behavior, 2: 43-54.

Messick, D. and Cook, K. (1983), Equity th.: psychological and sociological perspectives, Praeger.

Montani, S. and Portinale, L. (2005), Case Based Representation and retrieval with time dependent features, Case Based Reasoning research and development, Springer Berlin vol 3520 / 2005, Heidelberg, pp. 353 – 367.

Nas, T.F. (1996), Cost-Benefit Analysis: Theory and Application, CA., Sage, 1996.

Nikitakos, N. and Fikaris, G. (2007), Navigating a Ship using an Expert System, Hellenic Institute of Maritime Technology, 2007.

Psaraftis, Maritime Safety: To Be on not to Be Proactive, World Maritime University's, Maritime Affairs Journal, October 2002.

Sankar, K.P. and Shiu, S.C.K., (2004), Foundations of Soft Case Based Reasoning, Wiley Series of Intelligent Systems, Albus J., Meystel A. and Lofti Zadeh, Series Editors.

Sebag, M. and Schoenauer, M. (1993), A rule based similarity measure, Proceedings of the First European Workshop, Kaiserslautern, Germany, Springer Verlag, Berlin, pp. 119 – 131.

Sheffrin, S.M., (2003), Economics: Principles in action. Upper Saddle River, New Jersey 07458: Pearson Prentice Hall. pp. 15.

Simon, H.A, (1977), The New Science of Management Decision, NJ: Prentice Hall.

Smyth, B. and Keane, M.T. (1998) Adaptation-Guided retrieval: questioning the similarity assumption in reasoning. Artificial Intelligence, 102:2, pp. 249-293.

Soukoreff, R.W. and MacKenzie, I.S. (2001), Measuring errors in text entry tasks: an application of the Levenshtein string distance statistic, Extended Abstracts of the ACM Conference on Human Factors in Computing System, ACM New York, pp. 319 – 320.

Tversky, A., (1977), Features of similarity, Psychological review, vol 84 (4), pp. 327 – 352.

UK government, Ministry of Health, 14 criteria for excellence in policy making.

UK, HM Treasury, EGRUP, A framework for the Evaluation of regeneration projects and programmes, 1995.

UK, National Regeneration Agency, (2001) Additionality: A full guide, English partnerships.

Vos, J.F.J. and Achterkamp, M.C. (2006), A method for the identification of stakeholders. In: R.J. Jorna, Editor, Sustainable innovation. The organisational, human and knowledge dimension, Greenleaf Publishing, Sheffield (2006), pp. 98–110.

Watson I. and Marir F., (1994), Case Based Reasoning: A review, The Knowledge Engineering Review, Vol.9 No.4, Auckland, New Zealand.

Weber, G, (1994), Examples and remindings in a case based help system, Proceedings of the Second European Workshop, Chantilly, France, Springer Verlag, Berlin, pp. 165 – 177.

Witkin, R. and Altschuld, J. (1995), Planning and conducting needs assessments, Thousand Oaks, CA, Sage.

The economics behind the awarding of terminals in seaports: economic issues in the pre-bidding phase[1]

Theo Notteboom[1,2], Christophe Theys[1], Athanasios Pallis[3,4] and Peter W De Langen[5]

[1] Institute of Transport and Maritime Management Antwerp, University of Antwerp, Belgium.
[2] Antwerp Maritime Academy, Antwerp, Belgium.
[3] Department of Shipping Trade and Transport, University of the Aegean, Chios, Greece.
[4] Centre for Energy, Marine Transportation and Public Policy, Columbia University, NY, USA.
[5] Department of Corporate Strategy, Port of Rotterdam Authority, Rotterdam, The Netherlands.

Abstract

Terminal concessions in seaports have only recently attracted academic attention. Issues such as the allocation mechanisms (to be) used for granting those concessions, the determination of the concession term and concessions fees, as well as the inclusion of special clauses aimed at assuring that the terminal operator will act in the interest of the port authority and the wider community, are increasingly interesting both academically and for the port industry. This contribution gives an overview of the different phases of a typical terminal awarding process and distinguishes among pre-bidding, awarding (including prequalification and selection procedures) and post-bidding phases. The paper then focuses on the discussion of economic issues that might occur during the pre-bidding phase of the awarding procedure.

Keywords: Concessions, Seaports, Governance

1. Introduction

Port economic research has witnessed a strong surge in the last couple of years (see e.g. Cullinane and Talley, 2006, Pallis et al., 2009). Seaports worldwide are confronted with scarcity of land and large, often global, infrastructure users. The awarding of terminals to these users is prime task of landlord port authorities. Yet terminal concessions in seaports have only recently attracted academic attention (see e.g. Notteboom, 2007, Pallis et al., 2008, Theys et al., 2009). Issues such as the allocation mechanisms (to be) used for granting seaport concessions, the determination of the concession term and concessions fees, and the inclusion of special clauses aimed at assuring that the terminal operator will act in the interest of the port authority and the wider community (cf. throughput guarantees and environmental clauses), are interesting both academically and for the port industry.

In academic literature on public-private-partnerships (among which seaport terminals) and in practice distinctions are often made between various forms of contracts. The World Bank's Private Participation in Infrastructure Database[2] categorizes the different contracts primarily on the basis of the ownership, investment decisions and risk division. With *management and lease contracts* the private partner operates government-owned facilities for a pre-specified period. Management contracts thereby differ from leases in that in the former the public authority retains operational risk and pays the private partner a premium for his management services, while in the case of leases operational risk is transferred to the private entity which pays a (lease) fee to the public authority. In

[1] This paper is a partially reworked excerpt of another paper by the authors which discusses the different phases of the terminal awarding process in detail. That paper presents a classification scheme for awarding procedures, illustrates the applicability of established economic theories such as auction theory and defines an elaborate research agenda on the topic (see Theys et a.l, 2009).

[2] http://ppi.worldbank.org, accessed December 2008.

both cases, however, investment decisions remain the sole responsibility of the public partner. With *concessions*, on the other hand, private partners engage in substantial investments in existing or new (by the World Bank called *greenfield*) facilities and may transfer the facilities back to the government upon termination of the concession. Build-Operate-Transfer (BOT) projects are a well-known example of a greenfield project, but a vast range of public-private alternatives exist (see World Bank, 2007). The remainder of this paper will focus on awarding procedures in the landlord port authority model and refer to a terminal concession as "*a grant by a government or port authority to a (private) operator for providing specific port services, such as terminal operations*" (Notteboom, 2007).
A typical terminal awarding procedure consists of three phases:

- *Pre-bidding phase*: The awarding authority (typically a port authority or any other managing body of the port) makes the necessary preparations for the awarding taking into account prevailing regulatory conditions. The awarding authority has to decide on key issues related to the awarding procedure and make this information available to interested candidates. In this phase, the rules of the game are defined;
- *Awarding phase*: Candidates are screened, bids are evaluated and the most appropriate candidate is selected. The challenge for the awarding authority lies in making the right choice given the parameters set in the pre-bidding phase;
- *Post-bidding phase*: A legally binding contractual agreement is signed with the selected candidate and – in some cases – the company's performance is monitored during the contract term. If necessary, correcting measures are taken and disputes are settled.

Figure 1 provides an overview of these three phases of the awarding procedure and indicates which issues typically arise in each of them. Issues to resolve during *pre-bidding* concern decisions on what to award and under which conditions the awarding will take place. Prequalification is often (but not always) a first step in the *awarding phase* and consists of the selection of qualified companies out of the pool of interested candidates. It is based on a set of conditions developed by the awarding authority before the start of the awarding phase. Companies that have met these conditions are then allowed to one or more selection rounds in which the terminal concession is awarded.

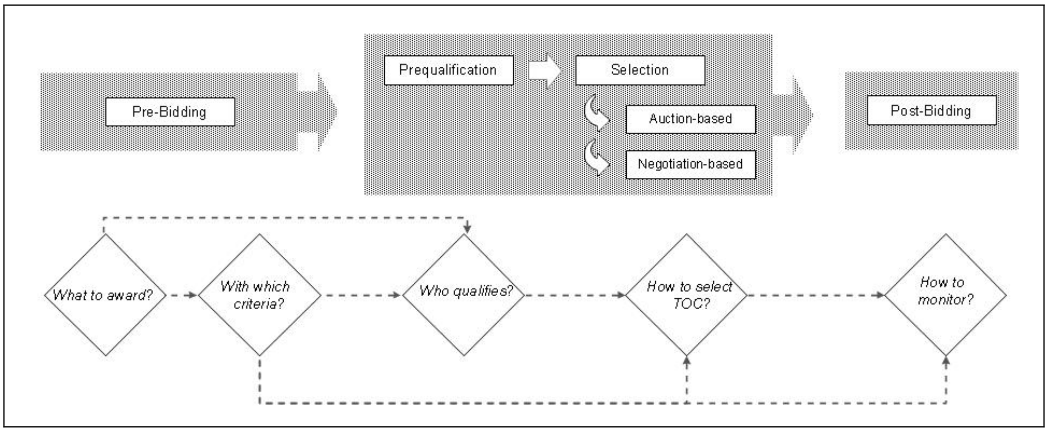


Figure 1: Schematic overview of awarding procedure and relevant issues

Awarding authorities can choose from a variety of different setups, but selection generally takes place on the basis of negotiations and/or auction-alike structures. Although the formal use of auctions is not common, many competitive selection procedures in seaports *de facto* use auctions because the terminal is assigned to the highest bidder – not necessarily in money terms but also with the highest 'score' on a number of criteria. Negotiations with the highest bidder after the (de facto or formal) auction are not rare and might even result in the reverse of the initial decision.[3]

[3] The concessioning of the two major container terminals in Greece, in particular developments after the auctioning (which was completed in May 2008) provide illustrative examples. In the case of the port of Thessaloniki the highest bidder was Hutchison Port Holdings. But following negotiations with the Port Authority, that lasted more than six months, it decided to withdraw its interest in December 2008. In the case of

The discussion of the different phases of the terminal awarding procedure is associated with the rationale behind the use of concessions. Since the seminal work of Demsetz (1968) competition *for* the market – such as through the competitive process – has been proposed in cases where competition *in* the market fails. In industries characterized by high economies of scale , natural monopolies leading to the highest possible market concentration are most cost efficient. Yet Demsetz (1968) demonstrated that such natural monopoly situations should not necessarily be inferior to (regular) competition among market players *in* a certain market. Given appropriate awarding mechanisms and a sufficiently large number of players competing for the right (or concession) to serve the market, the effect of such competition *for* the market – where the most efficient firm will be selected – can be similar to that of a competitive market within which many players operate. Hence, applied to a seaport setting, it is significant to analyze under which conditions competition *for* the terminal market is preferable over competition *in* the terminal market[4]. However, in the landlord model, port authorities use concessions regardless of the extent of competition *in* the market. For instance, in Rotterdam concessions are granted for new container terminals, in a market (the Hamburg-Le Havre range) that is widely regarded as rather competitive. In this case, there is both competition *for* and competition *in* the market.

In the remainder of the paper the use of concessions *is given* and the focus in Sections 2 to 7 is on the discussion of issues that might occur during the *pre-bidding phase* of the awarding procedures. Section 8 then presents short conclusions. For a detailed overview on the application of economic theories (including auction theory) to terminal concessions in seaports, and the research challenges related to not only pre-bidding, but also prequalification, selection and post-bidding the reader is referred to Theys et al (2009), of which this paper is a partially reworked excerpt.

2. The Site to be Concessioned

Once the decision has been taken to concession the operation of a (future) terminal, the awarding authority will have to describe the object of the concession. The port site that will be given in concession should be carefully delimited. Moreover, the technical and geological characteristics of the site and the physical, operative, commercial, juridical and labour aspects of the proposed concession have to be analyzed and filed. Such information typically ends up in later documents related to the awarding procedure (public auctions/tender documents or requests for proposal in a competitive bidding process), in order to avoid information asymmetry in the awarding process.

Given that the whole process might last for a considerable period of time, the stage of development of the terminal site to be awarded to private operator needs to be also defined. The awarding authority might opt to award either (a) an undeveloped site where the operator will have to develop infrastructure; (b) a greenfield site with infrastructure developed to site boundary; (c) an improved site with a quay line, paved yard but without buildings or handling equipment; (d) a site with all civil works completed but the operator supplies quay cranes and yard handling equipment; or (e) a fully developed site including quay cranes but the operator supplies yard handling equipment. This decision is closely related to the financial and operational risk to be transferred to the winner of the concession and might be deterministic for the number of bidders.

If the port authority does a poor job, candidates will have to verify the state of the premises and possibly perform surveys and/or field studies to prepare their bid. This incurs transaction costs and may reduce the number of bidders. In addition, any gaps left by the awarding authority during this stage provides the winner of the concession opportunities for manoeuvring during the later stages and even after the finalisation of the process

the port of Piraeus, the PA and Cosco Pacific signed the relevant contract in November 2008, yet the owner of the port (Greek government) has not ratified the agreement as further details are currently under discussion.

[4] Van Niekerk (2005) claims that in ports in developing countries with low cargo volumes, in remote countries that only serve natural hinterlands and in ports distant from major shipping routes concessions might not be the best instrument to realize the goals port authorities might have.

3. Multifunctional versus Dedicated Terminal Use

An awarding authority might want to decide on the main use of the terminal in the pre-bidding phase. Alternatively, the awarding authority might decide to just invite bids for a certain plot rather than specifying how it should be used.

The port authority could decide to allow a wide range of commodities on the premises. The likely result is a multifunctional terminal which demands less dedicated investments from the terminal operator and allows for a flexible use. However, the multifunctional use of a facility might make the selection process more complex as the port authority is likely to be confronted with candidates putting forward very different proposals in terms of the markets to be covered, the expected throughput volumes and required investments.

If the awarding authority clearly defines the use of the terminal (e.g. plot to be used for container handling only), then the port authority will, later on in the process, have an easier task in evaluating the technical and economic proposals of the candidates on an equal footing. The level of facility dedication to one commodity or market segment is thus one of the key decisions to be taken by the awarding authority in the pre-bidding phase[5].

While the type of terminal use is a clear practical issue, with a lot depending on the specifics of the port/terminal, or even throughput orientation (e.g. transhipment or not), the impact on the awarding process has never received attention in academic literature. Research has been done on multi-user versus single-user terminals. The question whether or not to allow single-user terminals forms an integral and crucial element in terminal leasing policy. Turner (2000) demonstrated that the position of terminal leasing policy as regards single-user terminals has a clear impact on system performance. Musso et al. (2001), Haralambides et al. (2002) and Cariou (2003) provide a more in-depth analysis on the concept, the pros and cons of dedicated/single-user terminals, as well as issues of pricing. Market reality has made some world ports, which were previously reluctant to allow dedicated terminals, now embrace the concept in view of binding cargo to the port.

4. Splitting and Phasing Decisions

The size of a terminal site to be concessioned is a complex issue. In the pre-bidding phase, the port authority will have to make key decisions. Assume a port authority has 200ha of port land available for container terminal development. Is this site best concessioned as one terminal or is the plot splitted in two or more sections, to be concessioned separately? In addition, the concession process can be phased. For example, the port authority could decide to award a first section of 100ha in year 1, 50ha in year 3 and the remaining 50ha in year 5. While the above strategic questions on splitting and phasing seem straightforward, the answers are not easy, given the far-reaching impact these decisions may have on the competitive profile of the port concerned. Two key problems can be identified in this respect.

[5] The dedication level also has an impact on the cost base for the terminal operator. For example, a port authority might stipulate that any warehouse developed on the site should be made suitable for the storage and handling of dangerous cargo. Such a requirement increases construction and operational costs for the multifunctional warehouse due to the regulatory requirements associated with the storage of dangerous cargo (e.g. thicker concrete floor, water treatment system, fire prevention measures). The terminal operator will thus be confronted with a higher cost base, even though the warehouse might never be used for the storage of dangerous cargo.

Terminal sections

Timing of the concessioning	**One**	**Two or more**
One point in time	No phasing No splitting	No phasing Splitting
Spread over several moments/years		Phasing Splitting

Figure 2: Options related to the splitting and phasing of a terminal site

The first problem relates to the way sections of a terminal plot are phased. A port authority could organize a separate competitive bidding procedure or auction process for each phase. Alternatively, a port authority might consider a second terminal phase as a mere extension of the initial terminal phase and decide to award the second phase to the incumbent terminal operator by direct appointment, thereby avoiding a competitive bidding process. Whether direct appointment is a viable option depends on (1) the relative dimensions of the phases and (2) the level of inter-terminal competition in the port and between the port and neighbouring ports. If the second phase of the terminal represents a marginal extension of the first phase then a competitive bidding process or auction for the second phase obviously does not make a lot of sense (at least when the incumbent terminal operator shows interest in also operating the second phase). In case the scale differences between the first terminal phase and subsequent terminal phases are small, then direct appointment to an existing operator is a less obvious choice. Port authorities who want to use direct appointment could opt for a gradual phasing of terminal extensions: the firm operating the first (larger) phase is given the chance to expand its facility berth by berth over time through direct appointment, so without having to face potential competitors through a competitive awarding process. Such a deliberate practice could emanate in collusive behaviour between the port authority and the incumbent terminal operator.

The second problem connected to the phasing and splitting of a terminal plot relates to the dynamics in intra-port competition. Consider a port with a captive hinterland served by an existing container terminal of 100ha operated by an incumbent firm X. The port authority plans to concession an additional plot of 200ha for container terminal development. If the port authority organizes a competitive bidding procedure for the new plot and makes the incumbent firm eligible to compete for the terminal concession, then two outcomes are possible:

- The incumbent firm wins. The incumbent thus strengthens its monopoly position and can extract monopoly rents. Especially when an incumbent firm does not face regulatory problems it is most likely to win the bidding procedure, as it will bid the economic rent away in the bidding process. Eventually the result is high revenues for the port authority, but high costs for port users. At the same time, the scale of the facilities is likely to generate economies of scale and scope, creating the potential for lowering the handling costs per unit for the operator.
- A new entrant is granted the concession, thereby introducing intra-port competition. Given the scale differences between the existing and new facility, the entrant in principle has a good starting position to outperform the incumbent firm.

The port authority might also opt to split up the 200ha plot in two separate sections: either the port authority then awards the two sections separately via a competitive process, or awards one section via a competitive to allow new entrants (incumbent is excluded from the competitive bidding or auction) and assigns the other section to the incumbent firm (to allow growth). Consequently, port authorities have, in principle, some leverage in shaping the competitive process within the port. The example discussed here concerned a port with a captive hinterland. Things become much more complicated when the neighbouring ports are vying for the same cargo.

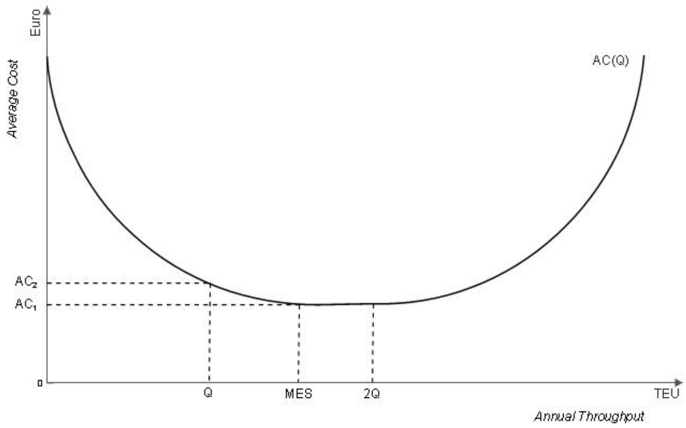


Figure 3: Illustrative minimum efficient scale for container terminals

The issues discussed above constitute the core of many legal disputes handled by competition authorities around the world. It inevitably raises further questions on the need/impact of intra-port competition and on the minimum efficient scale in terminal operations (see De Langen and Pallis, 2006). Hence, an evaluation of port authorities' practices in splitting and phasing of terminals demands more insight in the optimal/minimum size of a terminal from an operational cost perspective and in the cost/benefit balance linked to allowing new entrants vis-à-vis allowing incumbents to grow. As is illustrated in Figure 3 for a hypothetical container terminal, inappropriate splitting (e.g. from *2Q* to *Q*) could cause less efficient operations as a result of an increase in long-run average costs (*ceteris paribus*).

Port economic literature has addressed some of the above issues, to a limited extent and mostly in a qualitative manner. Defillippi (2004) focused on the benefits of turning mono-operator concession schemes to multi-operator schemes via subsidies, with the ultimate target being lower prices paid by users, whereas Flor and Defillippi (2003) focus on the role of the regulator to enact an access mandate appropriate for reversing monopolistic situations (i.e. by creating incentives to reach a Nash Equilibrium) and the relevant basic principles to arrive at the optimal charge. Yet quantitative approaches to the determination of the minimum efficient terminal scale and the shaping of intra-port and inter-port competition remain largely unexplored territory.

5. Risk and Investment Division

The infrastructure and superstructure investments linked to terminal development typically consist of:

- Land reclamation works, capital dredging and maintenance dredging;
- Quay-wall construction and maintenance;
- Mooring equipment and fenders;
- Investments in environmental mitigation and compensation (e.g. in the framework of the Bird and Habitat Directives of the European Commission);
- Electric installations and wiring on terminal;
- Paving of the terminal;
- On-terminal rail facilities;
- Roads on the terminal;
- Warehouses and technical buildings;
- Fencing and video surveillance (port security);
- Investments in truck gates;

- Office buildings;
- Terminal handling equipment (ship-to-shore cranes, yard equipment, etc.)

The port authorities will have to decide on the division of risks and investments related to the object of the concession. The general principle is that risks should be allocated to the party that can best manage them. In the port industry, commercial risks can best be managed by private operators. Contrary to other infrastructure investments, such as roads, terminal operators do have a substantial influence on demand, as they generally compete with others terminal operators, either in the same port or in different ports. As private operators are generally well positioned to manage these risks, all investments in terminal equipment are best left to these private firms.

For port infrastructure, this issue is more complex. In some cases, all risks are transferred to the private sector, and the awarding authority grants a concession to build and operate a terminal, the concessionaire has to make all investments, and consequently bears all the risks[6]. This option is especially attractive when the terminal to be concessioned is not part of a larger port expansion project, but will need to be built stand-alone. However, when a port authority develops a larger port expansion project, with various sites to be concessioned, the port authority is better placed to bear these risks.[7]

The awarding of a concession needs to address risk – as well as reward factors – in order to avoid unrealistic terms that may result in institutional entry barriers and a limited number of bids, or no bids at all. The type of risk and the hurdle rate of return needed for an operator to commit to a common user container terminal differ according to (a) the stage of development of terminal site leased (e.g. undeveloped, greenfield, improved site, fully developed) and (b) the nature of the market (e.g. established trade or not; type of trade; high or low entry barriers). The business case suggests that these situations (Table 1) determine whether, and under which conditions (i.e. length of concession, rates), private investments alone have the potential to be bankable or public funding is an essential mean to generate the interest of potential bidders (Bayne, 2006).

[6] In Asia governments often reduce the risk imposed to private operators that are granted concessions. The Korean government, for instance, used to work solely with a system of guaranteed minimum-maximum revenues for terminal operators investing in infra- and superstructure. Korean infrastructure projects generally include a government guarantee to cover for revenue drops below 90% of expected revenue in exchange for a revenue transfer to the government should operator revenues exceed a threshold set in advance. In the case of Busan New Port, however, the Pusan Newport Corporation negotiated a removal of this clause from the concession contract and accepted to bear full risk.

[7] This is for instance the case in the Maasvlakte II project in Rotterdam, which is a port expansion project of roughly 1000 ha with various different plots, designed for different types of users (container terminals, chemical plants and distribution activities).

Table 1: Risks involved for an Operator: Key Variables to be considered

Variable:	Stage of development of terminal site leased to private operator	Nature of the Market
Potential Conditions	Undeveloped site to develop infrastructure	No established regional trade; projections based on visionary cargo, or new free trade zone
	Greenfield site with infrastructure developed to site boundary	Established regional trade, substantial transhipment and lower barriers to entry
	An improved site with a quay line, paved yard but without buildings or handling equipment	Established hinterland general cargo trade but low container penetration factor
	A site with all civil works completed but supplies quay cranes and yard handling equipment	Established regional and national container trade but open to competition from other terminal operators within the same or nearby port
	A fully developed site including quay cranes but supplies yard handling equipment	Established container trade, need for facilities and high entry barriers

Source: Based on Bayne (2006); Drewry (2008).

This issue, along with the discussed issue of minimum efficient scale for multi-operator schemes, is closely related to the presence of positive general economic and financial conditions (i.e. low Inflation; GDP growth; user spending power, credit availability, low cost of money, and financial innovation), as well as the public funding for infrastructure projects. Dooms and Verbeke (2006) demonstrated that major port development projects in European landlord ports benefit to some extent from public financing, especially in the realm of basic infrastructure provision, including land reclamation, maritime access improvement and the construction of quay walls. Such public financing may not be desirable from an economic point of view, but is common practice in most port expansion projects around the world. The relation between public financing and the division of investment and risks in the ports industry, and the consequences of this relation for economic efficiency are themes to be addressed in academic literature,

6. Duration, Fees and Fee Structure

The port authority will have to make decisions on the duration, fees and fee structure before starting the awarding process. In most cases the term of the concession is determined by the port authority or government agency. In many parts of the world, legislators have developed rough guidelines on concession durations in view of safeguarding free and fair competition in the port sector. There are hardly any rules of thumb on the concession duration, and attempts made by the European Commission to create relevant rules have not met the agreement of all stakeholders (Farrell, 2001, Pallis, 2007). In general, the duration of the concession will vary with the amount of the initial investment required, the compliance with the development policy of the port and land lease and other easement rights. A port authority might opt for phased concession terms, with a base duration of e.g. 15 years and consecutive renewals (based on specific criteria) every 5 years. The port authority will also have to develop its views on a possible prolongation of the concession beyond the official term before entering into the granting procedure.

The duration of the agreement is of crucial importance both to terminal operators and port authorities. In general, long-term agreements allow private port operators to benefit from learning-by-doing processes and to achieve a reasonable return on investments (ROI). Port authorities try to find a balance between a reasonable payback period for the investments made by terminal operators on the one hand[8] and a

[8] As already discussed, this payback is strongly related to decisions regarding the risk undertaken by the operator. Inappropriate concession terms might prevent operators from making additional investments. This was recently

maximum entry to potential newcomers on the other. As long-term agreements act as entry barriers (see De Langen and Pallis, 2007), intra-port competition will only take place among the existing local port operators. However, even when concession periods are long, new players can still enter the market either through a merger or acquisition of a local operator or when a long-term concession or lease of a new terminal expansion is allocated to them (see Notteboom, 2002). While the importance of setting an optimal duration of a concession is well understood by academics and market players, it is highly remarkable that port-related academic literature does not offer any theories, guidelines and tools that can help in attaining an optimal concession term.

The port authority will also have to decide on the fees and the fee structure it will use throughout the awarding procedure (see also Notteboom, 2007). The awarding body will likely opt for centring bidding or negotiations (at least partially) on concession, but also has the option to work with a fixed, nonnegotiable fee structure. In setting the fees and fee structure, port authorities thus face a dilemma. On the one hand, port users demand a transparent, uniform and stable fee system. Such stability of the fee structure is important in view of investment decisions of the terminal operator. On the other hand, port authorities are tempted to apply the market mechanism in setting the fees for the use of valuable port land. In case fees are determined by the market, differences in fees among terminals in the same port can be substantial. Concession fees of two different container terminals in the same port, even operated by the same terminal operator, may then have significantly different concession fees. Moreover, the fee level for a new concession will largely depend on the market situation of the moment of awarding and the number of potential interested parties.

Granting authorities cannot really set a level playing field (in the sense of the same concession fees) as concessions can be, and are, sold. When the awarding authority grants a concession for a low price (in line with other concessions), the new owner may sell the terminal, and receive the rent of the low concession price. On the other hand, the mechanisms set by the awarding authority should avoid creating conditions for 'wildcat' operators, who might win at all costs and then renegotiate, or seek unforeseen post-award merging strategies.

The discussion of the pros and cons related to concession fee levels and the mechanisms set by the port, or any other awarding authority, the fees and companies strategies resulting from the competition between the bidding candidates (market mechanism) in order to win the concession, or any combination of these principles needs to receive in-depth attention in port economic literature.

7. Performance Targets and Final Asset Compensation

Port authorities have to design strategies in the pre-bidding phase on how they will deal with performance targets. The most common indicator relates to cargo throughput. The port authority can indicate upfront a minimum throughput to be guaranteed by the concessionaire (especially in case of existing berths/terminals). This should encourage the operator to market the port services to attract maritime trade and to optimize terminal and land usage. In case the terminal operator does not meet the objectives as set in the concession agreement, he will either have to pay a penalty to the port authority (e.g. a fixed amount per ton or TEU short) or, in the most extreme case, the concession will be taken away from him. Throughput guarantees in principle help to secure a reasonable level of land productivity, to reach high terminal utilization rates and can lower the entry barriers to newcomers (at least if the port authority follows a policy of retracting or reallocating certain parts of the terminal due to under-utilization). A port authority might also consider rewarding terminal operators who are doing much more than expected.

---

illustrated with the San Vicente International Terminal (Chile) where a 15+15 year concession term in combination with the agreed risk and investment spread appeared to be an unsound basis for investments in STS gantry cranes and RTGs. It was also illustrated in the recent decision of Hutchison Port Holdings to pull of its commitment to a 30-year concession of the Manta container terminal in Ecuador (even though it h as spent between $20m and $30m on developing the site as a hub port), reportedly claiming a unilateral government decision to restructure the required investment timetable to be followed by the terminal operator

Additionally, port authorities can follow different paths when it comes to dealing with the terminal superstructure at the end of the contract term. Although the decision on what to do with the superstructure is often made at the end of the concession term, it is useful for port authorities to develop views on this issue even as early as in the pre-bidding phase. Common approaches include the removal/destruction of the superstructure by the terminal operator at the end of the contract term or the transfer of the assets to the managing body of the port without any form of compensation. A port authority might opt for financial compensation of the terminal operator for the superstructure that was transferred at the end of the contract term. The design of this parameter of the awarding process has been neglected by research, even thought it is vital for avoiding the deterioration of the investment climate, and consequently the competitiveness of the terminal, in the later years of the concessions.

8. Conclusions

Terminal concessions in seaports demand a clear identification of the object of the concession, a choice of allocation mechanism to be used for granting those concessions, the determination of the concession term and concessions fees, as well as the inclusion of special clauses aimed at assuring that the terminal operator will act in the interest of the port authority and the wider community. The scope of the above issues needs to be determined prior to the start of the actual bidding procedures. In the later concession process, these issues will define the rules of the game. In this paper, we elaborated on the complex economic issues that are at stake during such a pre-bidding phase of the awarding procedure.

We have identified the following issues with a particular relevance for further port economic research:

- The impact of the type of terminal use on the awarding process;
- The problems related to the way sections of a terminal plot are phased and split, and their impact on the dynamics in intra- and inter-port competition;
- The division of risks and investments related to the object of the concession, and its links with institutional entry barriers;
- The determination of the duration of the concession, the concession fees and fee structure
- Modalities for the extension of a concession beyond the official term and final asset compensation in case the concession ends;
- The setting of performance targets.

This paper provides a conceptual discussion on each of the above issues and thus delivers input for the development of a much-needed academic research agenda on port-related concessions.

Acknowledgements

Christophe Theys gratefully acknowledges financial support from the Research Foundation Flanders (FWO). The participation of Athanasios A. Pallis has been supported by the Fulbright Foundation Research Scholars Program (2008-9).

References

Bayne, D. (2006), Port Economics-The Business Case, 3rd Intermodal Asia 2006, Brisbane, Australia.

Cariou, P. (2003), Dedicated terminals in container ports: a cost-benefit analysis, Research seminar: maritime transport, globalisation, regional integration and territorial development, Le Havre.

Cullinane, K. and Talley, W. (2006), Port Economics, Elsevier: London.

Defilippi, E. (2004), Intra-Port Competition, Regulatory Challenges and the Concession of Callao Port, Maritime Economics and Logistics 6: 279-311.

De Langen, P.W. and Pallis, A.A. (2006), Analysis of the Benefits of Intra-port Competition, International Journal of Transport Economics 33(1): 69-86.

De Langen, P.W. and Pallis, A.A. (2007), Entry Barriers in Seaport, Maritime Policy and Management 34(5): 427-440.

Demsetz, H. (1968), Why regulate utilities? Journal of Law and Economics, 11: 55-65.

Dooms, M. and Verbeke, A. (2006), The financing structure of large-scale, seaport development projects: a comparative analysis of recently completed and planned projects in the European Union, Proceedings of the IAME 2006 Conference (CD-Rom format), Melbourne.

Drewry, (2008), Global Container Terminal Operators: Annual Report, Drewry: London.

Farrell, S. (2001), If it ain't bust, don't fix it: the proposed EU directive on market access to port services, Maritime Policy and Management, 28(3): 307-313.

Flor, L. and Defilippi, E. (2003), Port Infrastructure: An Access Model for the Essential Facility, Maritime Economics and Logistics 5(2): 116-132.

Haralambides, H.E., Cariou, P. and Benacchio, M. (2002), Costs, Benefits and Pricing of Dedicated Container Terminals, Journal of Maritime Economics and Logistics 4(1): 21-34.

Musso, E., Ferrari, C. and Benacchio, M. (2001), Co-operation in maritime and port industry and its effects on market structure, Paper presented at WCTR Conference, Seoul.

Notteboom, T. (2002), Consolidation and contestability in the European container handling industry, Maritime Policy and Management 29(3): 257-269.

Notteboom, T. (2007), Concession agreements as port governance tools, In: Brooks, M.R. and K. Cullinane, (eds.), Devolution, Port Governance and Performance, Elsevier: London, pp. 449-467.

Pallis, A.A., (2007), EU port policy developments: implications for port governance, In: Brooks, M.R. and K. Cullinane (eds.), Devolution, Port Governance and Performance, Elsevier: London, pp. 161-176.

Pallis, A.A., Notteboom, T.E. and De Langen, P.W. (2008), Concession agreements and market entry in the container terminal industry, Maritime Economics and Logistics 10(3): 209-228.

Pallis, A.A. Vitsounis, T.K. and De Langen, P.W. (2009), Port economics, policy and management - Review of an emerging research field, Transport Reviews, forthcoming.

Theys, C., Notteboom, T.E., Pallis, A.A. and De Langen, P.W. (2009), The economics behind the awarding of terminals in seaports: Towards a research agenda. Submitted for publication in: Cullinane, K. (ed), The International Handbook of Maritime Economics and Business, Edward Elgar: Cheltenham.

Turner, H.S. (2000), Evaluating seaport policy alternatives: a simulation study of terminal leasing policy and system performance, Maritime Policy and Management 27(3): 283-301.

Van Niekerk, H.C. (2005), Port reform and concessioning in developing countries, Maritime Economics and Logistics 7(2), pp. 141-155.

World Bank (2007), The World Bank Port Reform Tool Kit (Second Edition), The World Bank Group: Washington D.C.

Air transport liberalization and its impacts on airline competition and air passenger traffic

Tae Hoon Oum[1], Anming Zhang[2] and Xiaowen Fu[3]

[1]University of British Columbia, and Embry Riddle University.
E-mail: tae.oum@sauder.ubc.ca
[2]University of British Columbia, Vancouver, BC, Canada, V6T 1Z2.
E-mail: anming.zhang@sauder.ubc.ca
[3]The Hong Kong Polytechnic University, Hong Kong.
E-mail: lgtxfu@polyu.edu.hk

Abstract

This study examines the impacts of liberalization policies on economic growth, traffic volume and traffic flow patterns, and investigates the mechanisms leading to those changes. Our investigation concludes that 1) liberalization has led to substantial economic and traffic growth. Such positive effects are mainly due to increased competition and efficiency gains in the airline industry, as well as positive externalities to the overall economy; 2) liberalization allows airlines to optimize their networks within and cross continental markets. As a result, traffic flow patterns will change accordingly. Strategic alliance is a second best solution and will have reduced when ownership and citizenship restrictions are relaxed; 3) there is a two-way relationship between the expansion of Low Cost Carriers (LCCs) vs. liberalization. The fast growth of LCCs leads to increased competition and stimulated traffic, calling for the removal of restrictions on capacity, frequency and pricing. In addition, development of LCCs in domestic market can promote liberalization policy by increasing the competitiveness of a national aviation industry. On the other hand, existing regulations hindered the growth of LCCs. Further liberalization is needed for the full realization of associated benefits.

1. Introduction and Background

International air transport operates within the framework of the 1944 Chicago Convention on international air transportation, under which airlines' commercial rights on international routes are governed by a complex web of more than 10,000 bilateral air services agreements (ASAs) between each country-pair. These ASAs regulate a wide range of conditions related to the provision of international air services. The WTO Secretariat (WTO 2006) identified seven features of ASAs as relevant indicators of openness for scheduled air passenger services. They are: 1) Grant of rights (air freedoms allowing airlines to provide services over designated markets), 2) Capacity clause (regulation on volume of traffic, frequency of service and/or aircraft types), 3) Tariff approval (whether fares need to be approved before applied), 4) withholding (which defines the conditions for the foreign carrier to operate, such as ownership and effective citizen control requirements), 5) Designation (which governs the number of airlines allowed to serve the market between two countries and on specific routes), 6) Statistics (that requires the exchange of operational statistics between countries or their airlines), and 7) Cooperative arrangements (which regulate the cooperative marketing agreements between airlines). After reviewing 2,299 ASAs in ICAO and WTO databases, Piermartini and Rousova (2008) indicated that the regulations used most frequently are on pricing, capacity and cooperative arrangements. In addition, while 60% of the ASAs allow multiple designations, the remaining 40% permit only single designation.

Since the deregulation of its domestic airline industry, the U.S. government has also pushed for the liberalization of international air markets. In 1979, the U.S. enacted the International Air

Transportation Competition Act, which formally laid down the principle of promoting liberalized bilateral ASAs with foreign countries. A major breakthrough was achieved when the first Open-Sky agreement was reached between the U.S. and the Netherlands in 1992, removing capacity and frequency constraints for aviation services between the two nations. As of 25 November 2008, the U.S. has open skies agreements with 94 countries in six continents, making it the open-skies hub nation of the world (US Department of State website, 2009).

During the period of 1988 to 1997, three air transport liberalization packages were implemented by EU countries, which created eventually a single aviation market for the EU community carriers by adding *Cabotage* rights in 1997. As of 11 January 2007, a total of 66 countries in all continents have recognized EU common market in their ASAs, allowing European air carriers to operate flights between any EU member states and these countries. In April 2007, the EU-US Open Aviation Agreement (OAA) was signed and went into effect on 30 March 2008. While similar agreements are being negotiated with other nations, efforts are made to further liberalize the international aviation market, which would remove remaining constraints such as ownership restriction.

Bilateral air services agreements remain the primary vehicles for liberalization of international air transport services for most countries. During the past decade, about one thousand bilateral air services agreements (including amendments and/or memoranda of understanding) were reportedly concluded. Over 70 percent of these agreements and amendments contained some form of liberalized arrangements, such as expanded traffic rights (covering Third, Fourth and in some cases Fifth Freedom traffic rights), multiple designations with or without route limitations, free determination of capacity, a double disapproval or free pricing regime, and broadened criteria of airline ownership and control. As the airline business evolves, some of the recent bilateral air services agreements have included provisions dealing with computer reservation systems (CRSs), airline codesharing, leasing of aircraft and intermodal transport. One notable development is the considerable increase in the number of bilateral "open skies" air services agreements, which provide for full market access without restrictions on Third, Fourth and Fifth Freedom traffic rights, designation, capacity, frequencies, codesharing and tariffs. As of February 2008, 142 bilateral "open skies" agreements have been reportedly concluded worldwide.

Despite of the fact that many liberalization agreements have been reached over the years, liberalization of the international aviation market remains a formidable challenge. Even with strong political will, the negotiation of liberalizing ASAs remains to be a lengthy process full of disagreements and bargaining. Many of the difficulties in liberalization efforts can be ascribed to stakeholders' different expectations on the effects of alternative policy / agreement scenarios. The resulting uncertainty of liberalization has prevented many governments from adopting substantial regulatory changes, and has given certain interests groups including national flag carriers strong influence over the negotiation process. Therefore, there is a need to review the actual effects brought by the liberalization process worldwide, and investigate the mechanisms leading to those changes. These efforts would, of course, facilitate policy makers in their efforts to address future liberalization initiatives.

This study aims to achieve the above objectives by investigating the following issues: Section 2 reviews the economic effects of liberalization on the air transport industry and economy. Section 3 studies the airline network competition and restructuring process with deregulation and liberalization, whereas Section 4 examines the impacts of low cost carriers on airline networks and aviation policy. The last section summarizes and concludes the study.

2. Economic Effects of Air Transport Liberalization

The evolving liberalization of international air transport regulation since the mid-1990s has played an important role in the growth of air transport industry by providing a favourable regulatory environment. Worldwide, the total number of annual passengers has grown by 46 percent in the past ten years, from 1.457 billion passengers to 2.128 billion per year (ICAO, 2007). It is estimated (ICAO

Secretariat, 2007) that, in 2006, about 31 percent of the country-pairs with non-stop passenger air services and about 49 percent of the seat capacity were offered between countries that have embraced liberalization either by bilateral "open skies" ASAs or by regional / plurilateral liberalized agreements and arrangements (compared with less than 4 percent and about 20 percent respectively in 1995, and about 16 percent and about 42 percent respectively in 2000). Numerous reports and papers from academia, governments and industries, confirmed that the liberalization efforts had brought significant welfare gains and economic growth world wide.

This section provides an overview of the economic effects of regulation and liberalization. A short summary of the origin and results of regulation is first given. We then review the the economic impacts of air liberalization on the aviation industry. Finally, a discussion on the relationships between air transport liberalization and overall economy is provided. While this paper focuses on the liberalization of international market, the U.S. regulation / deregulation process has also been discussed where appropriate since the regulation / deregulation practice in this market had served much as a prototype in the industry. In addition, the U.S. market has been extensively studied such that rich results and findings have been obtained.

2.1. Rationale and the Economic Effects of Air Transport Regulation

After the World War I, some state-owned enterprises and private airlines began to offer commercial air transport services to the public. However, with low demand and high risk of operation, commercial air transport would not have been sustainable without government support. As a result, the Kelly Air Mail Act of 1925 was passed in the U.S., allowing the Post Office to subsidize private air mail carriage by awarding contracts with payment exceeding air mail revenue on the routes. To oversee such a system, the Civil Aeronautics Board (CAB) was created as a regulator by the Civil Aeronautical Act of 1938. Charged with "the promotion, encouragement and development of civil aeronautics", the CAB aims to eliminate "unfair or destructive competitive practice" by regulating entry, rate levels and structures, subsidies and merger decisions (Caves 1962, Levine 1965, Borentein and Rose 2007).

Quite a few studies (Levine 1965, Jordan 1970 and Keeler 1972) found that the regulations imposed by CAB resulted in limited competition and high fares. Levine (1987) pointed out that fares in unregulated intra-state routes tend to have relatively high service level and load factors with remarkably lower fares. High fares maintained by regulation did not, however, lead to high industry profit. Airlines engaged in non-price competition with inefficiently high service quality (e.g., flight frequency, in-flight amenities) and newer, larger aircraft. This reduced airlines' load factor while increased average costs. In the years just prior to deregulation, the industry average load factors fell below 50% (Borentein and Rose 2007).

Similar pattern has been observed in the international market. The regulatory system on international air transport was formalized in the 1944 Chicago Convention. The United States, which was effectively the only country with sufficient financial resources, a large aircraft fleet and expertise after the World War II, attempted to promote competition on a multilateral basis. However, such an effort was not successful. Following the precedent of the first US-UK bilateral agreement in 1946 ("Bermuda I"), ASAs generally regulate services (passenger, cargo) and routes to be operated, and stipulate fare-setting mechanisms and capacity limit. In one sense, this bilateral system was an interesting solution to a competition issue: that is, countries at the time feared unilateral application of monopoly power by a trading partner. However, it introduced another set of competition problems by constraining entry, especially to routes between countries (Warren and Findlay, 1998). All these regulations have greatly hindered the growth of international travel. Such a situation only began to change gradually with the passage of the 1979 US International Air Transportation Competition Promotion Act (IATCPA), after which the U.S. began to explicitly promote liberalized bilateral ASAs with foreign countries.

As evidenced by the outcomes in both domestic and international markets, regulations were introduced with good intentions and objectives. Over time, however, policy makers found themselves drifting away from these original targets, with more and more regulations imposed to correct the undesirable effects. Many governments have realized that a better solution is to deregulate / liberalize the market, which have brought very positive economic effects to the air transport industry as well as the overall economy.

2.2. Economic Effects of Liberalization on the Air Transport Industry

While the net effects of previous liberalization events vary across the markets, there are some common changes brought to the air transport industry:

2.2.1. Increased Competition, Reduced Price and Traffic Stimulation Effects:

Most liberalization efforts have brought in significant traffic growth. Such traffic growth was mainly driven by two factors: First, liberalization removes constraints on pricing, route entry, service capacity and cooperative arrangements among alliance members. This allows airlines to compete more effectively and operate more efficiently, which reduces price and increases service quality in terms of flight frequency, frequent flier programs, etc. As a result, passenger traffic can be stimulated substantially. Secondly, liberalization allows airlines to optimize their network configuration. The implementation of hub-and-spoke networks enabled carriers to link small markets with their hub airports, expanding air services to new destinations. Maillebiau and Hansen (1995) developed a translog air travel demand function in a single aviation market in order to forecast the passenger increase between U.S. and five European countries: UK, France, West Germany, Netherlands and Italy. They estimated that the traffic growth from liberalization is 56% with an average benefit of $585 per passenger. Their results also found a decrease in airline yield of 35% and a 44% increase in accessibility.

This is not a surprising result. Button (1998) found that following the U.S. deregulation, during 1978-1988, passenger traffic increased by 55 percent while scheduled revenue passenger-miles grew by over 60 percent. The real costs of travel fell by about 17 percent on major routes.[1] Morrison and Winston (1986) estimated that the U.S. deregulation yield welfare gains of $6 billion to passengers and profit gains about $2.5 billion to stakeholders of carriers (including various labor unions). Table 1 compares the changes in prices of air travel vs. other goods and services in the United States during the 1978-2006 period. It shows that both domestic and international air services are two of the four items with the lowest nominal price increases during the 28-year period: 1.5-1.6 times the price of 1978 for air travel while college tuitions (private and public) increased by the factor of 7.5-8.5 times the 1978 levels.

[1] Borenstein and Rose (2007) found that between 1976 and 1986, the U.S. average domestic passenger yield declined in real terms at a rate of 3.4% per year, while revenue passenger miles increased at a rate of 8.2% per year. However, they pointed out that the price effects of the U.S. deregulation may have been overestimated. Instead, a major change was an increase in price dispersion. Price dispersion within carrier – routes more than doubled between 1979 and 2001.

Table 1: Price Changes of Air Travel versus Other Goods and Services

ITEM-U.S. Good or Service	Unit	1978	1990	2006	Growth
College Tuition: Public	Year	$688	$1,908	$5,836	8.5x
College Tuition: Private	Year	$2,958	$9,340	$22,218	7.5x
Prescription Drugs	Index	61.6	181.7	363.9	5.9x
New Single-Family Home	Home	$55,700	$122,900	$246,500	4.4x
New Vehicle	Vehicle	$6,470	$15,900	$28,450	4.4x
Unleaded Gasoline	Gallon	$0.67	$1.16	$2.59	3.9x
CPI (Urban-All Items)	CPI-U	65.2	130.6	201.6	3.1x
Movie Ticket	Ticket	$2.34	$4.22	$6.55	2.8x
First-Class Postage	Stamp	$0.15	$0.25	$0.39	2.6x
Whole Milk	Index	81.0	124.4	181.6	2.2x
Grade-A Large Eggs	Dozen	$0.82	$1.01	$1.31	1.6x
Air Travel: International	Mile	7.49¢	10.83¢	11.85¢	1.6x
Air Travel: Domestic	Mile	8.49¢	13.43¢	13.00¢	1.5x
Television	Index	101.8	74.6	22.3	0.2x

Sources: General Accountability Office (GAO, 2008), Airline Industry: Potential Mergers and Acquisitions Driven by Financial and Competitive Pressures, GAO-08-845 July 31, 2008.

2.2.2. Productive Efficiency Improvement

Liberalization has improves the productive efficiency of the airlines industry via several ways: First, liberalization allows airlines to optimize their network and pricing strategy. This improves airlines' operation efficiency and average load factor. As a result, average costs have been reduced steadily. Secondly, the increased competition following liberalization forces airlines to relentlessly improve their productive efficiency. Less efficient airlines are either merged or bankrupted, while new business models and innovations (e.g., low cost carriers, e-tickets and self service check-in) are nurtured when firms drive to achieve competitive edge. Oum and Yu (1998), Oum, Fu and Yu (2005) found that after deregulation, many remaining U.S. carriers have achieved global leadership in cost competitiveness. Fethi et al. (2000) found that the EU liberalization have improved airlines' efficiency significantly.

2.2.3. Effects on Employment in the Aviation Industry

As one would expect, the rapid growth brought by liberalization must lead to additional jobs in the aviation sector. Button (1998) estimated that with the substantial growth following the U.S. deregulation, the employment in the air transport industry increased by 32 percent during the 1978-1988 period. InterVISTAS (2006) estimated that the creation of the Single European Aviation Market in 1993 produced about 1.4 million new jobs in aviation and related industries; the 1998 UK – UAE (United Arab Emirates) liberalization created over 18,700 full-time equivalent positions in the UK side; and the 1986 Germany – UAE liberalization created 745 new full time positions in UAE and 2,600 new jobs in Germany.

It should be noted that the job creation process sometimes is accompanied with job relocation, when firms outsource certain functions to more cost effective regions. For example, with the liberalization / formation of European single aviation market, Lufthansa (LH) began to outsource certain functions to Eastern European countries. In 2005, LH built a new shared customer services center in the Czech Republic, and set up maintenance facilities for heavy checks in Hungary. The airline also plans to move most of its accounting and purchasing operations to Poland. In addition to cost cutting, outsourcing strategies are likely driven by the company's desire to explore overseas opportunities. Outsourcing operations abroad will reduce domestic production. However, a more competitive airline in the global market will achieve more service export for the country (e.g., Clougherty and Zhang, 2008).

2.3. Air Transport Liberalization and Overall Economy

There is a two-way relationship between air transportation and the overall economy. It has been well recognized that air transport and logistics, as other transport services, are so called "derived" demands. They are usually purchased as inputs or intermediate products for the consumption / production of some other services: passengers purchase air service because they need to go to the destination for business or leisure, whereas cargos are shipped such that they can be consumed / processed in the destination. Therefore, the demand for transport services is largely driven by the overall economy. Boeing (2008) attributes about two-thirds of traffic growth to the GDP growth, and the rest to other factors such as increasing trade, lower costs and improved services. ICAO estimated the income elasticity for air travel to be 1.27. That is, *ceteris paribus*, a 1-percent increase in GDP will lead to a 1.27-percent increase in air travel.

While air transport is, on one hand, driven by the global economy, it is, on the other hand, an important driver to the global economy. International Air Transport Association[2] (IATA) noted that air transport directly employs four million people worldwide and generates $400 billion in output. In addition, the efficiency and quality improvements in air passenger services contribute to the growth in sectors such as hotel and tourism. The free flow of people and information, together with improved air cargo operations, promote trade and improve the efficiency of the overall economy. That is, the aviation sector imposes significant positive externalities to other industries, contributing to economic and employment growth. Button et al. (1999) examined the link between high-tech employment in a region and whether the region is served by a hub airport. Using data from 321 U.S. metropolitan areas in 1994, the analysis found that the presence of a hub airport increased high tech employment by an average of 12,000 jobs in a region. Irwin and Kasarda (1991) examined the relationship between the structure of airline networks and employment growth in 104 metropolitan areas in the United States. They found that expansion of the airline network serving a region had a significant positive impact on local employment. The effect was particularly significant in the service sector. Furthermore, analysis using nonrecursive models confirmed that increases in the airline network were a cause rather than a consequence of this employment growth. In addition to job creation, air transport facilitates commerce communication and labor mobility. Button (2006) pointed out that in United States and Europe, more than 40% of air travels are for business purposes. The remaining trips are either for leisure or for visiting friends and relatives. Leisure travel promotes the hotel and tourism sectors, while visiting friends and relative trips provide the basis upon which social ties are retained and, as such, allow for an efficient and integrated labour market.

Air transport is ideal for the coordination of global supply chains, and thereby, improves the overall efficiency of the economy. As firms source around the world for most favorable inputs such as labor, land, technology and capital, manufacturing and factory locations can be sparsely distributed. Hummels (2006) found that the elasticity of air shipping costs with respect to distance declined dramatically, from 0.43 in 1974 to 0.045 in 2004. That is, doubling distance shipped caused a 43% increase in air shipping costs in 1974, but only a 4.5% increase in air shipping costs in 2004. As a result, the average air shipment is getting longer and the average ocean shipment is getting shorter.[3] Recent papers by Aizenman (2004) and Schaur (2006) have argued that air shipping may be an effective way to handle international demand volatility. Because air shipments take hours rather than weeks, firms can wait until the realization of demand shocks before deciding on quantities to be sold. That is, air shipping provides these firms with a real option to smooth demand shocks.

Same as other shipping modes, the efficiency and quality improvements of air transportation promote trade and economic growth. Two major barriers for trade are cost and time related to transportation. Limao and Venables (2001) find that a 10% increase in transport costs reduces trade volume by 20%.

---

[2] IATA 2005 annual report.

[3] Hummels (2006) pointed out that ocean shipped cargo traveled an average of 2919 miles in 2004, down from 3543 miles in 1975. In contrast, air shipped cargo traveled an average of 3383 miles in 2004, up from 2600 in 1975.

Recent studies find that a 10% increase in time reduces bilateral trade volumes by between 5% and 8% (Hausman et al., 2005; Djankov et al., 2005). While air transport is clearly superior to other shipping modes in terms of time, its perceived cost disadvantage has been reduced over the years. Swan (2007) found that since 1970, both price and production cost for air travel have been declining at about 1% annually. As shipments are of higher value and lighter weight, the *ad valorem* cost of air freight, i.e., the transport cost needed to move a dollar of goods, is also decreasing. Harrigan (2005) estimated that the relative cost of air transport has declined by 40% between 1990 and 2004. As a result, air cargo is of growing importance in cargo logistics, accounting for about 40% of international trade by value. Many countries have chosen to locate special economic zones and high tech parks near airports.

Some nations, such as the Netherlands and Singapore, achieved rapid economic developments by leveraging on their liberalized transport systems. Compared to its European neighbours such as France and Germany, Netherlands has a relatively small domestic market. Nevertheless, the country have been aggressive in liberalizing its transport sectors: in 1992 it signed the first open-sky agreement in the world with the U.S., promoting Schiphol airport as a major gateway for cross-Atlantic traffic, while facilitating its flag carrier at the time, KLM, to further expand its network coverage in Europe and North America. These efforts, together with its superior transport infrastructures,[4] have made the Netherlands not only a major European aviation hub nation, but also an ideal place to establish European Distribution Centers (Oum and Park, 2004). In terms of value, only 5% of the express cargo and retail logistics handled in the Netherlands are for local consumption (Datamonitor, 2005). With the establishment of their European Distribution Centers, many companies have chosen to also locate their billing centers, service depots, research centers or even European headquarters in the country. The well developed transport and logistics sector in the Netherlands has clearly enhanced the overall competitiveness of its economy.

3. Airline Network Competition and Liberalization

In markets not yet liberalized, there can be many constraints on airlines' network configuration. Bilateral air services agreements (ASAs) between two countries limit airports and route access, flight frequency and seat capacity. These regulations prevent carriers from optimizing their overall networks. The limitations imposed with a third country (i.e., limitations on beyond rights such as 5th freedom) will further constrain a carrier's network structure in a region. As many theoretical and empirical studies found, when these constraints are removed, airlines often choose to reconfigure their networks to achieve various objectives: to improve cost efficiency by exploiting "economies of traffic density"[5], to enhance service quality by initiating direct flights and/or by increasing flight frequency[6], to price more aggressively or to compete more strategically[7]. Many of these objectives are achieved by streamlining a carrier's multi-hub network.

3.1. Effects of Hub-and-spoke Networks and Airline Network Competition

The emergence and prevalence of hub-and-spoke network is one of the most common developments in deregulated markets, especially for airlines endowed with access rights to a single large market such as the United States and European Single Aviation Market. The formation of a hub-and-spoke network can affect both demand and cost.

[4] The Netherlands has the largest marine port in Europe (Rotterdam), superior inland river shipping to Germany and France, and extensive high speed rail and road connections to Western Europe.

[5] See, e.g., Caves et al. (1984) and Brueckner and Spiller (1994). Traffic density is calculated by dividing the total traffic volume by the carrier's network size. Network size is usually defined as the number of origin-destination pairs served by the carrier, or the number of nodes connected in its network.

[6] See, e.g., Morrison and Winston (1987), Berechman and Shy (1998), Brueckner and Zhang (2001), and Brueckner (2004)

[7] Borenstein (1989), Spiller (1989), Berry (1990), Bittlingmayer (1990), Brueckner and Spiller (1991), Brueckner et al. (1992), Zhang and Wei (1993), Oum et al. (1995), Zhang (1996), and Hendricks et al. (1997 and 1999)

The effect of hubbing on costs has been extensively studied in the literature (e.g., Caves et al. 1984, Brueckner and Spiller 1994, Hendricks, et al., 1995, 1999). Costs can go down due to higher traffic densities in hub-and-spoke (HS) operations than in fully connected (FC, or point-to-point) operations, although these cost savings might be offset by the travelers' circuitous routings via hubs.

Hubbing can also affect demand (which, in turn, affects revenues and profits) with its effect on passenger travel time and schedule delay time. Compared to non-stop services, an HS network increases the average passenger's travel time due to the extra connecting time at hubs and the circuitous routing of passenger trips. On the other hand, HS reduces a passenger's schedule delay time – i.e. the time between his desired departure and the actual departure time (Douglas and Miller, 1974) – by offering increased flight frequency. In addition, a HS network allows an airline to serve many additional city-pairs when a new spoke route is added to the network (Oum and Tretheway, 1990).

The hub-and-spoke network is an efficient way to serve destinations over large spatial distance. Airbus (2007) pointed out that, one source of connecting traffic is passengers who could in fact fly directly if they wanted to. For example, in 2006, 20% of those flying between Europe and Asia selected a connecting route, even though they could have taken a direct service. There are several reasons for this. Many passengers prefer connecting services to direct service due to the wider variety of schedules offered at major hubs, either in terms of flight frequency or number of destination cities. Airlines often offer lower prices for connecting services, which is a by-product benefit from global airline alliances (e.g., Oum et al., 2000). Passengers may also choose to fly via a hub to take advantage of a stay-over at an intermediate stop.

Airlines may form hub-and-spoke networks as a strategic response to competitors rather than to simply save costs. Oum et al. (1995) show that hubbing can be used as both an offensive and a defensive strategy in airline network rivalry. Another major benefit of HS networks is associated with a carrier's dominance at its hub airports, which allows it to achieve substantially higher mark-up above costs. Such a benefit to the dominant carrier is referred to as the "hub premium" in the literature, as has been confirmed in numerous studies including Borenstein (1989), Dresner and Windle (1992), Morrison and Winston (1995), Lee and Prado (2005), GAO (1989, 1990), Lijesen et al., (2001), DOT (2001). Such a benefit gives airlines a strong incentive to dominate an airport. Table 2 shows that during the fifteen years after the U.S. Domestic Airline Deregulation in 1978, all major network carriers have strengthened their market shares at their respective hubs.

In conclusion, the prevalence of HS networks after airline deregulation can be explained by cost advantages in production (economies of density) and/or revenue advantages achieved via demand stimulation (network complementarity). Even when there is neither cost nor revenue advantage, the threat of potential entry alone can give rise to an HS network as opposed to an FC network. Zhang (1996) further argues that, for strategic reasons, competing airlines would choose to develop HS networks using different hub airports.

Table 2: Increased Share of the Dominant Carriers at Concentrated Hub Airports, 1978-1993 Period

	1978		1993	
Airport	**Share**	**Carrier**	**Share**	**Carrier**
Atlanta	49.7	Delta	83.5	Delta
Charlotte	74.8	Eastern	94.6	USAir
Cincinnati	35.1	Delta	89.8	Delta
Dayton	35.3	TWA	40.5	USAir
Denver	32.0	United	51.8	United
Detroit	21.7	American	74.8	Northwest
Greensboro	64.5	Eastern	44.9	USAir
Memphis	42.2	Delta	76.3	Northwest
Minneapolis-St. Paul	31.7	Northwest	80.6	Northwest
Nashville	28.5	American	69.8	American
Pittsburgh	46.7	Allegheny	88.9	USAir
Raleigh-Durham	74.2	Eastern	80.4	American
St. Louis	39.4	TAW	60.4	TWA
Salt Lake City	39.6	Western	71.4	Delta
Syracuse	40.5	Allegheny	49.5	USAir

Source: Morrison and Winston (1995)

Upon the deregulation in 1978, major US carriers began to strategically plan their networks to strengthen their dominance in existing hubs and to expand continental market coverage. Such a process was accompanied with massive mergers, acquisitions and liquidations. For example, many airlines based in Central and Eastern United States acquired carriers based in Western United States.[8] This resulted in a massive consolidation of the industry which reduced the number of trunk airlines from over 25 before the 1978 deregulation to 6+ major national network carriers. As a result, all of the national network carriers have built up multiple hub networks in the United States.

While network carriers often utilize multiple hubs, they can not afford to have more than one hub in a region. Airneth (2005) observed that the closest distance between two major hubs in a successful dual-hub system in the United States is 900km, the case of Northwest's Minneapolis-St. Paul and Detroit. In 2008, Delta Airlines acquired Northwest, with a plan to reduce or close the hub functions of Memphis (NW's hub) and Cincinnati (Delta hub), since they are too close to Atlanta and Detroit hubs of the combined carrier. Such a restructuring would result in a network of four hubs in North America: Atlanta, Detroit, Minneapolis-St. Paul, and Salt Lake City. U.S Airways has also reduced drastically the hub functions of Pittsburg in the last five years since it is close to its own hub in Washington Reagan International Airport.

3.2. Airline Network Development and Policy Implication

If domestic and international markets are both fully deregulated, network carriers would be able to expand their multi-hub networks to global markets. Intercontinental mergers and acquisitions are likely to occur since they are usually cheaper and less time-consuming than developing a carrier's own network in other continents (Oum et al., 1993). The current discussions between European Commission and the U.S. on deregulating foreign ownership of airlines would have similar effects as a complete deregulation. In fact, such an agreement aiming to dismantle the limitations on foreign ownership may eventually lead to a complete dismantling of the bilateral ASA system.

Under the gradual liberalization scenario, there will be several driving forces for airlines to restructure their networks. First, full service airlines (FSAs) will consolidate via merger and acquisitions in domestic and intra-continental market, in order to strengthen their network and market positions in a

[8] For example, Delta acquired Western Airlines in order to expand their market coverage in western United States and to secure Salt Lake City as its western hub. American Airlines strengthened its Dallas –Ft Worth hub and acquired Air California. US Air acquired Piedmont and Pacific Southwest. On the other hand, Northwest acquired Republic in order to increase dominance of its Minneapolis-St. Paul hub and surrounding markets.

continent. Second, across different continents the next wave is to strengthen network and market linkages via global strategic alliances (Oum et al., 2000), as evidenced by the formation and growth of major airlines alliances such as STAR, SkyTeam and OneWorld. Since the airlines within each Strategic Alliance Group will retain their own identity, they will structure their networks in such a way to maximize their own profits. As a result, these airlines' international and intercontinental networks will be influenced heavily by the structure of their domestic/continental networks.

Previous alliance studies suggest that international alliances improved partners' operations and service quality, lowered fares and grew the market.[9] However, the future of these global alliances is not crystal clear. Since the existing alliances grew under a web of restrictive bilateral ASAs which barred cabotage and foreign ownership, they represented a "second best" approach to the realization of inter-firm synergies on both the cost and demand sides. (In effect, such realization is constrained by the existing restrictive international regimes; as a consequence, the observed benefits from alliances are lower than their full potential.). Therefore, the future growth of global airline alliances would be limited, if not approaching to zero, under a fully liberalized (both domestically and internationally) air transport market.

When restrictions on route entry, capacity and frequency are dropped in domestic and intra-continental markets, network reconfigurations are likely to be different among United States, Europe and Asia. The US carriers have complete freedom to restructure their domestic networks since 1978. Transborder open skies in Europe began in 1993, and the complete single market (including cabotage rights for all EU carriers) began in 1997. As a result, European airlines had less time to adjust their networks compared to their peers in the U.S. Most cross-border markets in Asia are still heavily regulated. As a result, most of the Asian carriers serve their principal city markets, rather than using their super airports as hubs. Such network pattern can be confirmed as in table 3: Many US airports serve as real hubs, with over 50% connecting ratios. In Europe, only Frankfurt airport has more than 50% connecting ratio. All other airports including London, Amsterdam and Paris have less than 50% connecting ratios. The Asian airports perform even less hub functions. Even the most active hub airport in East Asia, Hong Kong, has only slightly higher than 30% connecting ratio. Many Asian carriers are taking advantage of the restrictive international regulatory regime: with capacity restricted, airlines are able to charge higher prices to local traffic. Therefore, they have less incentive to use the scarce intra-Asia capacity to attract connecting passengers. In 2007, Narita and Incheon have only 17% and 12% connecting ratios respectively. As the international liberalization advances further and perhaps more rapidly in the future, Asian network carriers are likely to restructure their network and traffic routing patterns in such a way to increase hub functions of their major airports.

[9] For instance, the global alliances have facilitated competition among alliance networks, which significantly improved the efficiency of the international interlining market. Brueckner and Whelen (2000) found that fares are about 18-20 percent lower on international alliance and interlining routes.

Table 3: Percentage of Connecting Passengers at Major Airports, 2007

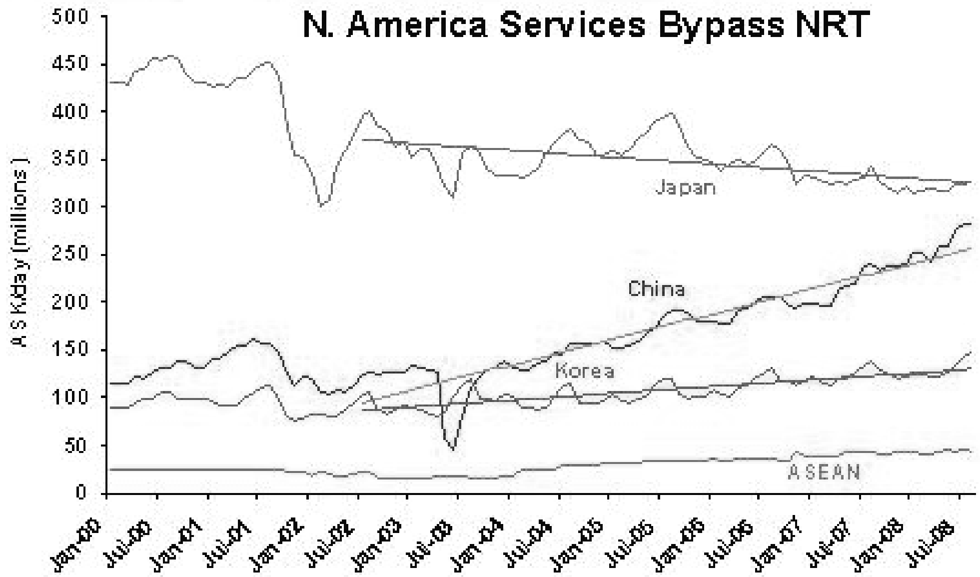
Airport	% Connecting Passenger	Airport	% Connecting Passenger
North America		**Europe**	
ATL	64.0%	AMS	41.3%
CLT	30.0%*	ARN	22.0%
DEN	43.0%	ATH	21.0%*
DFW	60.0%	CDG	32.0%**
DTW	48.4%	CPH	27.8%
EWR	30.6%	FRA	53.0%
IAD	20.7%*	LHR	36.0%**
IAH	51.2%	PRG	20.3%*
JFK	30.8%	VIE	31.9%
LAS	12.9%**	ZRH	33.8%*
LAX	3.9%		
MDW	25.0%**	**Asia**	
MEM	63.3%	CAN	20.1%
MIA	39.0%	HKG	33.3%
MSP	47.3%	ICN	12.1%
ORD	68.0%**	NRT	17.2%
PHL	37.0%*	PEK	n.a.
PIT	14.0%	PVG	16.3%
SEA	28.0%	TPE	11.0%
SFO	24.9%		
SLC	50.4%		
STL	23.9%		

Source: ATRS Airport Benchmarking Report, 2005-2007
* 2006 data ** 2005 data

The varying stages of openness in global aviation market imply that airline networks, and accompanying traffic flows, will experience shift in spatial pattern and market power. For example, Hong Kong had been much more liberalized than the neighboring economies including mainland China, Taiwan, Thailand and Vietnam, etc. Together with its fast growing economy, Hong Kong had secured leaderships for its airport and marine port in the region. However, with the gradual liberalization of mainland China, Hong Kong airport's hub status is facing serious challenge from nearby airports such as Guangzhou and Shenzhen. Since South Korean air carriers lost most of their domestic markets to high speed rail (KTX), the country has no choice but to adopt Singapore style policy to promote open skies regimes internationally, especially with China, Japan and Southeast Asian countries. It is noteworthy that South Korea has open skies ASA with the United States since 1998.

Due to historical reasons, Japan gave major bases of operations at Narita and other major Japanese airports to United, Northwest and Federal Express, and opened its markets to other US carriers substantially. However, Japanese government now realizes that the importance of economic integration with China and South Korea, and thus, the open skies regime in Northeast Asia is a more urgent task than signing open skies with the United States or Canada. Since both Tokyo-Narita and Tokyo-Haneda airports are expected to have substantially more slots in 2010, Japan expects to allocate a lion's share of these increases to Asian carriers, especially carriers of Northeast Asian subcontinent. An issue that worries Japanese government a lot is that there has been an increasing trend that Northeast Asia - North America air traffic are bypassing Tokyo-Narita (NRT) as shown in Figure 1.

Figure 1: North American – East Asian Traffic Diversion



Source: Swan (2007)

Even for countries with deregulated air transport markets, it is important to maintain their leadership in liberalization, thus that to keep their aviation sector competitive in the global market. Singapore, for example, has been working hard to maintain its leadership in the region in terms of air transport liberalization. As of 2006, Singapore has signed over 90 ASAs with other countries, compared to the 57 ASAs signed by Hong Kong.[10] Singapore also reached open-skies agreements with the U.S., New Zealand and the United Arab Emirates. In June 2006, the country became the first Asian nation to sign an open-skies agreement with the EU, which allows Singapore Airlines to fly anywhere within the 27 EU-nation bloc. Such aggressive and determined liberalization policy had helped the nation to maintain the competitiveness of its airports and airlines.

4. The Impacts of Low Cost Carriers and Implications on Aviation Policy

A strong trend that emerged with deregulation and liberalization in the United States, Canada and Europe was the disappearance of weaker airlines through bankruptcies or mergers but at the same time the birth of upstart competitors. Well-established brands like PanAm, Eastern Airlines, TWA and Canadian Airlines International disappeared, while LCCs such as Southwest and several new brands (e.g., JetBlue, Westjet, Ryanair, EasyJet) emerged and prospered. As pointed out by Transportation Research Board (1999), "Probably the most significant development in the U.S. airline industry during the past decade has been the continued expansion of Southwest Airlines and the resurgence of low-fare entry generally."[11] The "Southwest effect" – i.e., a rapid increase in traffic volume and a simultaneous fall in fares on routes where, or close to where, Southwest Airlines operates – has become widely known (US DOT, 1993; Richards, 1996). The price effects of LCCs were empirically estimated by, among others, US DOT (1993), Windle and Dresner (1995), Dresner et al. (1996), and Morrison (2001).[12] Franke (2004) suggested that Europe has a similar "Ryanair effect," whereas Zhang, et al. (2009) suggested that the "Southwest effect" might also exist in Asia.

LCCs such as Southwest Airlines and Ryanair grew under a deregulatory domestic environment – after the EU integration in the mid-1990s, the EU internal market has become a "domestic" market.

[10] It should be noted that the number of ASAs signed is not the sole indicator of market openness, since some ASAs signed may not be active. In addition, compared to other Asian economies Hong Kong has much better access to mainland China, a large and fast growing market.

[11] This statement was also quoted at the beginning of Morrison (2001).

[12] See Tretheway and Kincaid (2005) for a literature review on the effect of LCCs on air fares in the US.

In Asia, entry of LCCs was facilitated by domestic deregulations as well. While deregulation and liberalization have facilitated the growth of LCCs, the LCC experience has also promoted policy reform and liberalization. Until 1978, the US airline industry was regulated by the Civil Aeronautics Board. It was mainly through the experience of unregulated Southwest Airlines – which offered lower fares for intra-state (Texas) services than comparable regulated services between states – that the deregulation of market entry commenced in 1978 with the passage of the "Airline Deregulation Act" (Levine, 1987; Morrison, 2001). This has in turn stimulated Southwest's domestic expansion as the state borders did not matter any more.

Another case in which LCC experience stimulates policy liberalization is the ASEAN (the Association of Southeast Asian Nations) region where significant progress has been made lately. In July 2007, ASEAN countries reached an agreement under which unlimited flights between capital cities in ASEAN will start at the end of 2008. Furthermore, it was expected that ASEAN nations will sign an "open skies" agreement as early as December 2008 (Asia Times, 2008). These positive policy developments are due mainly to the positive effects of liberalization, both domestically and regionally, and of emerging LCCs. Consider the case of Malaysia. After maintaining a strict closed-skies aviation policy for many decades, more recently Malaysia has seen a boom in air traffic growth due to greater domestic competition led by AirAsia. This, together with the success of other regional LCCs, has prompted the Malaysian and other ASEAN governments to push for a more liberalized regulatory regime (Asia Times, 2008). Another major motivation for liberalization in these Southeast Asian countries is to boost tourism and business travel after the devastating Asian financial crisis of the late 1990s.[13] As a case of regional liberalization, consider the lucrative Singapore-Kuala Lumpur route. This route had for years been restricted by Malaysia to protect Malaysian Airlines, and was dominated by Malaysian Airlines and Singapore Airlines as a duopoly. In late 2007, the Malaysian government decided to allow AirAsia to operate on the route, paving the way for Tiger Airways (from the Singaporean side) to enter the route as well. The liberalization policy started with allowing two flights daily from each LCC, and then was extended to six daily flights in September 2008. As illustrated in Zhang, et al. (2009), the entry by AirAsia and Tiger Airways forced the two incumbent FSAs to significantly lower their fares, to the clear benefit of passengers.

Dobruszkes (2009) investigated airline competition in Europe following the liberalization in 1997. He found that traditional European airlines, especially the majors (Air France, British Airways, Lufthansa and KLM) have not benefited directly from the liberalization of European airspace in order to operate flights not centered on their country of origin. Their contribution to the usage of the 5^{th} -9^{th} air freedoms in Europe is less than 1% each. These carriers make greater use of the 5^{th} -9^{th} freedoms outside Europe, in particular on long –haul flights to the Far East that involve a stop over. In Europe, these carriers remain strongly rooted in their national centers. It is LCCs that have benefited most from the new air freedoms available as shown in table 4. Dobruszkes (2009) suggests that this may be due to the new mode of operation by LCCs, which facilitates the development of extra-national bases.

Table 4: Contribution of LCCs to the Use of 5th-9th Freedoms
(Europe 2005, excluding SAS from the total)

Air Freedoms	Flights	Seats
5th	0%	0%
6th	0%	0%
7th	63%	77%
8th	0%	0%
9th	24%	47%
Total	53%	71%

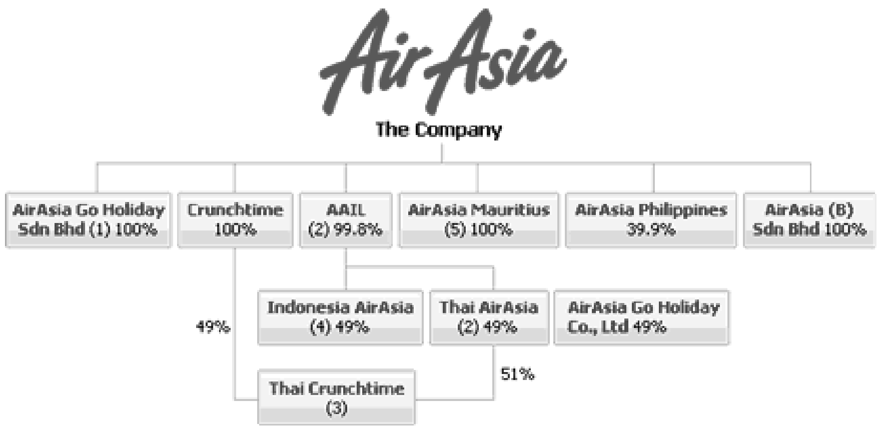
Source: Dobruszkes (2009)

[13] It is also interesting to note that statistics from the Tourism Office of Macau Government shows that after Viva Macau, a LCC, flied to Indonesia, Australia and Japan, visitor arrivals by air from these three countries have grown by 71%, 290%, and 300%, respectively.

Another important channel via which LCCs promote further policy liberalization is through the enhancement of the competitiveness of national carriers. Clougherty and Zhang (2008) identify three paths via which domestic rivalry (domestic competition) might influence international performance on the part of airlines. First, when there is an equivalence between the number of domestic and international competitors (that is, every domestic airline also serves international markets) then increasing the number of domestic competitors also increases the number of international competitors representing the nation. Accordingly, a strategic effect results as having multiple national competitors in world markets will enhance exports. Second, a "joint-economies of production" effect derives from the impact of domestic rivalry on the size of an incumbent firm's domestic operation, since size of domestic operation affects international performance in the airline industry (Clougherty, 2002, 2006). Third, domestic rivalry may also pressure firms to improve product quality and/or productivity, thus enhancing the competitiveness of home-nation airlines in international markets. In short, an additional rationale behind domestic deregulation and competition could well be the promotion of domestic carriers' competitiveness in international markets. Accordingly, the dramatic growth in domestic competition due to LCCs may significantly impact international competitive outcomes.

The large economic benefits of LCCs are so visible that their further developments tend to speed up the deregulation / liberalization process of domestic and international airline markets. On the other hand, as discussed in Zhang, et al. (2009), there are still a large number of visible and invisible barriers acting against growth of LCC activities in markets where LCCs are most needed. The organizational structure of AirAsia, arguably the most successful LCC in Asia, shown in figure 2, serves as a telling evidence of restrictions for an Asian LCC to grow its services cross national boundaries. In particular, given the restricted aviation regime in the region, AirAsia could extend its network and enter a new regional market only through joint venture (JV) arrangements or alliances: Thai AirAsia in Thailand and Indonesia AirAsia in Indonesia are two JV examples in which AirAsia holds a 49 percent share, so as to abide the national ownership restrictions of Thailand and Indonesia respectively.

Figure 2: Operating Companies for Air Asia Group



More recently, Tiger Airways (of Singapore) tried to establish JVs, namely, Tiger Airways Australia and Incheon Tiger Airways, in an attempt to expand its services to Australia and South Korea, respectively. While the Australian JV is in operation, the Korean project was called off in late December 2008 after more than one-year planning, citing by the "regulatory uncertainty" in Korea and a weak global economy (The Straits Times, 2008). The project would have been a tie-up with Incheon Metropolitan City, with the Singapore company taking a 49 percent stake. But from day one, the project faced local opposition. In August 2008, Korean LCCs (namely, Air Busan, Yeongnam Air, Jeju Air and Jin Air) jittery about the impending competition, filed a complaint with their country's

Ministry of Land, Transport and Maritime Affairs. They urged the government to put the brakes on the launch of the new carrier, claiming that it would in effect be controlled and run by Tiger, since the other shareholders had no airline experience. The airlines went so far as to say that the new airline would "attack Korea's aviation sovereignty" (The Straits Times, 2008).

The experiences from North America and Europe suggest that the benefits brought by LCCs are concrete, dramatic and lasting, and that they form a significant part of the gains from air transport liberalization. However, to fully gain such benefits, liberalization and deregulation need to be carried out.

5. Summary and Conclusion

As early as in the 1944 Chicago Convention, there had been proposals to liberalize the international aviation market. It took the industry half a century before the first Open Skies agreement got approved by the US-Netherlands governments in 1992. Although many Open Skies agreements have been reached in the following years, liberalization remains a formidable challenge. In addition, many of these liberalizations have been partial and incomplete, which needs further deregulations on ownership control and beyond rights etc. Many difficulties in liberalization can be ascribed to stakeholders' different expectations on the effects of alternative policy / agreement scenarios. The resulting uncertainty has prevented many governments from adopting substantial regulatory changes. This study examines the effects of past liberalization policies on economic growth, passenger traffic and low cost carriers. Our main conclusions are as follows:

- Liberalization has led to substantial economic and traffic growth. Such positive effects are mainly due to 1) increased competition in the aviation market, which reduces price and stimulates traffic growth; 2) productive efficiency gains as a result of carriers' optimization of their network operations and pricing strategy. In addition, the increased competitive pressure forces airlines to improve productivity, and eliminates inefficient carriers out of the market; 3) positive externalities to the overall economy including employment opportunities, trade promotion, better transport and logistics services etc. These impacts are not uniform across countries. However, there has been an increasing number of countries adopted (progressive) liberalizations. This suggests that countries involved have benefited from liberalization in general.

- Liberalization allowed carriers to optimize their networks to cover intra / inter continental markets. Hub-and-spoke networks have been extensively used by airlines to achieve cost advantages in production (economies of density) and / or revenue advantage. If ownership / citizenship restrictions are relaxed, market consolidation via merger and acquisition would allow airlines to strengthen their networks and market position. Strategic alliances allowed airlines to achieve "second best" network connection in markets where BSAs are still restrictive. Upon liberalization, the future growth of global airline alliance would be limited. Liberalization and network competition in international markets imply shift in traffic spatial pattern and market power. Therefore, it is important for countries to maintain their leadership in liberalization, thus that to keep their aviation sector competitive in the global markets.

- The prosperity of low cost carriers has brought significant impacts to the airline industry. There is a two-way relationship between LCC expansion vs. liberalization (and deregulation). The fast growth of LCCs leads to increased competition and reduced fare, which stimulate traffic substantially. These changes call for the removal of restrictions on capacity, frequency and pricing. In liberalized markets such as the EU single aviation market, LCCs have benefited most from the liberalization of beyond rights by establishing airport bases across borders. In addition, development of LCCs in domestic market can promote liberalization policy by increasing the competitiveness of a nation's aviation industry. On the other hand, existing regulations on route entry, ownership and effective citizen control have constrained the expansion of LCCs, and thereby, prevented the associated benefits to be fully realized.

The possibility of creating “destructive” or “excessive” competition had often been used as an excuse for regulation. Our investigation revealed that such negative effects were not material. Protection and regulation did not lead the airline industry to efficiency and profitability as hoped by policy makers. Instead, countries leading deregulation and liberalization scored various benefits for their aviation industry as well as the overall economy. Therefore, it is important for first-mover countries to maintain their leadership in liberalization, and it is urgent for countries still practicing tight regulation to catch up the wave of liberalization.

6. Acknowledgements

We would like to thank participants at the 2009 Sydney OECD- ITF (International Transport Forum) Conference and IFSPA 2009 Hong Kong Conference for helpful comments. Financial support from the OECD-ITF, Social Science and Humanities Research Council of Canada (SSHRC), and Hong Kong RGC Public Policy Research Grant (PolyU5002-PPR-5) are gratefully acknowledged. Special thanks to Chunyan Yu and Zhuo Lin for their comments and assistance to this study.

References

Airbus (2007), Global Market Forecast, 2007-2026.

AirNeth (2005), Strategies of multi-hub airlines and the implications for national aviation policies, AirNeth Workshop Report, 2005.

Aizenman, J. (2004), Endogneous pricing to market and financing cost, Journal of Monetary Economics 51(4), 691–712.

Asia Times, 2008. Prying open ASEAN’s skies. Asia Times, 6 July 2008.

ATRS (Air Transport Research Society), (2005): The ATRS Airport Performance Benchmarking Report: Global Standards for Airport Excellence, 3-volume report, published annually by the Centre for Transportation Studies, University of British Columbia, Vancouver, British Columbia. http://www.atrsworld.org.

Berechman, J. and Shy, O., (1998), The Structure of Airline Equilibrium Networks, in J.C.J.M. van der Bergh, P. Nijkamp, and P. Rietveld (eds.): Recent Advances in Spatial Equilibrium Modeling, Springer, Berlin.

Berry, S. (1992), Estimation of a Model of Entry in the Airline Industry, Econometrica, 60(4), pp.889-917

Bittlingmayer, G. (1990), Efficiency and Entry in a Simple Airline Network, International Journal of Industrial Organization 8, 245-57

Boeing (Boeing Commercial Aircrafts) (2008), Review of Boeing Commercial Airplanes Long-term Airplane Market Forecast Methodology and Airlines’s Underlying Requirement for Economic Profits, public distribution version.

Borenstein, S., (1989), Hubs and High Fares: Dominance and Market Power in the U.S. Airline Industry, RAND Journal of Economics 20, 344–365.

Borenstein, S. and Rose, N.L., (2007), How Airline Markets Work…or do they? Regulatory Reform in the Airline Industry, NBER Working Paper 13452

Brueckner, J.K. and Spiller, P.T. (1991), Competition and Mergers in Airline Networks, International Journal of Industrial Organization, 9, 323-342.

Brueckner, J.K. and Spiller, P.T., (1994), Economies of Traffic Density in the Deregulated Airline Industry, Journal of Law and Economics, 37, 379 - 415.

Brueckner, J.K., Dyer, N.J. and Spiller, P.T., (1992), Fare Determination in Airline Hub-and- Spoke Networks, Rand Journal of Economics, 23, 309 - 333.

Brueckner, J.K. and Whelen W.T. (2000), The price effect of international airline alliances, Journal of Law and Economics, 43, 503-545.

Brueckner, J.K. and Zhang, Y., (2001), A Model of Scheduling in Airline Networks: How a Hub-and-Spoke System Affects Flight Frequency, Fares and Welfare, Journal of Transport Economics and Policy, Volume 35, 195- 222.

Brueckner, J.K. (2004), Network Structure and Airline Scheduling, Journal of Industrial Economics, Volume LII, 291-312.

Button, K. (1998), Opening U.S. Skies to Global Airline Competition, Trade Policy Analysis, Trade Policy Analysis, Nov.24

Button, K., Lall, S., Stough, R. and Trice, M., (1999), High Technology Employment and Hub Airports, Journal of Air Transport Management, 5, pp. 53-59.

Button, K. (2006), Airports and International Economic Integration, ECMT-OECD working paper, presented in the 17th International Symposium on Theory and Practice in Transport Economics and Policy Berlin, 25-27 October 2006

Caves, R. (1962), Air Transport and Its Regulators: An Industry Study, Cambridge, MA: Harvard University Press.

Caves, D. W., Christensen, L.R., and Tretheway, M.W. (1984), Economies of Density versus Economies of Scale: Why Trunk and Local Service Costs Differ, Rand Journal of Economics, 15, 471 - 489.

Clougherty, J.A. (2002), US Domestic Airline Mergers: The Neglected International Determinants, International Journal of Industrial Organization, 20, 557-576.

Clougherty, J.A. (2006), The International Drivers of Domestic Airline Mergers in Twenty Nations: Integrating Industrial Organization and International Business, Managerial and Decision Economics, 27, 75-93.

Clougherty, J.A. and Zhang, A. (2009), Domestic Rivalry and Export Performance: Theory and Evidence from International Airline Markets, Canadian Journal of Economics, 42(2),pp 440-468.

Datamonitor (2005), "Express Logistics in the Netherlands: Industry Profile", Reference Code: 0176-2331, www.datamonitor.com

Djankov, S., Miguel, E., Qian, Y., Roland, G. and Zhuravskaya E., (2005), Who are Russia's Entrepreneurs?, Journal of European Economic Association, Vol. 3, No. 2-3, Pages 587-597.

Dobruszkes, F., (2009), Does liberalization of air transport imply increasing competition? Lessons from the European case, Transport Policy,16,pp.29–39.

DOT (U.S. Department of Transportation) (2001), Dominated hub fares, Domestic aviation competition series.

Douglas, G.W. and Miller, J.C. (1974), Quality Competition, Industry Equilibrium and Efficiency in the Price Constrained Airline Market, American Economic Review 64, 657-669.

Dresner, M., Lin, J. and Windle, R., (1996), The Impact of Low-cost Carriers on Airport and Route Competition, Journal of Transport Economics and Policy, 30, 309-328.

Fethi, M.D., Jackson, P.M., and Wayman-Jones, T.G. (2000), Measuring Efficiency of European Airlines: an Application of DEA and Tobit analysis, Discussion Paper, Efficiency and Productivity Research Unit, University of Leicester, UK.

Franke, M., (2004), Competition Between Network Carriers – Retreat Battle or Breakthrough to a New Level of Efficiency?, Journal of Air Transport Management 10, 15-21.

GAO (U.S. General Accounting Office) (1989), Barriers to Competition in the Airline Industry.

GAO (U.S. General Accounting Office) (1990), Airline competition: industry operating and marketing practices limit market entry.

General Accountability Office (GAO, 2008), Airline Industry: Potential Mergers and Acquisitions Driven by Financial and Competitive Pressures, GAO-08-845 July 31, 2008.

Harrigan, J. (2005), Airplanes and Comparative Advantage, NBER Working Paper No. 11688

Hausman, W.H., Lee, L.L. and Subramanian, U. (2005), Global Logistics Services, Supply Chain Metrics and Bilateral Trade Patterns, mimeo, World Bank, October.

Hendricks, K., Piccione, M. and Tan, G. (1995), The Economics of Hubs: The Case of Monopoly, Review of Economic Studies, 62, 83-99.

Hendricks, K., Piccione, M. and Tan G. (1997), Entry and Exit in Hub-spoke Networks", RAND Journal of Economics 28, 291–303.

Hendricks, K., Piccione, M. and Tan G. (1999), Equilibria in Networks, Econometrica, 67, 1407-1434.

Hummels, D. (2006), Global Trends in Trade and Transportation, ECMT-OECD Paper presented in the 17th International Symposium on Theory and Practice in Transport Economics and Policy, Berlin, 25-27 October 2006.

ICAO (2007), Outlook for Air Transport to the Year 2025, International Civil Aviation Organization, Montreal.

ICAO Secretariat (2007), Overview of Trends and Developments in International Air Transport, http://www.icao.int/icao/en/atb/epm/Ecp/OverviewTrends.pdf

InterVISTAS (2006), The Economic Impact of Air Service Liberalisation, report available at: http://www.intervistas.com/4/reports/2006-06-07_EconomicImpactOfAirServiceLiberalization_FinalReport.pdf

Irwin, M. D. and Kasarda, J.D. (1991), Air Passenger Linkages and Employment Growth among U.S. Metropolitan Areas, American Sociological Review 56(4), 524-537.

Jordan, W.A. (1970), Airline Regulation in America: Effects and Imperfections, Baltimore: Johns Hopkins Press.

Keeler, T.E. (1972), Airline Regulation and Market Performance, Bell Journal of Economics and Management Science. 3, pp.399-424.

Lee, D., and Prado, M. J. (2005), The Impact of Passenger Mix on Reported "Hub Premiums" in the U.S. Airline Industry, Southern Economic Journal, 72, pp.372-396.

Levine, M. E. (1965), Is Regulation Necessary? California Air Transportation and National Regulatory Policy, The Yale Law Journal. 74, pp. 1416-1447.

Levine, M.E., (1987), Airline Competition in Deregulated Markets: Theory, Firm Strategies and Public Policy, Yale Journal on Regulation 4, pp.283-344.

Lijesen, M., Rietveld, P. and Nijkamp, P. (2001), Hub premiums in European civil aviation, Transport Policy, Vol. 8, No. 3, pp.193-199.

Limao, N. and Venables, A.J. (2001), Infrastructure, Geographical Disadvantage, Transport Costs, and Trade, The World Bank Economic Review, Vol. 15, No. 3, pp: 451-479

Maillebiau, E. and Hansen, M. (1995), Demand and Consumer Welfare Impacts of International Airline Liberalization: the Case of the North Atlantic, Journal of Transport Economics and Policy, 29 (2), pp.115–136.

Morrison, S.A. and Winston, C. (1986), Empirical Implications and Tests of the Contestability Hypotheses, Journal of Law and Economics 30, pp. 53-66.

Morrison, S.A. and Winston, C (1995), The Evolution of the Airline Industry. Journal of Transport Economics and Policy Volume 35, Part 2, 254 Washington, DC: The Brookings Institution.
Morrison, S.A., (2001). Actual, Adjacent, and Potential Competition: Estimating the Full Effect of Southwest Airlines, Journal of Transport Economics and Policy 35, 239-256.

Oum, T.H. and Tretheway, M.W. (1990). Airline Hub and Spoke System, Journal of Transportation Research Forum, vol.30 (1990), pp. 380-393.

Oum, T.H., Taylor, A.J. and Zhang, A. (1993). Strategic Airline Policy in the Globalizing Airline Networks, Transportation Journal, vol.32, no.3, pp.14-30.

Oum, T.H., Zhang, A. and Zhang, Y. (1995), Airline Network Rivalry, Canadian Journal of Economics, 28, 836-857.

Oum, T.H. and Yu, C. (1998). Winning Airlines: Productivity and Cost Competitiveness of the World's Major Airlines, Kluwer Academic Press, New York, London; 1998), 240 pages. http://www.wkap.nl/prod/b/0-7923-8010-X

Oum, T.H., Park, J.H., and Zhang, A. (2000). Globalization and Strategic Alliances: The Case of the Airline Industry (Elsevier Science, 2000), 229 pages. ISBN 0 08 o43596 3. Accessible from http://www.elsevier.nl

Oum, T.H, Fu, X., and Yu, C, (2005): New Evidences on Airline Efficiency and Yields: A Comparative Analysis of Major North American Air Carriers and Its Implications, Transport Policy, 12, pp.153-164.

Oum, T.H. and Park, J.H. (2004), Multinational Firms' Location Preference for Regional Distribution Centers: Focus on the Northeast Asian Region, Transportation Research E :Logistics and Transportation Review , vol.40 (2004). 101-121.

Piermartini, R. and Rousova, L. (2008), Liberalization of Air Transport Services and Passenger Traffic, World Trade Organization - Economic Research and Statistics Division, Staff Working Paper ERSD-2008-06

Richards, K., (1996), The Effect of Southwest Airlines on US Airline Markets. Research in Transportation Economics, 4, pp.33-47.

Schaur, G. (2006), Airplanes and Price Volatility, mimeo, Purdue University.

Spiller, P.T. (1989). Pricing of Hub-and-Spoke Networks, Economics Letters, 30, pp.165-169.

Swan, W. (2007), Misunderstandings about Airline Growth, Journal of Air Transport Management 13, pp.3–8.

The Straits Times (2008), Tiger Airways Korea venture off, 31 December 2008, Singapore.

Transportation Research Board, National Research Council (1999), Entry and Competition in the US Airline Industry: Issues and Opportunities, Washington, DC: National Academy Press.

Tretheway, M.W. and Kincaid, I.S., (2005), The Effect of Market Structure on Airline Prices: A Review of Empirical Results, Journal of Air Law and Commerce 70, pp. 467-498.

US DOT, (1993), The Airline Deregulation Evolution Continues: The Southwest Effect, Office of Aviation Analysis, US Department of Transportation.

Warren, T. and Findlay, C. (1998), Competition Policy and International Trade in Air Transport and Telecommunications Services, World Economy, Vol. 21, pp. 445-456.

Windle, R. and Dresner, M., (1995). The Short and Long Run Effects of Entry on US Domestic Air routes, Transportation Journal 35, pp.14-25.

World Trade Organisation (WTO) (2006), Second Review of the Air Transport Annex: Developments in the Air Transport Sector (Part Two) Quantitative air services agreements review (QUASAR), Volumes I and II Note by the Secretariat, document S/C/W/270/Add.1".

Zhang, A. and Wei, X. (1993), Competition in Airline Networks: The Case of Constant Elasticity Demands. Economics Letters, 42, 253-259.

Zhang, A. (1996), An Analysis of Fortress Hubs in Airline Networks, Journal of Transport Economics and Policy, 30, pp. 293-308.

Zhang, A., Hanaoka, S., Inamura, H. and Ishikura, T. (2009), Low Cost Carriers in Asia: Deregulation, Regional Liberalization and Secondary Airports, Research in Transport Economics, Vol. 17, forthcoming.

Inventory policy comparison on supply chain network by simulation technique

Nam Kyu Park [1], B Dragovic[2] and Woo Young Choi [3]

[1] Department of Distribution Management, The Tong myung University, S. Korea.
[2] Maritime Transport and Traffic Division, Montenegro University, Montenegro.
[3] Department of Distribution Management, The Tong myung University, S. Korea.

Abstract

The aim of the paper is to solve the problem of customer reduction due to the difficulty of parts sourcing which impacts production delay and delivery delay in SC networks. Furthermore, this paper is to suggest the new inventory policy of MTS in order to solve the problem of current inventory policy. In order to compare two policies, a LCD maker is selected as a case study and the real data for 2007 years is used for simulation input. The maker uses MTO policy for parts sourcing which has the problem of lead time even if it has some advantage of inventory cost. Based on current process, the simulation program of AS-IS model and TO-BE model using ARENA 10 version is developed for evaluation. In a result, the order number of two policies shows that MTO is 52 and MTS is 53. However the quantity of order shows big difference such that MTO is 168,000 and MTS is 225,106. Particularly, the lead time of new inventory policy shows much shorter that that of MTO such that MTO 100 is days and MTS is 16 days. In spite of short lead time by MTS policy, new policy has to take burden of inventory cost per year. Total inventory cost per year by MTS policy is US $ 11,254 and each part inventory cost is that POL is US$ 1807, LDI is US$ 2,166 and Panel is US$ 7281. The implication of the research is that the company has to consider the cost and the service simultaneously in deciding the inventory policy. In the paper, even if the optimal point of deciding is put into tactical area, the ground of decision is suggested in order to improve the problem in SC networks.

Keywords: SCM, Simulation, Arena, MTO, MTS

1. Introduction

Supply chain management analyzes organically the whole process ranging from material supply to product delivery to the customers, checking and solving problems, and consequently trying to maximizing the profits of all the entities in a supply chain. To achieve effective and efficient supply chain management, the optimization of both production and inventory management of the organizations in the supply chain is greatly needed. Optimal production management and proper level of inventory lead not only to the cost reduction of the whole supply chain, but also to the profit increase of each individual company. Like this, an optimal production and inventory management policy is a critical factor to the success of supply chain management, but it is not an easy task. In order to achieve effective and efficient supply chain management, the level of inventory is required to remain to a minimum, reducing backorders, and simultaneously maintaining a high service level. But inventory reduction and a high service level are contradictory to each other, and so, finding the best trade-off between these two goals is not an easy task. If exact forecasting is possible, we will be able to reduce inventory to a minimum, while fully satisfying customer's demand, but it is practically impossible. This study has used two research methods: theoretical research based on previous studies and simulation method. The previous studies are including literature review and theoretical research through the analysis of related data, while conducting expert interviews for related data collection. Based on previous studies and related data, this study has designed both a production management model and an inventory management model, and then developed a simulation model. ARENA (Version 10) has been used as a simulation language. In order to observe inventory changes thoroughly, various order quantities and lead times have been used, and then we have analyzed their results.

2. Literature Review

2.1. Previous Studies on MTO Production

The research on the production planning under the MTO (Make-To-Order) production environment can be divided into two: research at the strategic level and research at the operational level. The researches at the strategic level are mainly focusing on the decision support system of corporate organizations in terms of a market environment. Carravilla & Sousaa (1995) have divided the production planning under the MTO environment of shoes industry into three stages and have suggested a decision support system suitable to the purpose of each stage. Corti et al. (2006) have checked the possibility of production according to the orders and delivery dates, and have suggested a model which enables the coordination of both production capacity and production range. Haskose et al. (2004) have made researches on how much work volume affects the capacity of production system, mainly focusing on analyzing its effect in terms of material use and material procurement period.

2.2. Previous Studies on MTS

B.C. Giri and W.Y. Yun (2005) have presented a model for determining optimal production volume including the studies on production system failure and repairs which come from the uncertainty of production system. Hanen Bouchriha et al. (2005) have made researches on lot sizing in order to minimize the costs for the paper production process. Georghios P. Sphicas (2006) has intensified his studies on the existing EOQ (economic order quantity) model and EPQ (economic production quantity) model by adding the assumption on the permission range of backorder.

3. Case Study of the Selected Company

The "A" company selected for this case study is a solid company located at Icheon of Gyeonggi Province. The company is producing about 60 items including the main products such as TFT-LCD (Thin Film Transistor – Liquid Crystal Display), LCD TV, Navigation and MP3 and 4. Its annual sales amount to about 100 billion Korean won. The company has another production base and logistic center in China, and 70% of the company's total production comes from the first factory and the rest 30% from the second. The company is outsourcing its raw materials and parts to the outside companies. Therefore, it procures materials and parts from other partner companies, and manufactures its products through assembly lines, and produces its final products. In this study we have selected one of its main products - a navigation module - for our case study, as it is practically impossible to use all the products for simulation test.

3.1. Production and Logistics Process

Let's take a look at the production and logistics processes of A company. As shown in the <Figure 1>, it purchases materials such as a Glass, Pol, and LSI, and those items are to undergo the IQC (Incoming Quality Control) inspection and then move on to the panel production lines. After panel production, module bonding, COG (Chip ON Glass), TCP (Tape Carrier Package), FPC (Flexible Printed Circuit) processes are to be followed. Again in-house inspections such as UV coating, air cleaning, and foil fixing are to be conducted and then go to the warehouse.

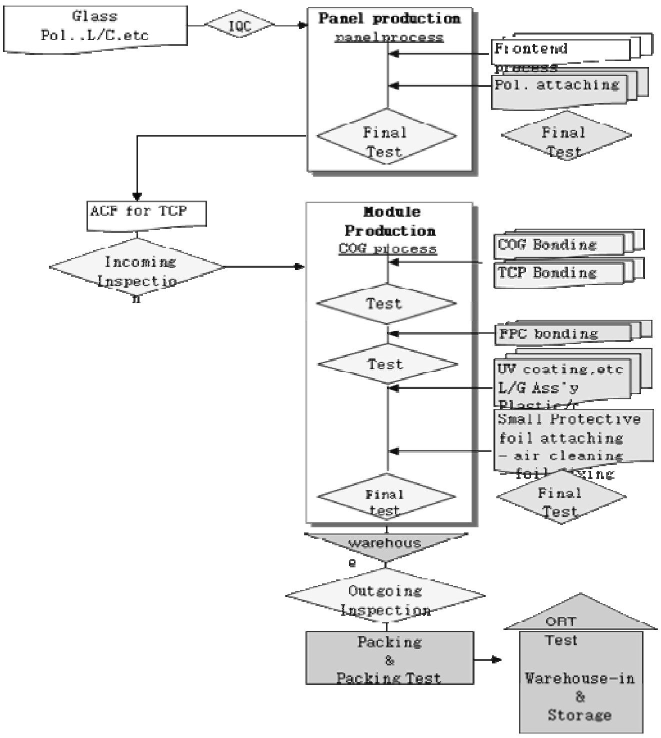


Figure 1 : "A" company's production and logistics processes

3.2. Problems of Logistics Process

Through intensive and wide range of interviews with persons in charge of A company, this study has pointed out 9 kinds of problems such as "limited effectiveness of production planning," "long lead time of main parts," and "loss of customers due to delivery delay." These problems have been summarized in the below <Table 1>.

Table 1 : "A" Company Problems on business area

Section	Problems
Production Planning	. Limited effectiveness of production planning (production planning: once a month)
	. No simulation test because of production planning
	. Low yield of DSTN(Double Super Twisted Nematic) product
	. Production troubles (such as defective products) give rise to a material supply problem.
Materials Purchase	. Long lead time of major parts (such as POL and IC)
	. Occurrence of a material supply problem gives rise to long waiting time, thus causing loss in job handling.
	. Lack of the proper linkage between parts order and final products delivery schedule
Product Delivery	. Loss of customers due to delivery delay
	. Occurrence of a production problem causes additional delivery costs due to emergency delivery.

4. Input Data Analysis of Simulation Modeling

In order to generate simulation input data, the following data for the selected product of A company have been collected for a full year from January 2007 to December 2007: the interval of order arrival time, order quantity, order quantity of each part, and handling hours of production and each process. The collected data has been analyzed by Arena Input Analyzer.

Table 2 : Manufacturing plant & Delivery Prcess Simulation Data

Section	Expression
Order Recive	-0.5 + LOGN(6.42, 6.01)'
Customer Order Amount	NORM(3.74e+003, 1.66e+003)
warehousing of Parts & inspection	Triangular(-0.5, 3, +0.5)
Produce	Constant(5)
Finished product	Triangular(±1, 2)
Delivery to Customer	Normal(±0.5, 2)

Table 3 : Parts Simulation Data

Section	Expression
Part Order	Panel: Order for the same amount of finished products
	LDI: Order for the same amount of finished products
	POL: Double order of finished products
Panel Lead-Time	Triangular(±1,2, 15)
LDI Lead-Time	Triangular(±1,2, 29)
POL Lead-Time	Triangular(±1,2, 31)

5. MTO Simulation Modeling

In order to define the MTO logistics process, this study has analyzed the production process of the A company as shown in the following <Figure2 >. First of all, if the company receives an order, it will make production planning, while checking whether it has enough finished products or materials inventory for the order. If there are enough finished products, it will immediately deliver them to the customers, but if not, it has to place an order with outside companies for parts. In the case of parts, likewise, if it has enough parts inventory, it will soon start to manufacture, but if not, it has to place an order for parts. However, basically the MTO production method has no inventory of finished products or parts. Therefore, the above-mentioned process is not performed, but only for an exceptional case has its process model been made. In the field workplace, 10% of additional parts order is being placed in consideration of the losses that can occur in the production or on the move. But this study is based on the assumption of having no loss. The parts such as panel, LDI, and POL will be supplied to the production line of the company, and according to its production planning, final products will be produced and inspected, and then delivered to the customers.

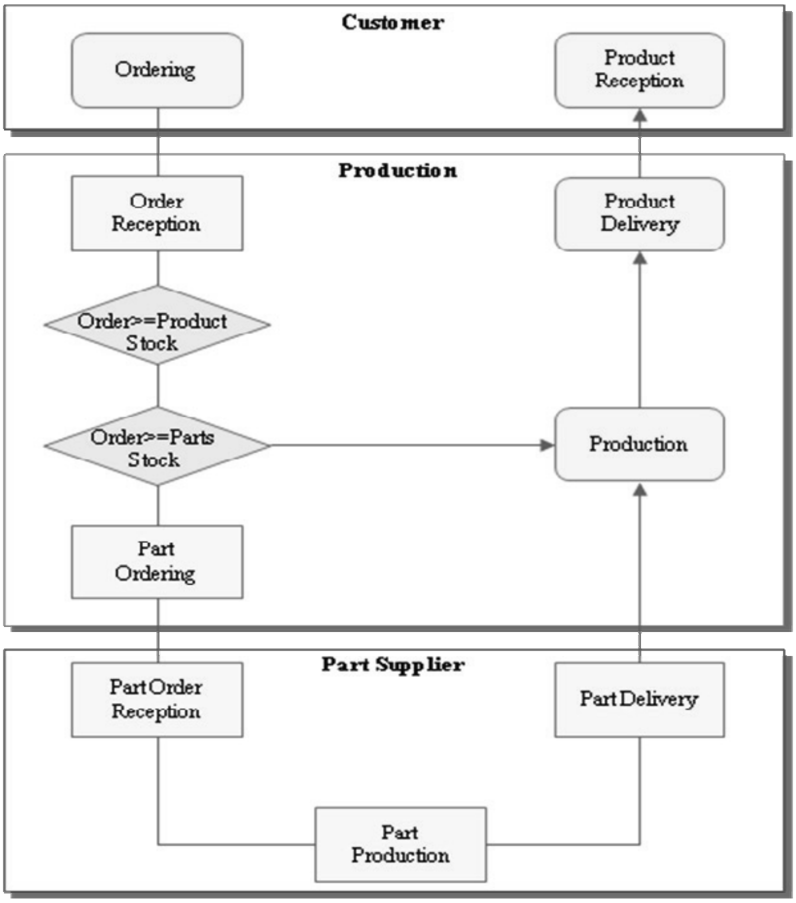


Figure 2 : MTO Logistic Process Modeling

5.1. Validity Test of MTO Simulation Model

The performance of simulation has been conducted on a daily basis, and its results have been summarized on an annual basis. The simulation results of the MTO model are illustrated in the following <Table 4 >. The differences between the current method and the MTO simulation model are as follows: the number of orders per year is 52, both are the same. In terms of annual order quantity, the current method amounts to about 194,000, and the MTO simulation model 211,000, showing a difference of 17,000. In terms of annual production volume, the current method amounts to 165,000, and the MTO simulation model 168,000, showing a difference of 2,700. In terms of the procurement period of purchase orders, the current method requires 96 days, and the MTO simulation model 100 day, showing the difference of 4 days. Finally, in terms of job performance rating within the delivery period, the former shows 88%, and the latter 89%. Based on these results, we can find out that the MTO simulation model is well reflecting reality.

Table 4 : Present condition of Production and MTO Simulation Result Compare

Section	Annual Orders(Number)	Amount of Orders received(EA)	Annual Production(EA)	Lead-time(Day)	Deadline of Delivery(%)
Current approach	52	194,464	165,680	96	88
MTO Simulation	52	211,582	168,460	100	89

6. MTS Simulation Modeling

Ordering methods available in the MTS include quantity ordering, regular ordering, and Min-Max. In order to simplify the problem, this study has used quantity ordering, that is, EOQ and ROP (Re-Order Point).

6.1. MTS Simulation Input Data

The collected data on the company's main parts such as panel, LDI, and POL have been used as MTS simulation input data, which is shown in the following <Table 5>.

Table 5 : MTS Model Basic Input Data

Section	POL	LDI	Panel
Product price	0.5$	2$	8$
Stock maintenance	20%	20%	20%
AVG Order(1 times)	7,853(EA)	3,777(EA)	3,777(EA)
Daily Demand	1,1190(EA)	538(EA)	538(EA)
Deviation of demand(Year)	503(EA)	242(EA)	242(EA)
Lead-Time(AVG)	31(Day)	29(Day)	15(Day)
Ordering Cost	20$	74$	296$

By using the data on taking and placing orders during the past one year, this study has calculated EOQ and ROP, and its results are shown in the following <Table 6>.

Table 6 : EOQ and ROP(Each Parts)

Section	POL	LDI	Panel
Economic Order Quantity(EOQ)	9,037 (EA)	8,525 (EA)	8,525 (EA)
Reoder Point(ROP)	41,209 (EA)	18,642 (EA)	10,255 (EA)

6.2. MTS Inventory Management Model

The MTS simulation model is much different than the MTO simulation model. As shown in the <Figure3>, the inventory management model has been made independently. This inventory system is based on the quantity ordering, always checking ROP, and placing an order by way of EOQ, if it goes below ROP.

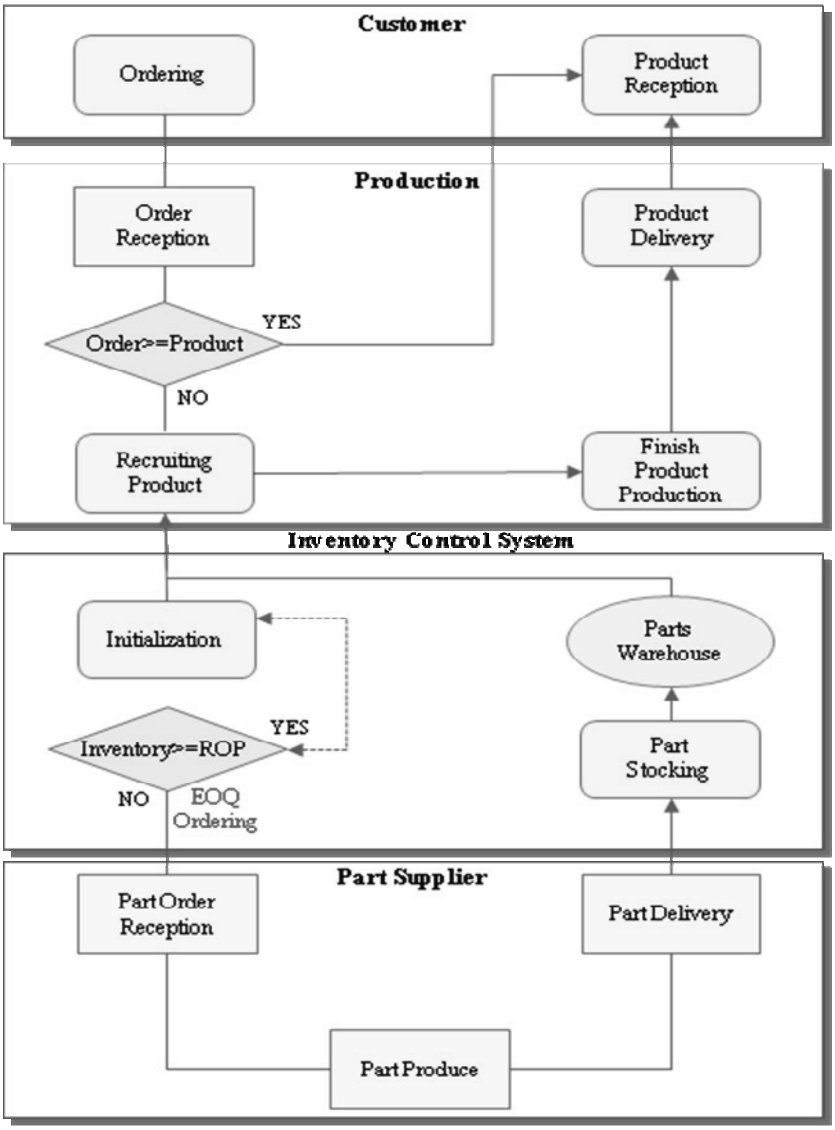


Figure 3 : MTS Logistic Process Modeling

6.3. MTS Simulation Results

The MTS simulation has been conducted on a daily basis. In order to test the results and also to reduce the range of error, the simulations have been conducted for ten years. The results of simulations are shown in the following <Table 7>, which includes the number of orders per each time, quantity of orders fulfilled, production volume, average procurement period.

Table 7 : MTS Simulation Result

Section	The number of orders	Order throughput	Annual production(EA)	Lead-Time(Day)
AVG	53	53	225,106	16

7. Inventory Policy Comparison between MTO and MTS

This study has made a comparison of the results of both inventory policies: MTO inventory policy and MTS inventory policy. The results are as follows. In terms of the number of orders per year, MTO is 52, and MTS is 53. In terms of order quantity, both MTO and MTS show a similar result. But in terms of annual production quantity, as shown in the <Table 8>, MTO is 168,000 and MTS 225,106, showing a difference of about 57,000. Meanwhile, in terms of procurement period, MTO is 100 days, and MTS is 16 days, making a great difference of 84 days. The reason is that, in the case of MTS inventory management model, it always has proper inventory available for every order, thus bringing productivity enhancement and reduction of procurement period.

Table 8 : MTS Simulation Result

Deadline of Delivery(%)	Section	Annual Orders(Number)	Amount of Order Received(EA)	Annual Production(EA)	Lead-time(Day)
89	MTO	52	211,582	168,460	100
100	MTS	53	225,106	225,106	16

However, MTS increases inventory volume, which causes an inventory cost to increase, as illustrated in the <Table 9>.

Table 9 : Annual Inventory costs of Each Parts

Section	POL	LDI	Panel
Annual Demand	408,374	196,409	196,409
Product price ($)	0.5	2	8
Stock maintenance (%)	20	20	20
Ordering Cost ($)	20	74	296
Fill up Order(EA)	9,037	8,525	8525

8. Conclusions

This study has dealt with the SCM (supply chain management) of domestic LCD module manufacturer, investigating how much the difficulty of parts procurement can affect production and delivery period, so that it may cause the loss of customers. So, this study has focused on solving this problem and simultaneously has tried to suggest the merits and demerits of a new alterative method. This study has developed the MTO simulation system and MTS simulation system of the case company. By using both simulation models, we have compared and evaluated the results of both systems, presenting the problems to be solved and to be considered. According to the results of MTO inventory policy and MTS inventory policy, in terms of annual production quantity, MTO method is 168,000, and MTS method is 225,106, showing a difference of about 57,000 but in terms of procurement period, MTO is 100 days, and MTS is 16, making a great difference of 84 days. This great difference comes from the fact that MTS has proper inventory available for any order, consequently bringing productivity improvement and procurement period curtailment. In spite of this great achievement in terms of lead time, however, MTS has to bear considerable inventory costs in terms of inventory level. According to the total annual inventory costs of each part in the MTS method, POL is US$ 1,807, LDI is US$ 2,166, and Panel US$ 7,281, and so the total inventory costs run to US$ 11,254. Through this study we have found out the following suggestions. Business management is always facing continuous decision making process to choose a better alternative. In the case of an inventory problem, management has to make a decision on whether it will curtail inventory costs at the expense of customer service, or whether it will improve customer service at its expense.

9. Acknowledgements

This work was supported by the grant No. B0009720 from the Regional Technology Innovation Program of the Ministry of Knowledge Economy(MKE)

References

Carravilla & Sousaa (1995), Hierarchical production planning in a Make-To-Order company : a case study, European Journal of Operational Research, 86: 43-56.

Corti et al. (2006), A capacity-driven approach to establish reliable due dates in a MTO environment, International Journal of Production Economics, 104: 536-554.

Haskose et al. (2004), Performance analysis of make-to-order manufacturing systems under different workload control regimes, International Journal of Production Economics, 90: 169-186.

Giri, B.C. and Yun, W.Y. (2005), Optimal design of unreliable production–inventory systems with variable production rate, European Journal of Operational Research, 162: 372-386.

Hanen Bouchriha et al. (2007), Lot sizing problem on a paper machine under a cyclic production approach, International Journal of Production Economics, 105: 318-328.

Sphicas, G. P. (2006), EOQ and EPQ with linear and fixed backorder costs: Two cases identified and models analyzed without calculus, International Journal of Production Economics, 100: 59-64.

Foreland-based regionalization: integrating intermediate hubs with port hinterlands

Jean-Paul Rodrigue [1] *and Theo Notteboom* [2]

[1] Department of Global Studies & Geography, Hofstra University, Hempstead, New York 11549, USA.
E-mail: Jean-paul.Rodrigue@Hofstra.edu
[2] Institute of Transport & Maritime Management, University of Antwerp, Keizerstraat 64, B-2000 Antwerp, Belgium.
E-mail: theo.notteboom@.ua.ac.be

Abstract

The development and changes in port hinterlands have received a lot of attention since they represent substantial opportunities to improve the efficiency of global freight distribution. Port regionalization was a concept brought forward by Notteboom and Rodrigue (2005) to articulate the emerging port hinterland dynamics in light of containerization, supply chain management and the setting of inland terminals. This paper expands the concept by looking at a particular dimension of the regionalization paradigm concerning the evolving role of intermediate hubs.

It is argued that in addition to hinterland-based regionalization there is also a foreland-based regionalization where intermediate hubs capture a maritime hinterland. This intensity and viability of processes of foreland-based regionalization are depending on multiple geographical, technical and market-related factors. The paper will identify and analyze these underlying parameters. By doing so, the paper assesses whether foreland-based regionalization is just a transitional phase in port development or, alternatively, represents an emerging functional characteristic of contemporary freight distribution.

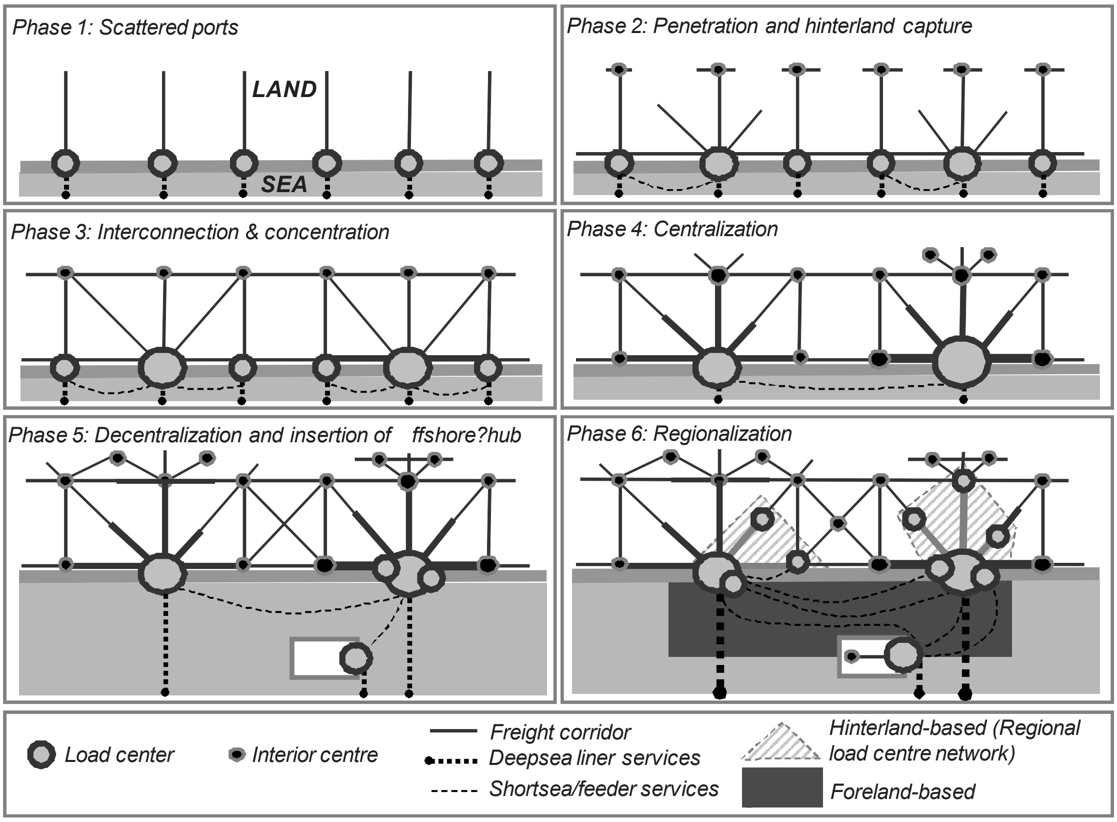
Keywords: Port Regionalization, Hinterland, Intermediate hubs (Offshore).

1. Introduction: the Port Regionalization Thesis Revisited

The development and changes in port hinterlands have received a lot of attention since they represent substantial opportunities to improve the efficiency of global freight distribution by improving its most costly segment. Port regionalization was a concept brought forward by Notteboom and Rodrigue (2005) to articulate the emerging port hinterland dynamics in light of containerization, supply chain management and the setting of inland terminals. The phase of regionalization brings the perspective of port development to a higher geographical scale, i.e. beyond the port perimeter.

Two main forces have triggered regionalization, one global and the other local. The first force concerns globalization where regionalization enables the development of a distribution network that corresponds more closely to fragmented production and consumption systems. Supply chain management can be accommodated on the maritime side with economies of scale and frequency of service along major pendulum routes linking gateways and hubs. On the inland side freight has to find its way to (and from) a variety of locations, which requires spatial deconsolidation (or consolidation). The second force concerns local constraints such as congestion and limited amount of land that impact port growth and expansion and which can partially been circumscribed by regionalization. Many freight activities that used to take place in proximity of port terminal facilities can take place further inland with the setting of a network of inland terminals. Jointly, these forces are part of two phases that have transformed port systems in recent decades.

Figure 1: A model on port system development, including a revised sixth phase



The port system development model as presented by Notteboom and Rodrigue (2005) is depicted in figure 1. The first phases (1 to 4) are well evidenced by the traditional port growth theories (Taaffe et al., 1963; Barke, 1986; Hayuth, 1981). The subsequent two phases have seen noticeable transformations in the port – hinterland relationships.

Phase 5 depicts a phase of decentralization and the insertion of intermediate hubs. The growth in the volume of containerized traffic in the 1980s permitted a greater number of ports to participate as load centers by competing over extended hinterlands since a critical mass was reached. The outcome was a relative level of decentralization along several maritime ranges. Additionally intermediate hubs were built to accommodate modern containership drafts on sites having land for future expansion, with lower labor costs and with terminals owned, in whole or in part, by carriers or port operators. In some cases intermediate hubs were developed in offshore location, on small islands with a very small local cargo base. Initially, these terminals solely focused on accommodating transhipment flows, but in time other freight distribution activities started to locate in vicinity.

Phase 6 consists of a regionalization phase, i.e. the integration of inland freight distribution centres and terminals with gateway ports. The outcome is the formation of a regional load center network with an improved efficiency of inland freight distribution. Gateways achieve a higher level of synchronization with their hinterlands through specialized high capacity corridors of circulation serviced by rail or barges.

In this development phase the port system consequently adapts to the imperatives of distribution systems and global production networks while mitigating local constraints. The port regionalization phase is characterized by a strong functional interdependency and even joint development of a specific load centre and (selected) multimodal logistics platforms in its hinterland, ultimately leading to the formation of a regional load centre network that is aimed at meeting the requirements of global logistics and production networks

In the meantime the port regionalization model has been applied to concrete cases, e.g. see Notteboom (2006) for an application to the port of Antwerp. The phasing and particularly the order of the phases have received some further elaboration in more recent academic work. The models on port development portray a high degree of path dependency in the development of ports at a regional scale. Path dependence implies that port systems evolve by building on previous phases and 'memory effects'. Port systems would follow a similar evolutionary development path. Notteboom (2007) argued, however, that port development processes also show a certain degree of contingency. Strategies and actions of market players and other stakeholders may deviate from existing development paths. Both path dependency and contingency explain why port systems around the world do not develop along the same lines or follow the same sequence of stages as suggested in the models on port system development. The result is some level of disparity among port system developments around the world.

More recently, Rimmer and Comtois (2009) stated there is no need for an additional sixth phase because they see regionalization as nothing else than decentralization. We argue that the regionalization phase is more than a simple decentralization phase as it involves the expansion of the hinterland reach of the port through a number of strategies linking the port more closely to inland freight distribution centres in a functional way.

We do however agree with the view of Rimmer and Comtois (2009) that there is a danger of becoming too preoccupied with the land-based network, without incorporating the realities in the maritime space. The port regionalization phase focuses strongly on the hinterland side of the spectrum (i.e. the changing dynamics between ports and inland centres towards the formation of regional load centre networks), while the issue of maritime networks and, in particular, the role of intermediate hubs in the port regionalization phase have not been thoroughly developed.

Concomitantly with the phase of regionalization, transport terminals have seen a higher level of integration within freight distribution systems, either as a buffer where they can be used for temporary storage or as a constraint inciting various forms of satellite/inland terminal use and inventory in transit practices (Rodrigue and Notteboom 2009). This emerging practice of 'terminalisation' (higher integration of intermodal terminals in supply chain management) reinforces the regionalization thesis but also underlines that the regionalization thesis looks at only one side of the maritime / land interface, with the actual role of intermediate hubs being unclear.

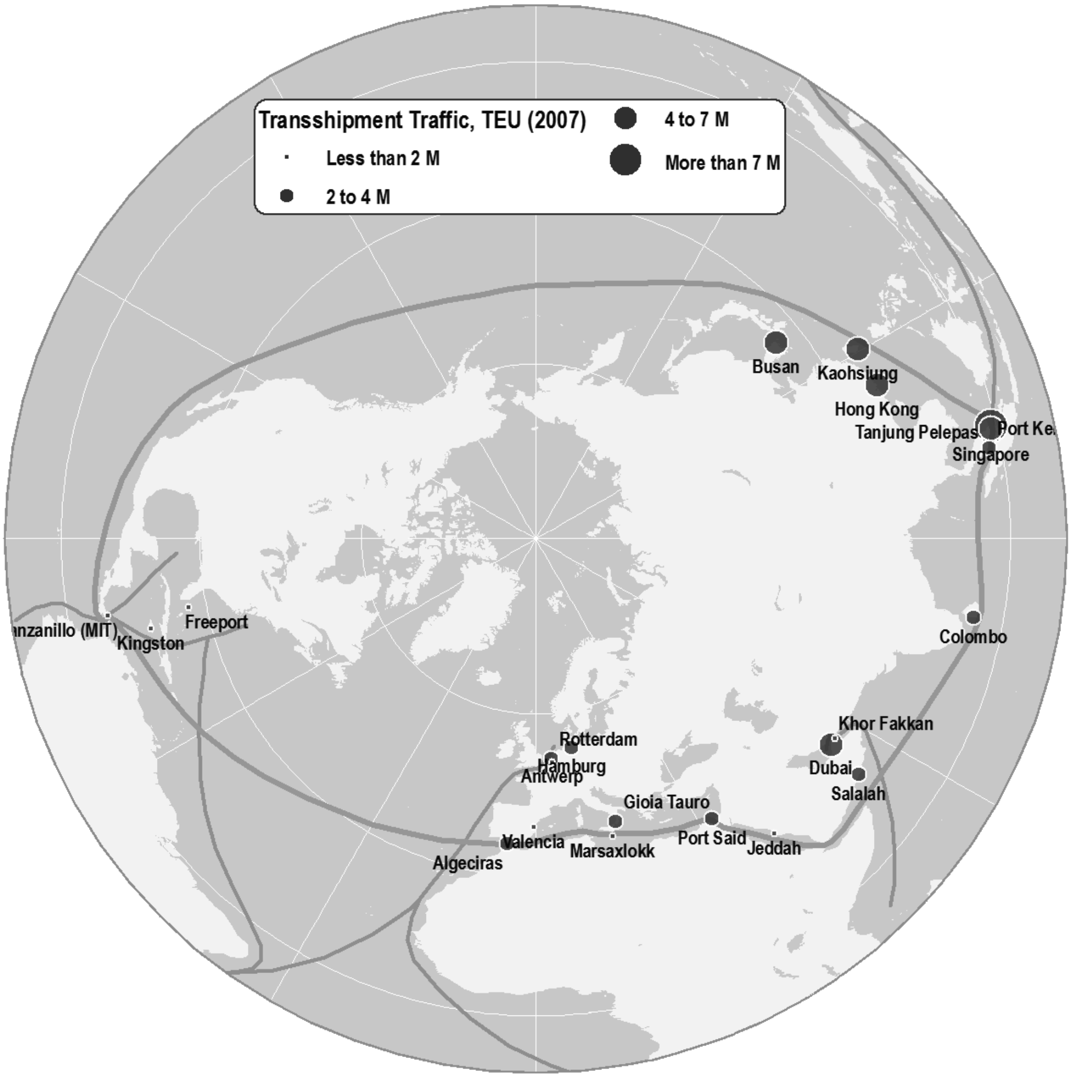
This paper extends the concept of regionalization by looking at a particular dimension of the paradigm concerning the evolving role of intermediate hubs that are capturing maritime forelands to create added value. A distinction is made between hinterland-based regionalization and foreland-based regionalization as suggested in phase 6 of figure 1. The next section discusses the role and vulnerability of intermediate hubs as part of global shipping networks. Then, the foreland-based regionalization concept is introduced as a way for intermediate hubs to acquire a more sustainable position in supply chains and vis-à-vis partner ports in the networks they serve.

2. The Role and Function of Intermediate Hubs

2.1. General Discussion

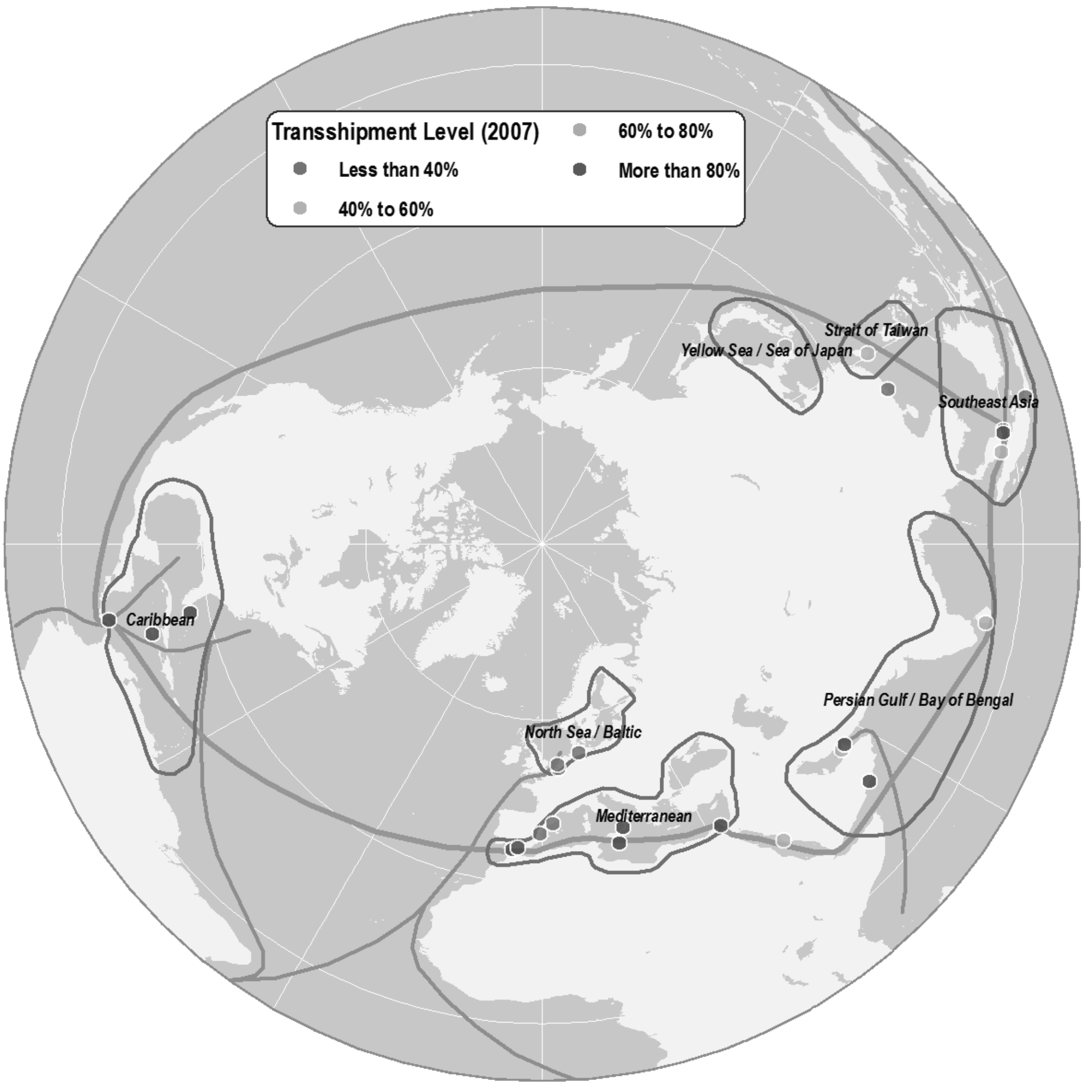
With globalization, the function of intermediacy has become increasingly prevalent for long distance transportation, particularly freight distribution (Fleming and Hayuth, 1994). Intermediate hubs have emerged since the mid 1990s in many port systems around the world: Freeport (Bahamas), Salalah (Oman), Tanjung Pelepas (Malaysia), Gioia Tauro, Algeciras, Taranto, Cagliari, Damietta and Malta in the Mediterranean to name but a few (figure 2). These hubs tend to have an excellent nautical accessibility, are located in proximity of major shipping routes, with some at the intersection of longitudinal and latitudinal routes to accommodate interlining/realay flows, and often have land for future expansion. Terminals are typically owned, in whole or in part, by carriers or global terminal operators which are efficiently using these facilities. An abundant literature exists on the role of intermediate hubs in maritime hub-and-spoke systems (see e.g. Baird, 2006; Fagerholt, 2004; Guy, 2003; Wijnolst and Wergeland, 2008; McCalla et al., 2005).

Figure 2: World's Main Intermediate Hubs, 2007



Yet, the creation of intermediate hubs does not occur in all port systems, but around specific regions which are ideally suited for maritime hub-and-spoke distribution patterns as a result of geographical, nautical and market-related factors (figure 3). Some markets seem to offer the right conditions for the emergence of more than one transhipment hub (e.g. the central Med). Other port systems do not feature any offshore hub development. In the US, many impediments in American shipping regulations gravitating around the Jones Act have favored a process of port system development with limited (feeder) services between US ports and the absence of US-based transhipment hubs (Freeport and other ports in the Caribbean to a limited extent take up this role). Also, northern Europe up to now does not count any real transhipment hub, let alone an 'offshore' hub. Hamburg, the North-European leader in terms of sea-sea flows, has a transhipment incidence of merely 40%, far below the high transshipment shares in the main south European hubs (85% to 95%). It is generally expected that the transhipment shares in newcomer ports Flushing and Wilhelmshaven might slightly exceed 40%. The only concrete plan for a real North-European offshore hub relates to a proposed transhipment facility at the natural deep-water harbor at Scapa Flow in the Orkney Islands. Baird (2006) argued that alternative port sites such as Scapa Flow could provide a superior and more competitive location from which to support the fast expanding transhipment markets of northern Europe. Notwithstanding such a plea, market players have not adopted the idea of bringing a north European offshore hub into reality.

Figure 3: World's Main Transshipment Markets, 2007



It is clear that the insertion of intermediate hubs within global shipping networks does not make the mainland load centers redundant. On the contrary, intermediate hubs multiply shipping options, which is the main reason why their rationale and dynamics has always been approached from a shipping line's perspective. As such, they play a fundamental role in the optimization of vessel movements, namely as hubs, relay or interlining locations within global shipping networks. As hubs, which account for about 85% of the global transshipment traffic, they are the points of convergence of regional shipping, essentially linking separate hierarchies and interfacing global and regional freight distribution systems. Both as relay and interlining locations, which account for 15% of the global transshipment traffic, they connect the same hierarchy levels and improve connectivity within the network. Some intermediary locations strictly perform cargo handling functions and have a non-existent hinterland.

Yet, the other benefits of intermediate hubs such as linking gateways, improving the connectivity of maritime shipping and acting as intermediary locations within global systems of circulation tend to be underestimated. It is the hubbing function of intermediate terminals that requires clarification in light of their role within regional port systems. It thus appears that the regionalization thesis needs to be expanded further to better reflect the distribution strategies of the hinterland and the foreland that are connected to a port gateway.

2.2. The Role of Intermediate Hubs in Global Shipping Networks

Intermediate hubs emerge in places where the hub-and-spoke and interlining/relay solutions offer clear advantages over direct port calls at mainland ports. They are particularly located along the equatorial round-the-world route (Ashar, 2002; De Monie, 1997). The global beltway pattern focuses on a hub-and-spoke system that allows shipping lines to provide a global grid of east/west, north/south and regional services. The large ships on the east/west routes will call mainly at transhipment hubs where containers will be shifted to multi-layered feeder subsystems serving north/south, diagonal and regional routes. Some boxes in such a system would undergo as many as four transshipments before reaching the final port of discharge. Much of the discussion on the hub port system (and consequently on regional container distribution) has focused on the deployment of large mainline vessels. Less attention has been paid to feeder vessels and to the hinterland transport modes, and yet both are fundamental in the dynamics of intermediate hubs. Part of the problem is that some ports, particularly those which serve as direct port of call, or even cater for some transhipment traffic, feel that serving feeder vessels means a loss of status.

Carriers have some feeder options available: direct feeders between hub and feeder port or indirect feeders via line-bundling loops including more than one feeder port. The first strategy has the lowest transit time but typically requires more feeders and smaller feeder vessels. Alternatively, indirect feeders benefit from economies of feeder vessel size, but incur longer distances and longer transit times. The carrier's choice between one or more direct calls at mainland load centres with the mother vessel or an indirect call via a feeder vessel is determined by factors such as the diversion distance, nautical conditions (such as draft), the volumes of containers involved, the possibility to combine transhipment activities with a strong cargo-generating power of the port's regional hinterland, the related costs, port productivity and the strength of the individual carrier in the markets served (see e.g. Zohil and Prijon 1999).

When point-to-point markets cannot support a direct service in the container trade, shipping lines need to make decisions about transhipment locations. A wide range of literature discusses the location and selection of optimal transhipment locations (see e.g. Fleming, 2000; Waals and Wijnolst, 2001; Baird. 2006; Ng, 2006; McCalla et al., 2005; Hayuth and Fleming, 1994; Aversa et al., 2005; McCalla, 2008; Lirn et al., 2004; Tai, 2005) and related hub operations (see e.g. Henesey, 2006). Hub-and-spoke networks would allow considerable economies of scale of equipment, but the cost efficiency of larger ships might be not sufficient to offset the extra feeder costs and container lift charges involved.

In referring to the Asian hub/feeder restructuring, Robinson (1998) argues that a system of hub ports as main articulation points between mainline and feeder nets is being replaced by a hierarchical set of networks reflecting differing cost/efficiency levels in the market. High-order service networks will have fewer ports of call and bigger vessels than lower order networks. Increasing volumes as such can lead to an increasing segmentation in liner service networks and a hierarchy in hubs.

2.3. The Vulnerability of Intermediate Hubs

Intermediate hubs as part of elaborate hub-and-spoke networks are often vulnerable to market changes. Two developments undermine the position of intermediate hubs as transhipment facilities: (i) container growth or decline, and (ii) new entrants to the transhipment market.

2.3.1. Container Growth and Decline

Port systems typically observe an increasing container volume as a result of worldwide growth in containerization. Particularly, the Asian and European container port systems have witnessed strong growth in the last decade, although volumes have plummeted from mid 2008 due to an economic downturn. Container port systems typically have to cope with higher volumes on specific routes, thereby supporting the development of direct end-to-end or line-bundling services that bypass transhipment hubs. In other words, the insertion of hubs can turn out to be just an intermediate stage in connecting a region to the global liner shipping networks. Once volumes for the gateway ports are sufficient, the utility of hubs diminishes. In extreme cases, a hub can become a redundant node in the network.

The Mediterranean Sea provides a good example. In the West Med, extensive hub-feeder container systems and shortsea shipping networks emerged since the mid 1990s to cope with the increasing volumes and to connect to other European port regions (see figure 4). Terminals are typically owned, in whole or in part, by carriers which are efficiently using these facilities. Marsaxlokk on Malta, Gioia Tauro, Cagliari and Taranto in Italy and Algeciras in Spain act as turntables in a growing sea-sea transhipment business in the region. These sites were selected to serve continents, not regions, for transshipping at the crossing points of trade lanes, and for potential productivity and cost control. They are typically located far away from the immediate hinterland that historically guided port selection.

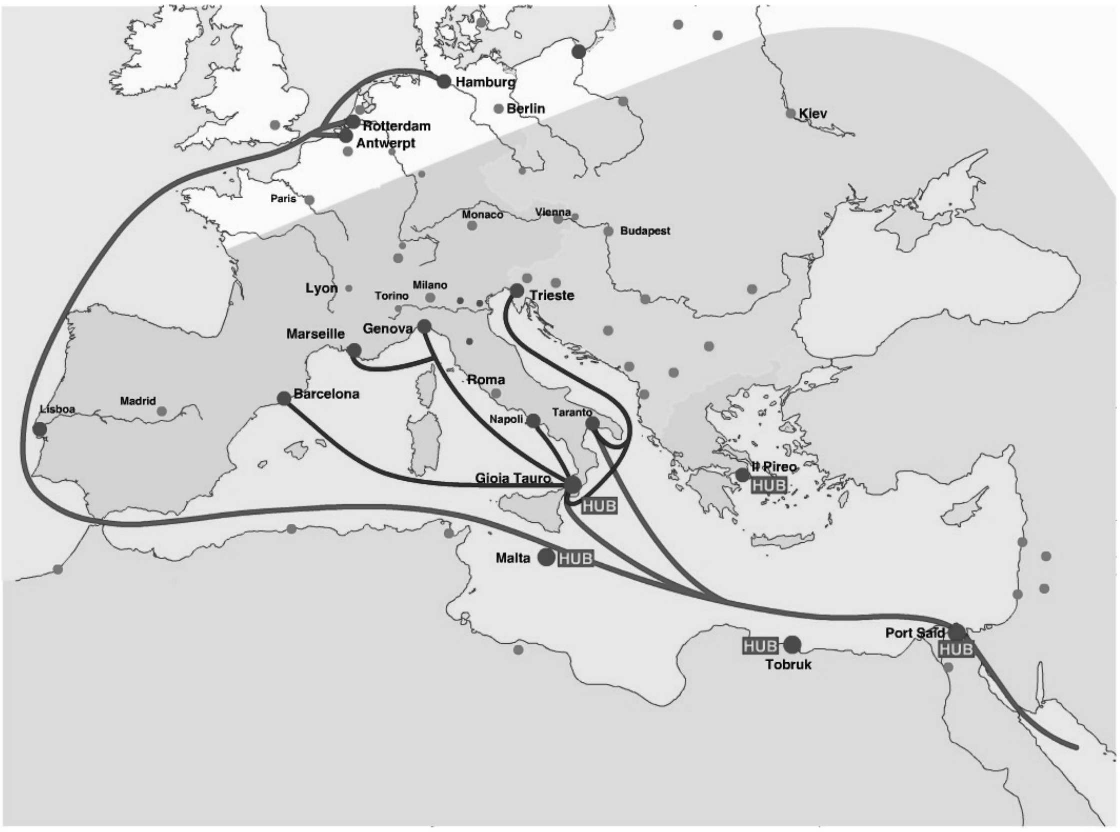


Figure 4: Transshipment hubs in the Mediterranean

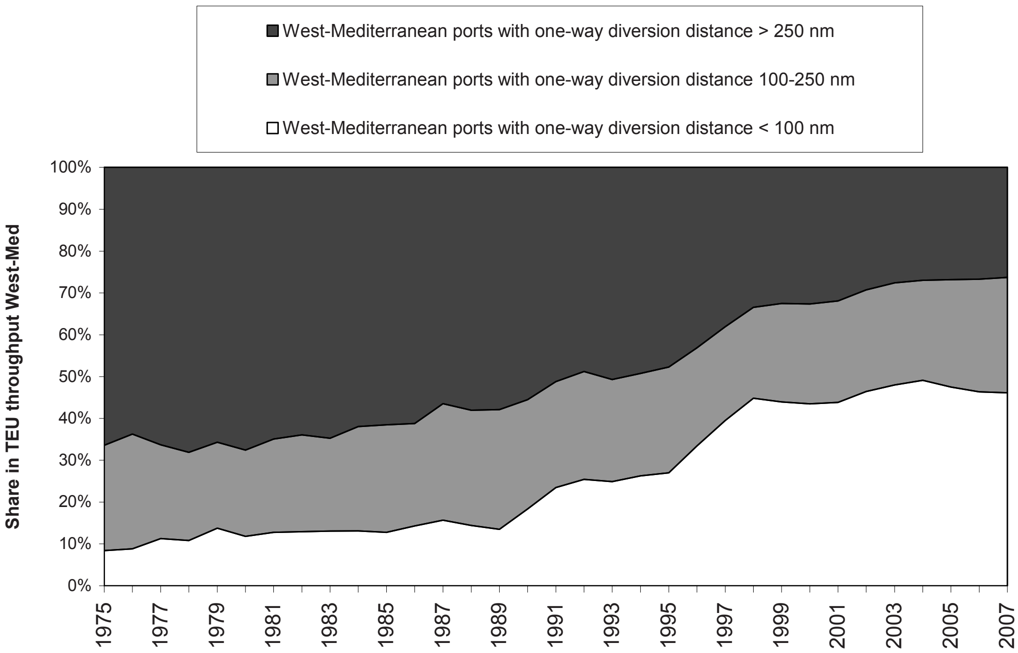


Figure 5: The market shares of ports in the West Mediterranean. Ports grouped according to the diversion distance from the main shipping route (1975-2007)

Source: Notteboom (2008)

The market share of the transhipment hubs in total European container throughput peaked in 2005 (12.2%) but since then started to decline to 11.4% as volume growth in mainland ports allowed shipping lines to shift to direct calls. While some shipping lines still rely on the hub-and-spoke configuration in the Med, others decided to add new line-bundling services calling at mainland ports directly. Maersk Line, MSC and CMA-CGM are modifying their service patterns, giving increasing priority to gateway ports. In reaction, mainly Italian transhipment hubs are reorienting their focus, now serving Central and East Med regions. Algeciras (stronghold of APM Terminals of the AP Moller Group) relies much on east-west and north-south interlining and is facing competition from newcomer Tanger Med where APM Terminals has also set up business recently. The net result of the above developments has been a slight decline in the market share of the West Med hubs in recent years (figure 5).

2.3.2. New Entrants

The transhipment business remains a highly 'footloose' business, mainly because along long distance shipping lanes there are several site options for intermediate hubs. New entrants in the market inevitably lead to a distribution of transhipment volumes over more players and nodes. There are two ways of entry with implications on existing transhipment facilities. First of all, competing shipping lines might decide, once volume has reached a certain threshold, to introduce direct services to gateway ports in the region on those links that support direct services. Such a cherry picking strategy can have a negative effect on cargo volumes on the spokes of the existing transhipment network and might even lead to the collapse of the whole hub-and-spoke system. Secondly, new transhipment facilities might be set up in the vicinity of existing hubs by competing terminal operators or shipping lines. The introduction of a competing network solution undermines established hub-and-spoke networks of incumbents in the region.

Consequently, bundling networks such as hub-and-spoke systems are vulnerable to changes in traffic volumes caused by new entrants in one or more of the spokes. The conversion of one link out of the hub-and-spoke network to a direct service, as a result of cherry picking by a newcomer or volume increases in the network, negatively affects the profitability and operational efficiency within the bundling system. As a result, once a hub-and-spoke network is installed, the shipping lines involved are continuously challenged to shift to a (downsized) system of direct calls or line-bundling systems.

Recent developments in the liner shipping industry have revealed the weaknesses of pure hub-and-spoke systems versus multi-port itineraries. As the feasibility of a hub-and-spoke system using intermediate hubs only bears fruits at a specific level of cargo volume, the system lacks flexibility to cope with strong increases in cargo volumes or a sudden collapse in volumes (due to market entry). A system of more loops with smaller vessels can provide an answer to these challenges. Carriers continuously review their strategy with respect to liner shipping networks, and this could very well introduce a tendency towards less transhipment and more direct port of calls (even for the bigger vessels). Gilman (1999) rightly stated that the networks operated by large vessels will continue to be based on end-to-end services. Hub-and-spoke systems are just a part of the overall scene.

3. Foreland-based Regionalization: In Search of Competitive Advantage

The vulnerability of intermediate hubs is partly the result of the narrow focus of these nodes on transhipment of containers only. The terminal operators in hubs mainly compete on basic resources such as location, nautical accessibility and terminal infrastructure (post-panamax cranes) and on the terminal productivity they can offer in terms of short vessel turnaround times. These sources of competition can rather easily be imitated by competitors in neighbouring locations, so that it is very hard to create a sustainable competitive advantage. Basic terminal handling activities no longer constitute the basis for competitive advantage as terminals have access to similar terminal technology (60% of the world's gantry cranes are provided by ZPMC in Shanghai, NAVIS is a world leader in IT system for terminal operations) which makes imitation more feasible. In developing a case study on the North Europe transhipment market, Ng (2006) demonstrated that the opinion of shipping lines in port attractiveness seems to be in accordance to their decisions on transhipment hub choices. By means of a Likert-style questionnaire, the study revealed shipping lines not only look at the monetary in assessing the attractiveness of transhipment facilities. Other factors, notably, time efficiency, geographical location and service quality also need to be taken into consideration. The importance of qualitative factors on top of infrastructure/equipment and pure cost factors is also echoed in recent studies on port selection criteria (see e.g. Lirn et al., 2004; Wiegmans et al., 2008; Chang et al., 2008 and Notteboom, 2009).

We argue that intermediate hubs have a better chance of consolidating their position in liner service networks when they are able to offer value-added services that go beyond mere transhipment activities. These additional services should be based on factors which are difficult to imitate by competitors, because they rely on a complex process of continuous improvement and enhancement and are strongly embedded in supply chain networks. Among the activities that could be involved, container transloading and customization offer opportunities to adapt products to regional market characteristics, while keeping the advantage of mass production upstream. Depending on the concerned supply chains, some intermediate locations may be able to integrate global production networks more effectively. For instance, a location such as Algeciras would be able to integrate supply chains having Asian, African and South American components. The hub or relay transhipment functions are thus multiplied with their closer integration within supply chain management. There is also an opportunity, particularly if transloading is already taking place, to interface different container leasing markets, enabling a better usage of containerized assets. Maritime shipping companies can maintain their assets in constant circulation within their networks while regional leasing companies can also maintain their assets within a more bounded, but also more flexible (e.g. demurrage) market. Repositioning costs are consequently reduced. Another multiplier could be the possibility to custom-clear cargo bound to a regional market at the intermediate hub, particularly if the hub is within the same trading block such as the European Union, NAFTA or ASEAN.

If a core competence is based on a complexity of technologies and skills it will be difficult for competitors to imitate and it will therefore have a higher probability of generating a competitive advantage and a less vulnerable position for the intermediate hub concerned. Depending of regional geographical characteristics, namely deviation from shipping lanes, there are several reasons why intermediate hubs are likely to play a more important role that goes beyond the conventional hub-feeder, interlining and relay functions.

First of all, economies of scale in maritime shipping have reached a level seriously undermining the serviceability of some ports, even those with significant hinterlands. Hubs offer advantages of consolidation. The consolidation of flows in a hub allows the development of artery routes served by large mainline vessels. The current economic downturn can help hubs, in the sense that shipping lines are looking for considerable savings (given the drop in freight rates), so they have an immense appetite for savings through route optimization.

Secondly, intermediate hubs are increasingly aware of the vulnerability of the transhipment business since bound to the strategies of shipping lines. Transhipment traffic and relay/interlining traffic are in essence linked to the decisions of the shipping line with regard to liner service network optimization. Transhipment activities do in principle not take place because of shippers' requirements. However, the development of transhipment activities can give incentive to shippers to redesign their own distribution networks. This leads to strategies at extracting more value from cargo passing through a stronger anchoring within the regional port system. For example, the development of distribution networks centered around main distribution facilities near the hub (value-added logistical activities) generates additional cargo flows to shipping lines. Hubs with a local cargo base have an edge here compared to ‘offshore’ locations, as the former are in a position to combine the transhipment/relay/interlining function with a gateway function.

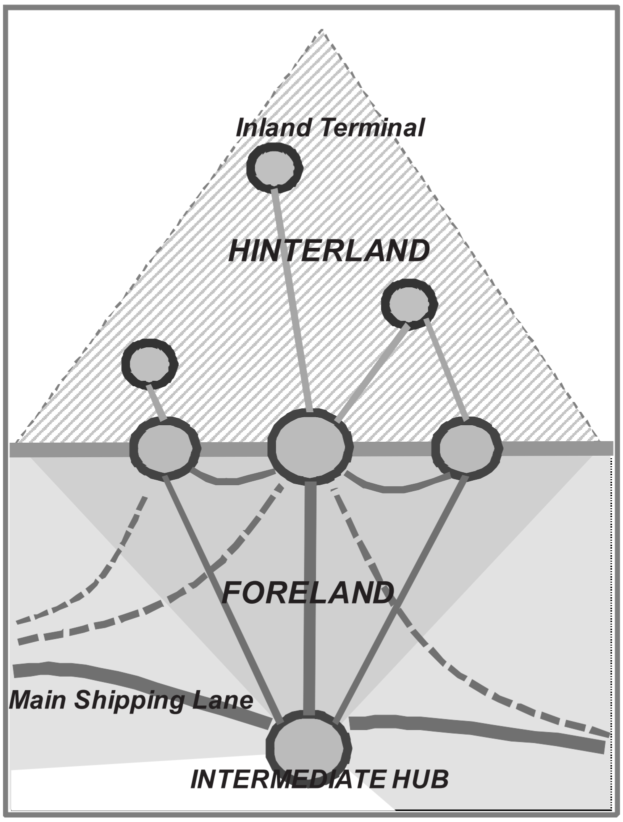


Figure 6: A graphical conceptualization of foreland-based regionalization

On the maritime segment, economies of scale and regional hinterland access need to be better reconciled. The concept of foreland-based regionalization refers to the integration of intermediate hubs in regional shipping networks where the maritime foreland of the intermediate hub is functionally acting as a hinterland (see figure 6). For various reasons (e.g. deviation, small volume, niche hinterland), some ports are not well connected to the global long distance shipping network and show limited opportunities to improve this connectivity since maritime shipping companies must consider an effective configuration of the networks that tend to focus on major gateways and intermediate hubs. The intermediate hub enables a level of accessibility and this accessibility incites them to look beyond their conventional transshipment role. This includes actions to extract more value of cargo passing through and as such get more economic rent out of a transhipment facility. Such strategies have led some transhipment hubs such as Gioia Tauro and Algeciras to develop inland rail services to capture and serve the economic centres in the distant hinterlands directly, while at the same time trying to attract logistics sites to the ports.

Theys et al., (2008) developed a model for determining seaport-located logistics activities. Their empirical application to the port of Busan clearly demonstrates that, in case ports are involved in providing logistics services to transhipment cargo, hinterland characteristics remain of fundamental importance in terms of logistics attractiveness. First of all, if ports are serving both their own hinterland as overseas markets, the extent to which a corridor is developed will influence the number of logistics facilities in the port that serve their own hinterland markets. For poorly developed hinterland corridors, the port will be ideally located for logistics facilities targeted to the hinterland, on top of the logistics centers that were already based in the port area for serving overseas markets. In case of well-developed corridors, on the other hand, logistics facilities for import/export cargo might move closer to consumption and production areas in the hinterland. Secondly, the hinterland connections of (smaller) ports that are served by feeder connections also affect the location possibilities that companies have for logistics facilities. When a port tranships cargo to smaller ports with a poorly developed hinterland, economies of scale and scope will make this port particularly attractive for locating logistics facilities targeted to those overseas areas. The overseas ports and their hinterlands will thus likely have relatively few logistics facilities. On the other hand, when the overseas port has a strongly developed corridor serving larger parts of its own hinterland, that port actually reduces the logistics attractiveness of the port through which it is fed. Logistics facilities then will be located either in the area surrounding the overseas port, or in its hinterland. When over time cargo throughput in such ports increases, they might be directly served.

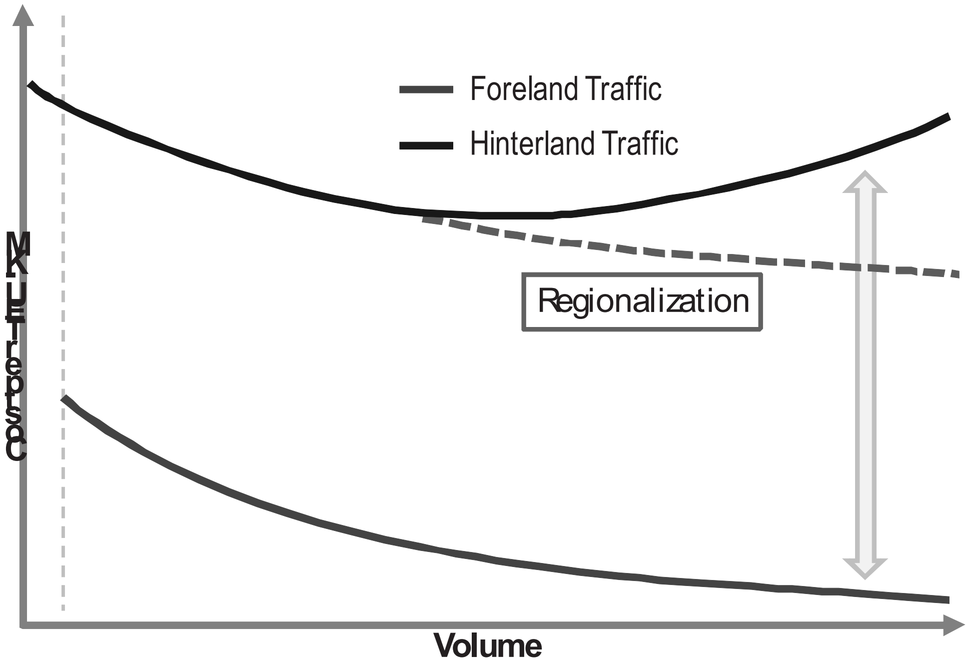
Some smaller gateway ports are not eager to be part of foreland-based regionalization strategies as it confirms their secondary or tertiary role in the global maritime shipping network. Ports tend to prefer direct ship calls within long distance pendulum networks. Yet, this is not always logistically feasible since the port volume may not justify a sufficient frequency. The positive outcome of foreland-based regionalization is that it enables to support a level of traffic which otherwise would not be feasible considering the network configuration of maritime shipping companies. Shippers must provide a network structure to cope with economies of scale as well as intermediate locations minimizing deviation from major shipping lanes with geography playing an important role in combination with liner network dynamics. Foreland-based regionalization is thus a step forward in reconciling the operational characteristics of forelands and hinterlands.

4. Foreland-based Regionalization: Reconciling Forelands and Hinterlands

Maritime shipping and inland freight distribution have evolved at two different momentums, both improving the level of functional and geographical integration of the global economy. Containerization has attained a large diffusion within supply chains and several niche markets, namely commodities, still have potential for development. Still, the role of maritime shipping in global supply chains remains to be further considered within the economic and transport literature. The design of liner shipping networks and hinterland networks are strongly intertwined (Notteboom 2004). Shipping lines have been able in the past to limit ports of call partly as a result of advances in large-scale intermodal transport combined with absorption pricing systems, but the load centres are only as competitive as the inland and relay links that connect to it. The optimal network design is not only function of carrier-specific operational factors, but also of shippers' needs (for transit time and other service elements) and of shippers' willingness to pay for a better service. The more cost efficient the network becomes from a carrier's perspective, the less convenient that network could be for the shippers' needs in terms of frequency and flexibility.

Thus, as indicated by figure 7, there is a growing disparity between the maritime and inland sides of the freight distribution equation as traffic volumes reach a certain level. Economies of scale of maritime shipping are a logical process, particularly as volumes increase, leading to larger loads between ports of call. Such a process is less clear for inland transportation since containerized freight must be broken into much small loads because of the geographical fragmentation of production or consumption and supply chain management. At some traffic level, diseconomies of scale result in congestion, difficulties to meet expected levels of service and higher costs.

Figure 7: Cost per TEU-KM for hinterland and foreland traffic



Any change, let it be market or technical, challenges the balance in terms of capacity and level of service between forelands and hinterlands. It also forces additional strategies to reconcile them, particularly if a change has impacted one system more than the other and thus led to additional imbalances. Over this, the full impacts of a new class of containerships above the 10,000 TEU threshold is subject to much speculation, namely their impacts on the number of ports of call and on the configuration of shipping networks. While hinterland-based regionalization has permitted inland freight distribution to keep up with volume and network configuration changes within maritime shipping, the new imbalances brought by new mega containerships may reinforce foreland-based regionalization.

At a regional level, several small or medium-sized ports may realize that it is in their long term interests to have a higher level of integration with an intermediate hub, even if it comes at the expense of less long distance pendulum services calls. Foreland-based regionalization would support export-oriented strategies with a better connectivity of more marginal (or in their early stage of growth) ports to global shipping networks and thus international trade. There are also site constraints, environmental factors or simple market potential that may put a limit to the volume generated by the hinterland of some ports.

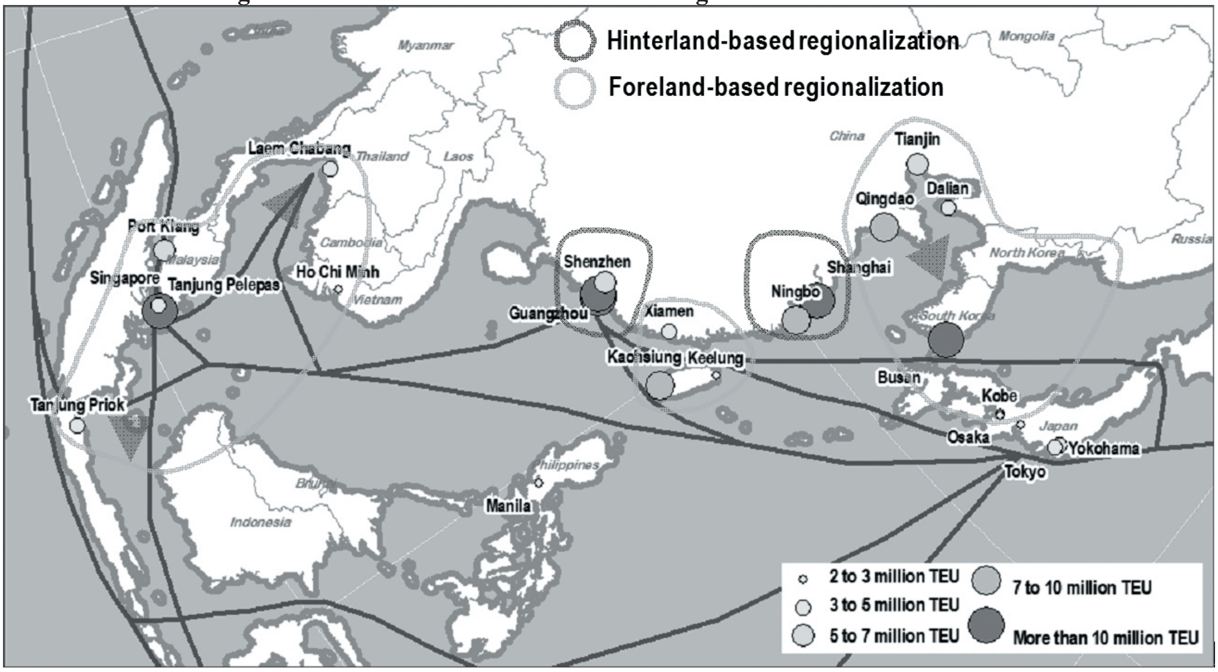
On the intermediate hub side, the volatile long distance transhipment traffic would be complemented with a more stable and secure regional traffic. Both the foreland and the hinterland are mutually self reinforcing since hinterland stability can anchor the volatility of the transshipment function, particularly in light of footloose operators. A better reconciliation between forelands and hinterlands would help insure that returns on investments are higher, subject to less fluctuations and that the competitiveness of a maritime range is improved.

The reinforcement of regional economic integration through trade agreements could lead to a better recognition of intermediate hubs within a common international trade and transport policy. The intermediate hub, through foreland-based regionalization, could become a functional gateway (port of entry) where shipments are cleared through customs before reaching its final ports of destination by a feeder service. Such a strategy would also help anchor at the intermediate hub added value activities, such as free trade zones, at a traffic level that would promote the cost effectiveness of regional distribution.

While foreland-based and hinterland-based regionalization can co-exist in the same region, each port system shows a dominance of one of the two approaches. A few examples:

- The container ports in the Baltic Sea in Europe rely on a foreland-based regionalization model to get access to the major trade routes. The major container ports in the Le Havre-Hamburg range (i.e. Hamburg, Bremerhaven, Rotterdam and to a lesser extent also Antwerp, Zeebrugge and Le Havre) serve as intermediate hubs. Mainline vessels call at these main ports while the Baltic ports receive the connecting feeder services. The observed routing practices are a result of the long diversion distance to Baltic ports for mainline vessels and the limited nautical accessibility of the Baltic Sea (Baltimax vessels). As container volumes in the region are growing strongly, major Baltic ports such as Saint-Petersburg (1.66 million TEU in 2007), Kotka in Finland (570,000 TEU) and Gdynia in Poland (614,000 TEU) are among the ports aiming for direct calls of deepsea vessels. Up to now, the market continues to rely on the hub-and-spoke solution centered on the intermediate hubs in the Le Havre-Hamburg range. The existing model goes hand in hand with the establishment of major European Distribution Centers (EDC) in the Benelux countries (near the main ports Rotterdam, Antwerp and Zeebrugge) and in the northern part of Germany. These EDCs also serve the Scandinavian and Baltic markets, mainly via the existing feeder networks. The port of Aarhus in Denmark and Gothenburg in Sweden, both situated near the entrance of the Baltic, have ambitions to increase their intermediacy role, both in maritime services and logistic activities, between the overseas markets and the Baltic;
- Next to their intermediacy vis-à-vis the Baltic port system, the main ports in the Le Havre-Hamburg range rely heavily on hinterland-based port regionalization strategies to serve the hinterland regions in Western-Europe and Central- and Eastern Europe. The respective port authorities have set up task forces together with various stakeholders (carriers, shippers, transport operators, labour and government bodies) to identify and address issues affecting logistics performance and to enhance collective actions and coordination (see also Van Der Horst and De Langen 2008 and De Langen and Chouly 2004; Notteboom 2006 and Notteboom 2008). The port authorities act as facilitator. Apart from the port authorities, also other organizations are adopting a role as facilitator in port regionalization issues (cf. representative bodies such as Alfaport in Antwerp and Deltalinqs in Rotterdam);
- Among the major trade gateways of Pacific Asia (figure 8), two major port clusters are dominated by hinterland-based regionalization where transshipment traffic is more limited: Hong Kong and Shanghai. This conventionally fits the export-oriented strategies of the major manufacturing clusters of the Chinese coast with facilities located in proximity to port terminals to insure an efficient access to global shipping networks since hinterland access tends to be poor. There is also the emergence of a foreland-based regionalization with three major intermediary hubs linked to an array of smaller ports; Pusan, Kaohsiung and Singapore. The intermediary role of Pusan and Singapore is reinforced by a detour index that the Yellow Sea and the Gulf of Thailand respectively impose on pendulum services. Supply chain integration set in place by foreign direct investments (e.g. Korea – Northern China and Taiwan – Central China) can also favor the use of the national intermediary hub. Also of significance are the prospects of economic integration, which could lead to the reconfiguration of some regional shipping networks, such as within ASEAN. It remains to be seen if the re-establishment of direct shipping links between Taiwan and mainland China will favor a foreland-based regionalization having Kaohsiung as the main hub, but this seems a likely possibility considering the existing high level of economic integration and the importance of Taiwanese shippers (e.g. Evergreen).

Figure 8: Foreland and hinterland-based regionalization in Pacific Asia



5. Conclusion: An Unfolding Paradigm?

The evolution of the role and function of intermediary hubs has resulted in a process that has been defined as foreland-based regionalization, which is the integration of intermediate hubs in regional shipping networks where the maritime foreland of the intermediate hub is functionally acting as a hinterland. Doing so insures greater traffic stability at the intermediate hub and enables smaller port to have access to global shipping networks.

In addition to the conventional risk of footloose operators switching traffic to another intermediate hub, market change risks (such as lower volumes) underline the need for intermediate hubs to mitigate these risks through a higher integration with their feeder ports. Concomitantly, there is growing evidence that containerization is entering a phase of maturity (Notteboom and Rodrigue, 2009) and that future volumes are likely to grow at a lower rate than previously observed. This trend is reinforced by global recessionary forces that have been unfolding since 2008, forcing maritime shipping lines to reconsider their networks. Decrease in shipping rates as well as lower cargo volumes along several pendulum routes are strong incentives to drop some lower volume port calls and place a greater focus on a hub-feeder structure for regional freight markets. Fewer assets are used while service frequency is maintained. In this context, foreland-based regionalization could become a more important strategy of regional competition within global freight distribution systems than it currently is.

Several issues remain to be addressed to substantiate the foreland-based regionalization thesis. Particularly, smaller ports may be reluctant players in such a system as it can be perceived that it would limit their growth opportunities and their capacity. Ports, whatever their size, aim at direct calls and some could attempt to also become an intermediary hub. In the present port competition model, also smaller ports are frequently tempted to making investment decisions of a speculative nature. Many container ports make significant investments without any degree of assurance that traffic will increase and shipping lines will retain their loyalty. Their only belief is that a lack of investments will certainly not increase traffic. Such supply-driven strategies could lead to overcapacity and a more vulnerable intermediate hub, since regional competition is more acute to secure this role. Thus, a good strategy for smaller ports is to link to

There is also the question pertaining to if foreland-based regionalization is simply a transition phase in port system development. There is evidence than it may not be the case and that foreland-based regionalization would be a distinct phase on its own. The massification of flows linked to vessel sizes leaves limited options outside major gateways and hubs. The prospects of higher energy prices would also favor intermediate locations as higher capacity and lower speed would be a characteristic of long distance shipping. Since maritime shipping companies and port operators are the actors making key asset management decisions, the emergence of intermediary hubs and the foreland-based regionalization they entail appears to be a conscious decision to establish, when suitable, such a network configuration linking regional and global shipping networks.

Further research should be done on the measurement of foreland-based regionalization. Network-related indicators (e.g. transhipment percentage, connectivity via relay/interlining operations) obviously provide a first indication of the reliance of a port system on intermediate hubs. However, empirical work on the foreland-based regionalization concept demands more and new indicators to measure the vulnerability of intermediate hubs in the network and the capacity of the hubs to extract economic rent through the development of value-added logistics services.

References

Ashar, A. (2002), Revolution now, Containerisation International, January 2002.

Aversa, R., Botter, R.C., Haralambides, H. and Yoshizaki, H. (2005), A Mixed Integer Programming Model on the Location of a Hub Port in the East Coast of South America, Maritime Economics and Logistics, 7, 1-18.

Baird, A.J. (2001), A new economic evaluation of the hubport versus multiport strategy, Proceedings of the IAME 2001 conference, Hong Kong, 138-166.

Baird, A.J. (2006), Optimising the container transhipment hub location in northern Europe, Journal of Transport Geography, 14(3), 195-214.

Barke, M. (1986), Transport and trade, Oliver & Boyd, Edinburgh.

Chang, Y.T., Lee, S.Y. and Tongzon, J.L. (2008), Port selection factors by shipping lines: Different perspectives between trunk liners and feeder service providers, Marine Policy, 32 (6), 877-885.

De Monie, G. (1997), The global economy, very large containerships and the funding of mega hubs, Port Finance Conference, London, June 26-27, 1997.

Fagerholt, K. (2004), Designing optimal routes in a liner shipping problem, Maritime Policy and Management, 31(4), 259–268.

Fleming, D.K. and Hayuth, Y. (1994), Spatial characteristics of transportation hubs: centrality and intermediacy, Journal of Transport Geography, 2(1), 3-18.

Fleming, K. (2000), A geographical perspective on the transhipment function, International Journal of Maritime Economics, 2 (3), 163-176.

Gilman, S. (1999), The size economies and network efficiency of large containerships, International Journal of Maritime Economics, 1(1), 39-59.

Guy, E. (2003), Shipping line networks and the integration of South America trades, Maritime Policy and Management, 30(3), 231–242.

Hayuth, Y. (1981), Containerization and the load center concept, Economic Geography, 57(2), 160-176.

Hayuth, Y. and Fleming, D.F. (1994), Concepts of strategic commercial location: the case of container ports, Maritime Policy and Management, 21, 187-193.

Henesey, L. (2006), Overview of Transhipment Operations and Simulation, MedTrade conference, Malta, April 6-7.

Lirn, T.C., Thanopoulou, H.A., Beynon, M.J., and Beresford, A.K.C. (2004), An Application of AHP on Transhipment Port Selection: A Global Perspective, Maritime Economics and Logistics, 6, 70-91.

McCalla, R., Slack, B. and Comtois, C. (2005), The Caribbean basin: adjusting to global trends in containerization, Maritime Policy & Management, 32(3), 245-261.

McCalla, R. (2008), Container transshipment at Kingston Jamaica, Journal of Transport Geography, 16, 182–190.

Ng, K.Y. (2006), Assessing the Attractiveness of Ports in the North European Container Transhipment Market: An Agenda for Future Research in Port Competition, Maritime Economics and Logistics, 8, 234-250.

Notteboom, T. (2004), A carrier's perspective on container network configuration at sea and on land, Journal of International Logistics and Trade, 1(2), 65-87.

Notteboom, T. (2006), Port regionalization in Antwerp, in: Notteboom, T. (ed.), 2006, Ports are more than piers, De Lloyd: Antwerp, 307-328.

Notteboom, T. (2007), Path dependency and contingency in the development of multi-hub port regions and multi-gateway port regions, in: Notteboom, T., De Langen, P., Ducruet, C. (eds), 2007, Ports in proximity: competition, cooperation and integration, Conference Proceedings, University of Antwerp / Erasmus University of Rotterdam, 83-104.

Notteboom, T. (2008), The relationship between seaports and the intermodal hinterland in light of global supply chains: European challenges, Discussion Paper No. 2008-10, OECD - International Transport Forum, 44 p.

Notteboom, T. and Rodrigue, J.P. (2005), Port Regionalization: Towards a New Phase in Port Development, Maritime Policy and Management, 32(3), 297-313.

Notteboom, T. and Rodrigue, J.P. (2009), The Future of Containerization: Perspectives from Maritime and Inland Freight Distribution, GeoJournal, 74(1), 7-22.

Notteboom, T. (2009), Complementarity and substitutability among adjacent gateway ports, Environment and Planning A, 41(3), 743-762.

Rimmer, P.J. and Comtois, C. (2009), China's container-related dynamics, 1990–2005, GeoJournal, 74(1), 35-50.

Robinson, R. (1998), Asian hub/feeder nets: the dynamics of restructuring, Maritime Policy and Management, 25(1), 21-40.

Rodrigue, J.P. (1999), Globalization and the synchronization of transport terminals. Journal of Transport Geography, 7, 255-261.

Rodrigue, J.P. and Notteboom, T. (2009), The Terminalization of Supply Chains: Reassessing the Role of Terminals in Port / Hinterland Logistical Relationships, Maritime Policy and Management, 36(2), 165–183.

Taaffe, E.J., Morrill, R.L. and Gould, P.R. (1963), Transport expansion in underdeveloped countries: a comparative analysis, Geographical Review, 53, 503-529.

Tai, H. (2005), Analysis of hub port choice for container trunk lines in East Asia, Journal of the Eastern Asia Society for Transportation Studies, 6, 907 – 919.

Theys, C., Ryoo, D.K. and Notteboom, T. (2008), Towards a generic framework for the development of logistics in seaports: lessons from the Busan case, Journal of International Logistics and Trade, 6 (2), 1-22.

Waals, F. and Wijnolst, N. (2001), Malacca-max: container shipping network economy, Proceedings of the IAME 2001 conference, Hong Kong, 191-207.

Wiegmans, B., Van Der Hoest, A. and Notteboom, T. (2008), Port and terminal selection by deep-sea container operators, Maritime Policy and Management, 35(6), 517-534.

Wijnolst, N. and Wergeland, T. (2008), Shipping innovation, Dynamar: Alkmaar.

Zohil, J. and Prijon, M. (1999), The MED rule: the interdependence of container throughput and transhipment volumes in the Mediterranean ports. Maritime Policy and Management, 26, 175-193.

Accounting implications for lease classification in acquiring transportation assets

Owan Tang

Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hong Kong.

Abstract

This article investigates the Accounting Standards on lease classification. Global financial tsunami makes bank borrowing difficult. When the US cut interest rate to near zero at Dec 16, 2008, the banks simply refused to respond to rate cuts. To raise funds, transportation companies shift the fund raising options from bank borrowing to equity financing. To attract public investors, companies make promise in the prospectus of not taking risky actions – translate in accounting terms – to maintain a healthy debt-equity ratio.

To bypass such promise, management buys assets and creatively classifies the transaction as a lease rather than a purchase. With such classification, management can engage in risky asset investment behavior regardless on the promise made in the prospectus.

This article studies the current standards set by the IASB (International Accounting Standards Board) and AASB (Australian Accounting Standard Board) on lease classification. In Australia, before 1983, accounting standards were prepared by the accounting profession and had no legislative backing. Since 1983, the Corporation Law has compulsory required Australian public companies to comply with approved accounting standard. Therefore, a violation of the accounting standards represents a non-compliance of the Corporation Law. The legal consequence makes lease misclassification an important topic for transportation companies.

Key words: Lease classification, accounting standards, IASB, AASB

1. Introduction

It is expected that transportation companies, such as airlines and ocean carriers, become difficult to borrow money from banks under the impacts of financial tsunami starting at the end of 2008. When the US cut interest rate to near zero at Dec 16, 2008, the banks simply non-respond to the rate cuts. To raise funds, transportation companies have to shift their fund raising options from bank borrowing to equity financing. To attract public investors, transportation companies need to make promise in the prospectus of not taking risky actions – translate in accounting terms – to maintain a healthy debt-equity ratio.

Purchase of a transportation asset, such as an airplane by an airline or a container vessel by an ocean carrier, would create a huge financial impact on a transportation company's debt-equity ratio. In order to keep the debt-equity ratio unchanged, transportation companies use long term leases rather than the traditional debt financing to acquire the assets. In other words, to the detriment of the public investors, management bypasses the debt-equity promise they made, buys assets and creatively classifies the transaction as a lease rather than a purchase. With such classification, management can engage in risky asset acquiring behavior in contrary to the promise made in the prospectus.

To counter with such misclassification accounting practice, the International Accountant Standard Board (IASB) issues the *IAS 17 Leases* to guide the accounting professionals. The analysis of *IAS 17 Leases* is beneficial because it has taken the task of designing a set of global accounting standards.[1]

When the acquisition of an asset is accomplished by a long term lease rather than through debt financing, it would create at least the following four aspects of differences:

1.1. Difference in financial accounting consequences
1.2. Difference in legal rights of the parties under commercial law
1.3. Difference in income tax consequences
1.4. Difference in legal obligations of the parties under bankruptcy law

In this article, the author will focus only on the difference in financial accounting consequences. This article studies the current standards set by the IASB and AASB (Australian Accounting Standard Board) on lease classification. The study of the Australian accounting standards is helpful to Hong Kong because both Australia and HK share the common foundation in building their accounting standards – both were once British colonies, and adopted the UK accounting standards from the very beginning. In term of subsequent developments, both Australia and HK have decided to harmonize their accounting standards with that of the IAS standards in the same year of 2005. Therefore, the development of Australian accounting standards provides a good reference to HK.

In Australia, the Corporation Law has compulsory required Australian public companies to comply with approved accounting standard. Therefore, a violation of the accounting standards represents a non-compliance of the Corporation Law. The legal consequence makes lease misclassification an important topic for transportation companies.

2. Incentives to Misclassify Leases

The simple term "lease" could cover different types of contracts. Since there exists a multitude of definitions in local GAAP and fiscal legislations, the author adopts the definition used in *IAS 17*, the international accounting standard for leases, where a lease is defined as an agreement whereby the lessor conveys to the lessee, in return for payment, the right to use an asset for an agreed period of time.

The classification of lease arrangement determines the accounting treatment of the transaction. The following table summaries the different treatments[2]:

	Finance Lease	Operating Lease
Asset	- Lessee records the leased item as asset at the inception of the lease - Lessee then reduces the asset amount by depreciation	Nil

[1] IASB's membership was representative of accounting standards boards, rather than of professional accounting bodies, across national borders; therefore, it is in the best position to undertake the task of drafting a set of global accounting standards.

[2] *IAS 17* specifies that "At the commencement of the lease term, the lessee shall recognize the finance lease as an asset and a liability in its balance sheet. The reported amount shall equal to the fair value of the leased item or, if lower, the present value of the minimum lease payments. The reported amount shall be determined at the inception of the lease. The discount rate used in calculating the present value of the minimum lease payments is the interest rate implicit in the lease. The lessee can add any initial direct costs to the recognized asset amount." Australian equivalent can be found on AASB 117-20.

Liability	- Lessee records the entire rental payments as a liability at the inception of the lease - Lessee then reduces the liability through each rental payments	Nil
Expense	- Lessee records only the depreciation expense, not the rental payment as expense	Lessee treats rental payments as expenses

The rationale of recognizing both the asset and liability for the leased item in a finance lease is that if such transaction is not reflected in the lessee's balance sheet, the economic resources and the level of obligations of lessee's company will be understated, thereby distorting the financial ratios. Therefore, the solution of rectifying such understatement is by recognizing in the lessee's balance sheet both as an asset and as an obligation to pay future lease payments.

The financial consequences of the different treatments may create a strong incentive for managers to classify a lease as operating lease rather than as a finance lease. Finance lease classification may create the following adverse impacts on a lessee's financial statements:

2.1. By recognizing the asset at the inception of the lease, it increases the amount reported in the non-current assets, and reduces the return on assets ratio.

2.2. By recognizing the present value of the entire future lease payments as a liability, it increases the amount reported in the non-current liabilities. This will in turn adversely affects the debt-equity ratios and liquidity-solvency ratios.

2.3. By recognizing the entire future lease payments as a liability, this may result in breaching the debt covenants, causing debts to become due immediately.

2.4. The subsequent depreciation and interest expenses may exceed the rental payments, this may reduce the reported profits.

2.5. Depreciation and interest expenses are not deductible for tax purposes.

To check with the incentives to misclassify leases, US issued the SFAS No. 13.[3] However, it fails to remove the underlying incentives of lease misclassification. For example, the US historically allows special purpose entities (SPE), such as Enron, to finance synthetic lease with 97% debt and 3% equity. The rule also allows a lessee to avoid consolidating the SPE if the 3% equity financing is provided by an outside entity unrelated to the lessee or lessor. Enron abused the 3% rule to avoid consolidating many of its SPE, and as a result, about $16 billion of debt was not shown on its balance sheet and hidden from investors.[4]

Drafters of accounting standards and academics have long debated about the complexity of the accounting treatments toward lease contracts.[5] For example, in Canada, when calculating minimum

[3] The provisions of SFAS No. 13: "derive from the view that a lease that transfers substantially all of the benefits and risks incident to the ownership of property should be accounted for as the acquisition of an asset and the incurrence of an obligation by the lessee and as a sale or financing by the lessor. All other leases should be accounted for as operating leases". (FASB 1976, para. 60)

[4] Batson, N.: 2003, Second Interim Report of Neal Batson, United States Bankruptcy Court, Southern District of New York, Case No. 01-16034 (AJG), Chapter 11, pp. 9–12.

[5] McGregor, W 1996, *Accounting for leases: a new approach – recognition by lesses of assets and liabilities arising under lease contracts*, FASB, July. The IASB and the US Financial Accounting Standards Board (FASB) published the Summary Report of the Leases Working Group Meeting on February 15, 2007 to recognize the difficulty in defining the operating and finance leases, and its negative impact of reporting economically similar transactions differently. http://www.fasb.org/board_meeting_minutes/10-07-08_leases.pdf

lease payments, all executory costs are excluded; while *IAS 17* only excludes service charges, taxes and reimbursements of expenses paid by the lessor on behalf of the lessee when calculating minimum lease payments. The US experience has shown that neither the Accounting Research Committee of the AICPA nor the Accounting Principles Board was able to build a consensus with this issue. FASB saw the subject of accounting for leases as one of its first priorities; nevertheless, despite multiple revisions including nine FASB amendments, six FASB Interpretations, 12 FASB Technical Bulletins, there is universal agreement that SFAS No. 13 fails to achieve its stated objectives.

3. Scope of Application - IAS 17 Leases

IAS 17 does not apply to licensing agreements, such as patents and copyrights.[6] Therefore, *IAS 17* applies to the tangible asset of an airplane, but it does not apply to the ultrasonic scanning systems used in structural flaws inspection of airplane parts during the repair stage.

Besides, *IAS 17* does not apply to service contracts.[7] In other words, when airplane manufacturer leases an airplane to an airline company, with an obligation to provide maintenance service; *IAS 17* does not apply to the service portion of the lease because it does not transfer the right to use the airplane.

4. When to Make the Classification?

IAS 17 specifies that lease classification should be made at the inception of the lease.[8] Lease renewal is not subjecting to the process of classification. If at any time the lessee and the lessor agree to change the provisions of the lease in a manner that would have resulted in a different classification, the revised agreement is regarded as a new agreement over its term, and such revised agreement is subject to classification.

However, changes in estimates (for example, changes in estimates of the economic life of the leased property) do not give rise to a new classification of a lease.

5. Economic Life under *IAS 17*

IAS 17 defines economic life in two ways: (1) it refers to the economically usable period of an asset; or (2) it refers to the number of production units expected to be obtained from the asset.[9]

Economic life is one of the important elements for classifying a lease under *IAS 17*. Since *IAS* 17 uses the risk and reward factors as the criteria to classify a lease, and risks and rewards are closely connected with an asset's economic life. *IAS* 17 defines risks as the possibilities of losses from idle capacity or technological obsolescence because of changing economic conditions.[10] Rewards are defined as the expectation of profitable operation over the asset's economic life.[11]

Since finance lease gives rise to depreciation expense for each reporting period through out the economic life of the leased item. The calculation of depreciation expense is particularly relevant for airliners because aircraft components have different economic life in tax law. The mechanical structure of an aircraft has a longer useful life expectancy up to say 25-years. Aircraft engines, on the other hand, have a shorter economic life, about 10-years. For landing gear, it normally has only 7-years of estimated useful-life expectancy.

---

[6] Australian equivalent can be found on AASB 117-2(b).
[7] Australian equivalent can be found on AASB 117-3.
[8] Australian equivalent can be found on AASB 117- 13.
[9] Australian equivalent can be found on AASB 117- 4.
[10] Australian equivalent can be found on AASB 117- 7.
[11] Ibid.

IAS 17 reminds accounting practitioners not to make judgment solely based on the legal form of a lease; in other words, economic life analysis provide a better alternative in judging the financial substance of a leasing arrangement. For example, a lawyer may conclude an arrangement is a lease and not a purchase because the lessee does not acquire legal title to the asset; an accountant will focus on the facts that if the lessee acquires the economic benefits of an asset for the major part of its economic life, the accountant should classify the arrangement as a purchase, in other words, a finance lease.

6. Guiding Principles of Classification - Finance Lease under *IAS 17*

Under *IAS* 17, a finance lease transfers substantially all the risks and rewards of an asset incidental to ownership.[12] Whether a lease is a finance lease or an operating lease depends on the substance of the transaction rather than the form of the contract.[13] *IAS* 17 illustrates with following five examples that would "normally" be classified as finance lease:

6.1. If a lease transfers ownership of the asset to the lessee by the end of the lease term, it is likely to be a finance lease.[14]

6.2. If a lessee contains an option to purchase the asset at a price that is sufficiently lower than the fair value, it is likely to be a finance lease. *IAS* 17 defines fair value as the amount for which an asset could be exchanged between knowledgeable and willing parties in an arm's length transaction.[15] (Checking the existence of an option is rather straight forward, because it can be ascertain within the four corners of the lease contract, however, the difficulty for an accountant is to make a professional judgment on whether at the inception of the lease, the lessee would likely to exercise that the option in the future.)

6.3. If a lease period covers the major part of the economic life of the asset, even the title is not transferred, it is likely to be a finance lease. *IAS* 17 defines lease period as the non-cancellable period, which includes the option to continue the lease.[16] (The difficulty for an accountant is to make a professional judgment on whether at the inception of the lease, the lessee would likely to exercise that the option in the future.) If the lease period is 12 years, and the economic life of an airplane is 20-years, it covers only 60% of the economic life and hardly be classified as a finance lease. However, if majority of the benefits is received within the first 12 years, the accountant has to make a professional judgment on whether it constitutes "major part of the asset's economic life".

6.4. If the present value of the minimum lease payments amounts to substantially all of the fair value of the leased asset, it is likely to be a finance lease. The calculation of the minimum lease payments is discounted to the inception of the lease.

6.5. If the leased assets are of a specialized nature that only the lessee can use them without major modifications, it is likely to be a finance lease.

7. Judgmental Nature of Lease Classification under *IAS 17*

Lease classification is essentially an exercise of professional judgment. The readers of *IAS 17* must bear in mind that the five examples shall not be seen as a definitive list of all situations that an accountant should consider; there may be other situations that give rise to a lease being classified as a

[12] Australian equivalent can be found on AASB 117- 8.
[13] Australian equivalent can be found on AASB 117- 10.
[14] In AASB 117- 4, it specifies that even title is not eventually transferred, so long as the lease transfers substantially all the risks and rewards, it still can be classified as a finance lease.
[15] Australian equivalent can be found on AASB 117- 4.
[16] Ibid.

finance lease that are not listed. *IAS 17* intentionally uses the word "normally" to remind the readers the existence of other possible exceptions.

8. Likelihood to identify Lease Misclassification

Lease misclassification is a type of fraudulent financial reporting; because by deliberately classifying a lease as operating lease, a company can understate its debt and make its reported profit looks better than it should be. Fraudulent financial reporting is a matter of grave social and economic concern (Kaminski, 2004). For auditors, a failure to detect false financial statements would cause legal liability and severe negative impacts to professional reputation, for instance, the failure to detect Enron's accounting fraud eventually led to the collapse of Arthur Andersen.[17]

Analytical procedures (APs)[18] have been recognized as a useful tool for detecting accounting fraud (Thornhill, 1995). APs involve the procedures of analyzing trends, ratios, and reasonableness tests derived from an entity's financial data (Albrecht, 2004). In the US, an auditor must perform APs in audit planning with an objective to identify the unusual events, amounts, ratios and trends (AICPA, 1988). However, accounting academics found that only 4 out of 24 fraud cases could be detected by APs (Blocher,1992). Therefore, without the exercise of ethnical sense and good judgment from the financial statements preparers, the likelihood to identify lease misclassification by external auditors may not be high.

Lease classification would involve a significant degree of second-guessing; and many preparers, instead of exercise their professional judgment, would rather choose to play the safe game by urging the accounting bodies to issue more bright lines rules, so that they could get ready answers to solve the detailed fact patterns they face. However, too much bright lines rules would eventually lead to a rules-based approach, which in turn adding complexity to the financial reporting system. Say Robert Herz, the FASB Chairman, that the U.S. system has too many rules and bright lines, too much detail; and they undermine professionalism, both in the preparation of financial statements and in auditing. (Kranacher, 2007)

In fact, in its July 2003 report to the US Congress, the US Stock Exchange Commission (SEC) urged the adoption of a principles-based accounting standards system. Under the ideal system, the FASB should only articulate the broader principles, then explain them, where possible, with real-world examples. The FASB should not get into every possible fact pattern or create unnecessary exceptions, in other words, FASB should try its best to stay away from creating bright lines (Kranacher, 2007).

9. Training on Professional Judgment

The final issue this article will address is the training of professional judgment. In lease classification, the preparer of financial statements has to determine at the inception stage, whether a lease arrangement be classified as finance or operating, this calls for the exercise of professional judgment. The issue here is whether we can bring students to the expert level though university training.

Even in the very ideal scenario, it will take about 10 years of sustained practice for a practitioner to reach expert level performance in professional judgment of his field (Ericsson, 1996). The ideal learning environment would be one that allows a great deal of relatively quick and clear feedback. In hospitals (and in medical school), there is usually an outcome where the patient gets better or worse in response to a treatment, so the medical student can get feedback and learn from it. However, accounting practice is not a field that could generate quick and clear outcome feedback. In accounting,

---

[17] Before the Enron accounting fiasco, Arthur Andersen reached a high of 28,000 employees in the US and 85,000 worldwide; after the Enron instance, the firm has downsized to only 200 employees, based primarily in Chicago, and most of their attention is on handling the lawsuits and for the orderly dissolution of the company.
[18] APs refer to the variety of techniques an auditor use to assess the risk of material misstatements in financial records.

we have to rely on a delayed review process to get feedback. Recent research indicates that auditors have very poor understanding of the technical knowledge of other auditors (Kennedy and Peecher, 1997; Tan and Jamal, 2006). Therefore, says professor Karim Jamal from University of Alberta, that as educators, we cannot bring our students to the expert level; we should get them ready to move out into a world of practice where they will reach peak performance some years later. However, we can't leave it all up to accounting practice, because of the comparatively weak learning environment that it provides. Professor Jamal did a research on reviewers' biases in accounting field and found that there existed numerous biases in reviewers' evaluation of their subordinates' work (Tan and Jamal, 2001).

10. Conclusion

This article investigates the accounting implications for lease classification, with a particular focus on IASB and AASB treatments on lease classification. When a manager structures a contract for leased asset in a manner that the company can enjoy benefits similar to outright ownership, but in a way to keep both the leased asset and related liabilities off the company's financial records, the unethical motives for lease misclassification contribute to many of the accounting debacles. [19]

To check with the lease misclassification, *IAS* 17 outlines a judgmental framework for accounting practitioners for decision making. The author submits that an Australian company that makes lease classification which complies with the Australian standards outlined in AASB 117 will substantially be in compliance with that standards issued by the International Accounting Standards Board. A reading of the *IAS* 17 shows that 20 percent of its paragraphs were devoted to the classification of finance lease, which aims not at creating a bright line lease classification rule for the accounting practitioners. The author further submits that the quality of professional judgment and the ethical behavior will serve the public much better, so as to faithfully perform the accountants' stewardship responsibility, than a technical expertise to just merely meeting a bright line rule for financial reporting.

In terms of training for professional judgment, the author submits that accounting academics need to do a better job in school because the actual accounting practice environment is negatively affected by delayed review process and biased reviewers, a poorer learning environment for developing professional judgment.

11. Acknowledgements

This research was supported under **Project Code: G-SAA8** by The Hong Kong Polytechnic University

References

AICPA (1988), Statement on Auditing Standards (SAS) No. 56: Analytical Procedures, American Institute of Certified Public Accountants, New York, NY.

Albrecht, W.S. and Albrecht, C. (2004), Fraud Examination and Prevention, South-Western Thomson Learning, Mason, OH.

Batson, N. (2003), Second Interim Report of Neal Batson, United States Bankruptcy Court, Southern District of New York, Case No. 01-16034 (AJG), Chapter 11, pp. 9–12.

[19] Thomas J Frecka, *Ethical Issues in Financial Reporting: Is Intentional Structuring of Lease Contracts to Avoid Capitalization Unethical?* Journal of Business Ethics, Dordrecht: Jun 2008, Vol. 80, Iss. 1; p. 45 (15 pages)

Blocher, E. (1992), 'The Role of Analytical Procedures in Detecting Management Fraud', Institute of Management Accountants, Montvale, NJ.

Bosco, W. (2006), Lease Accounting: Separating Myth from Reality. Equipment Leasing Association, White Paper.

Byrnes, N. (2006), 'You May Be Liable for That Lease; FASB's Review of Lease Accounting Standards Could Really Hammer Retailers', Business Week, June 5, 2006.

CPA Australia. (2008), Accounting Handbook 2008, N.S.W.: Pearson Prentice Hall.

Davis-Friday, P.Y. and Frecka, T. J. (2002), What Managers Should Know About Earnings Management – Its Prevalence, Legality, Ethicality, and Does It Work?, Review of Accounting and Finance 1, 57–71.

Dieter, R. (1979), Is Lessee Accounting Working?, The CPA Journal, August 1979, pp. 13–19.

Ericsson, K.A. (1996), The Road to Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports and Games, Mahwah, NI: Lawrence Erlbaum Associates.

Financial Accounting Standards Board. (1976), Accounting for Leases, Statement of Financial Accounting Standards No. 13 (Norwalk, Connecticut, FASB).

Frecka, T. (2008), Ethical Issues in Financial Reporting: Is Intentional Structuring of Lease Contracts to Avoid Capitalization Unethical?, Journal of Business Ethics, Dordrecht: Jun 2008, Vol. 80, Iss. 1; p. 45.

Imhoff, E.A., Lipe, R.C. and Wright, D.W. (1991), Operating Leases: Impact of Constructive Capitalization, Accounting Horizons 5, 51–63.

Kaminski, K., Wetzel, S., and Guan, L. (2004), Can Financial Ratios detect Fraudulent Financial Reporting?, Managerial Auditing Journal, Vol. 19 No 1, 15–28.

Kennedy, J., and Peecher. M.E., (1997), Judging auditors' technical knowledge, Journal of Accounting Research 35(2): 279-93.

Kranacher, M., (2007), An Exclusive Interview with FASB Chairman Robert H. Herz., CPA Journal, November 2007.

Lim, S.C., Mann, S.C. and Mihov, V.T. (2003), Market Evaluation of Off-Balance Sheet Financing: You Can Run But You Can't Hide, EFMA 2004 Basel Meetings Paper, Universitat Basel, Basel, Switzerland. December, 2003.

Lipe, R. (2001), Lease Accounting Research and the G4 + 1 Proposal, Accounting Horizons 15, 299–310.

McGregor, W. (1996), Accounting for leases: a new approach – recognition by lesses of assets and liabilities arising under lease contracts, FASB, July 1996.

Tan, H.T., and Jamal, K. (2001), Do auditors objectively evaluate their subordinates' work?, The Accounting Review 76 (1): 99-110.

Tan, H.T., and Jamal, K. (2006), How well do auditors know how others view them?, Contemporary Accounting Research 23 (3): 761-87.

Thornhill, W.T. (1995), Forensic Accounting – How to Investigate Financial Fraud, Richard D. Irwin, New York, NY.

Wyatt, A.R. and Gaa, J.C. (2004), Accounting Professionalism: A Fundamental Problem and the Quest for Fundamental Solutions, The CPA Journal 74(3), 22–32, March 2004.

Australian perspectives - port state control on marine pollution

Owen Tang

Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hong Kong.

Abstract

This article investigates the legal regime on port state control toward marine pollution. The discussion mainly focuses on marine pollution by the spills of oil and discharge of ballast water caused by substandard vessels. Since Australia is one of the most environmental conscious nations in the Asian-Pacific region, this paper explores and evaluates how the Australian domestic legal system incorporates the international legal regime on port state control in preventing marine pollution, with a particular emphasis on its legal treatment to detain foreign substandard vessels from assessing open seas.

Key words: Flag state control, port state control, Memoranda of Understanding, Australian distinction between unseaworthy and substandard vessel, power of detention, *Australian Navigation Act*, *Australian Pollution Act*.

1. Introduction

Oceans form an integral part of the global ecosystem. Oceans cover 71% of the earth's surface, contain 80% of the life on earth, and produce 70% of the world's oxygen supply.[1] Merchant ships for international trade are the major users of the oceans. Merchant shipping represents a significant part of international transport of goods through oceans, it accounts for more than 95% of world trade by weight.[2] Besides, oceans are essential to the international energy trade - about 60% of all oil produced for human energy needs is transported over the oceans.[3]

One of the major risks associate with shipping activities is the pollution of the marine environment due to accidental or deliberate discharge of oil, the sea receives approximately 3.2 million tons of oil annually, and close to half of this amount enters as a result of tanker operations or other accidental spills from ships.[4]

The likely damage upon a maritime casualty will affect not only the interested parties to a maritime adventure, namely the shipowners, carriers, and cargo owners; but the marine environment and nearby coastal communities. Accidental oil spills create very damaging effects to the coastal marine environments as the currents and tides in the coastal areas are not as strong as in the open sea. Researches indicate that at the site of an accident, the concentration of marine pollution will be highest, and will decrease progressively with distance by dispersion and dilution. Natural dispersion is fastest in the open sea, where the currents, tides, and winds are strongest, and slowest in stagnant waters.[5]

[1] D.W. Toews, J.J. Kay, and E. Lister (2008), *Ecosystem Approach: Complexity, Uncertainty, and Managing for Sustainability*, New York: Columbia University Press.
[2] V. Lun (2006), *Shipping and Transport Logistics*, Singapore: McGraw-Hill Education.
[3] United States Department of Energy, *International Energy Annual (2008)*.
See www.energy.gov/about/index.htm (last accessed in April, 16, 2009).
[4] Impact of Oil and Related Chemicals and Wastes in the Marine Environment, GESAMP Reports and Studies No. 50.
[5] Secretary of State for Transport (1994), Safer Ships, Cleaner Seas, The Report of Lord Donaldson's Inquiry into the Prevention of Pollution from Merchant Shipping, 141, at 23.

Discharge of ballast water constitutes another type of marine pollution.[6] The organisms from foreign waters, which are carried with the ballast water pose the risk of introducing a marine invasive species into the unique marine environment around the ports. In Australia, with approximately 95 per cent of Australia's commodities being transported by sea. Each year around 150 million tonnes of ships' ballast water is discharged into Australian ports by 10,000 ship visits from some 600 overseas ports. Most shipping into Australia arrives from the northern Pacific area, with the greatest volume of ballast water being discharged by bulk carriers.[7] The environmental impacts of invasive species can be significant.

This article will first explain why the flag state control legal regime is structurally unreliable in enforcing the international convention of safety measures on shipping activities; then the author will elucidate how port state control is the better sensible alternative to prevent marine pollution. Australia is one of the most environmental conscious nations in the Asian-Pacific region, this paper will evaluate how the Australia carries out the international port state control legal regime in preventing marine pollution, with a particular emphasis on how its legal treatment on foreign substandard vessels.

2. Maritime Common Law – The Negative Aspects of Liability Limitation Law

The Liability Limitation Law provides that if the ship causes injury without the owner's "privity or knowledge", then the liability of the shipowner is limited to the value of the ship and its "freight pending".

Historically, the Liability Limitation Law was designed to encourage investment in shipping and was particularly helpful to shipowner as a means of providing a ceiling for liability prior to the widespread acceptance of limitation of liability through incorporation.[8] The "privity" requirement has been linked to the fault-based notions of unseaworthiness,[9] so the basic idea can be recapitulated as a measure for protecting innocent shippers against unlimited liabilities. One of the most celebrated attempts to seek this liability shelter is the petition by the owner of the Torrey Canyon (responsible for a spill off the coasts of England and France in 1967 causing extensive environmental damage and cleanup costs of several million dollars) that sought to limit liability to fifty dollars, the value of a single surviving lifeboat.[10]

3. Shipping Conventions on Ship Safety

The sinking of the Titanic in 1912 led to the inter-governmental cooperation on formulating uniform laws for the safe operation of international shipping.[11] Today, the majority of laws regulating the construction, maintenance and operation of ships were generated under the auspices of the International Maritime Organization (IMO).

---

[6] Ballast water is carried in unladen ships to provide stability. At the ships' destination, the cargo is loaded and the ballast water is pumped out, organisms from foreign waters then establish populations in the surrounding waters of the loading ports. Many iron ore and coal carrying ships arrive empty of cargo and fully ballasted, so enormous volumes of foreign water are pumped into Australian ports.
Australian Museum Home Page www.austmus.gov.au (last accessed in April, 16, 2009).

[7] Australian Marine Environment Protection Committee (MEPC) (2008), *Briefing on Harmful Aquatic Organisms in Ballast Water*, 58th session: 6 - 10 October 2008.

[8] G. Gilmore & C. Black (1975), *The Law of Admiralty* §§ 10-1 to 10-3, at 818-21 (2d ed.).

[9] See *Tug Ocean Prince v. United States*, 584 F.2d 1151, 1155 (2d Cir. 1978).

[10] See *In re Barracuda Tanker Corp*., 281 F. Supp. 228 (S.D.N.Y. 1968). The claims "were eventually settled for three million pounds and the American limitation proceedings were discontinued."

[11] E Jansen (1991), *Governments' Responsibilities To Ensure That Ships Meet International Convention Standards in D Sanders* (Ed.) *The Management of Safety in Shipping*, The Nautical Institute, London.

The IMO regulations were mainly focusing on two broad areas: (1) The proper design, construction and maintenance of the structure and equipment of the ship [the hardware side]; and (2) The proper operation of the ships [the software side].[12]

4. Flag State Enforcement of International Safety Conventions

Historically, the flag states bear the primary legal obligation to regulate the safe operation of merchant ships. The government of the flag state is responsible for promulgating laws and regulations to effectuate its international obligations.[13] Flag state is a bearer of the primary legal obligation because international safety conventions[14] can only have effect at the intergovernmental level, and they cannot be enforced at the individual ship level. For example, in Australia and most of the common law jurisdictions, ratification of an international treaty does not automatically give the treaty document legal effect in domestic law. To have practical effect, the domestic legislature must incorporate the treaty provisions into its domestic law. Therefore, flag state is the national entity that can exert the greatest degree of legal control over the individual ship level. As recognized by the UNCLOS, a state is the best authorized body to "effectively exercise its jurisdiction and control in administrative, technical and social matters over ships flying its flag".[15] However, many of these nations do not possess sufficient resources to regulate the registered fleet. The mere reliance on flag states to prevent maritime casualties on the marine environment is impractical.

In fact, merchant ships in the recent times are registered under different "flag nations", and not all nations take their flag responsibilities very seriously. Some nations even have their vessel registries run by private corporations.[16] The "Flags of Convenience" ("FOC") nations have contracted out the administration of their fleets. The problem of substandard shipping is correlated with how seriously an individual ship register is administered.[17] The administration structure of FOC attracts irresponsible ship owners to shop around for ship registries with the lowest standards to avoid the costs of compliance with international regulations.[18]

5. Port State Control

In additional to the Limitation Act and FOC phenomenon, even responsible flag states would not have unlimited resources to enforce relevant international treaties on its fleet, which scattered throughout the world, on a continuous basis.[19] Therefore, port state control would have a key role to play even in the ideal world that all flag states intend to comply with their full responsibilities.

---

[12] The measures include the navigational rules, the training and certification of crews and criteria for safe handling of dangerous cargo.

[13] United Nations Convention on the Law of the Sea (UNCLOS).

[14] Conventions, protocols, codes and resolutions agreed under the auspices of the IMO, ILO or other similar multilateral or bilateral inter-governmental "gatherings".

[15] Article 94(1) of UNCLOS.

[16] FOC registers are often owned and managed by foreign nationals with headquarters located outside the flag state. For example International Registries Inc. which used to manage the Liberian Ship register (on 1 January 2000, IRI ceased acting for the Liberian registry) and now manages the Marshall Islands Registry, has its headquarters in Reston, Virginia, USA close to Washington DC. It was founded by Edward Stettinus, a former US Secretary of State and is a privately held company owned and operated by its senior employees.

[17] For example, in Australia, Flag State Convention (FSC) inspections are restricts to only the surveyors of the governmental agency - the Australian Maritime Safety Authority (AMSA). The AMSA guidelines specify that FSC inspections will be conducted at six monthly intervals for Australian flag ships. If tankers are of 15 years old, it requires FSC inspection at three-month-intervals. FSC inspection at three-month-intervals is mandatory for passenger ships regardless of age. *See*: *AMSA PSC Procedures*, *ITS63 Ship Inspection, Targetting of Ships* at 2.3.

[18] G.C. Kasoulides (1993), *Port State Control and Jurisdiction*, Dordrecht, Boston, at 185.

[19] Safer Ships, Cleaner Seas, The Report of Lord Donaldson, see supra note 5, at 57.

It is well established that when a vessel is in port, it will be subject to the laws of the "host" nation because the vessel is located within the sovereign territory of the coastal state. Port state control, therefore, has been traditionally limited to regulation of ships which have "moored" at a port, and this includes ships which have anchored, berthed alongside, are at a single point mooring or at an offshore facility.

Both UNCLOS and International Maritime Organization (IMO) made rules to strengthen the legal regime of port state control. The UNCLOS radically expands a port state's authority on investigating[20] and instituting proceedings against breach of international conventions committed by vessels outside a state's coastal jurisdiction,[21] namely its internal waters, territorial sea or exclusive economic zone.

In considering the party with most to lose as a consequence of maritime casualty is the coastal state adjacent to the site of the maritime accident, it seems logical that a mere reliance on flag state control to ensure compliance with relevant legislation is not sufficient, and an additional "check" by the port states is necessary. Port state control becomes an effective counterforce to the ever present external environmental threat posed by unseaworthy ships. The cost of port state control is well justified through a fee structure imposed for inspections and fines levied for breaches.

6. Enforcing of International Obligations of Port State Control

Under International Law, the concept of port state control requires a foreign vessel to comply both the laws of its own flag state but also those of the port state. In other words, even if the flag state is not a party to a particular international convention; if the law of a port state makes compliance of that particular international convention mandatory, the port state can enforce the foreign vessels for compliance if they are within the port state's sovereign territory.[22]

Conversely, if a state ratifies an International Convention, it will have an obligation under International Law to enforce relevant provisions as part of its port state control procedures, irrespective of whether such provisions are contrary to domestic legislation.

The real issue regarding substandard shipping and maritime casualties is less related to insufficient international legislation, but more to the fact that the relevant legislation is not properly complied with. Therefore, the problem is one of compliance and enforcement and not one of lack of detailed rules and guidelines. Now the question becomes: If a central government ratifies an international obligation, can the local government (where the port locates) refuse to comply?

In the United States, the issue relates to the concept of vertical preemption and its legal effect on state laws. For example, the *US Federal Water Pollution Control Act* disclaims an intent to preempt the states "from imposing any requirement or liability with respect to the discharge of oil."[23] As a result,

---

[20] Article 218 of UNCLOS provided that:

(1) When a vessel is voluntarily within a port or at an off-shore terminal of a state, that state may undertake investigations and, where the evidence so warrants, institute proceedings in respect of any discharge from that vessel outside the internal waters, territorial sea or exclusive economic zone of that state in violation of applicable international rules and standards established through the competent international organization or general diplomatic conference.

[21] Port state jurisdiction is provided in Article 25 of UNCLOS that:

(1) The coastal state may take the necessary steps in its territorial sea to prevent passage which is not innocent.

(2) In the case of ships proceeding to internal waters … the coastal state also has the right to take the necessary steps to prevent any breach of the conditions to which admission of those ships to internal waters … is subject.

[22] For example whether a ship is seaworthy or not will be determined according to the provisions of municipal law, which clearly illustrates the importance of uniformity at an international level.

[23] See Section 311(o), 33 U.S.C.A. § 1321(o), reading:

the state would be free to provide additional requirements or penalties not specified by federal law. In the US, the states regularly impose damage assessment charges, penalties, cleanup costs, and other obligations not specified by federal law.

In Australia, the *Commonwealth v State of Tasmania* (*Tasmanian Dams Case*)[24] affirmed the pre-eminence of international treaty obligations over contrary state legislation.

The *Tasmanian Dam Case* was a landmark decision in Australian constitutional law, which centered around the proposed construction of a hydro-electric dam on the Franklin River in Tasmania, which was supported by the Tasmanian government (the state legislation), but opposed by the environmentalist groups. A four to three majority of the court held that the federal government had legitimately prevented construction of the dam.

One of the legal issues concerns Section 51(xxix) of the *Australian Constitution*, which gives the federal parliament the power to make laws with regard to external affairs. Section 51(xxix) is a nebulously defined provision. The federal government passed a law[25] under this provision to prohibit the Tasmanian government to clear and excavate the area for building the dam. The Australian federal government claimed that the law was enacted for fulfilling the obligations of an international treaty[26] to which Australia was a party. The Tasmanian government argued that the *Australian Constitution* gave no authority to the federal government to make such regulations. Both governments put their case to the High Court of Australia in 1983.

The High Court recognized that the fact that when the *Australian Constitution* came into effect in 1901, there were few international organizations such as the United Nations in existence. The external affairs power under the Australian Constitution was intended to be ambiguous, which would give it the capability of expansion. The High Court further explained that so long a federal law implements an international law or treaty, it is sufficient that it acquires the international character under Section 51(xxix).

7. Australia's Role in the Tokyo Memoranda of Understanding (Tokyo MOU) Regional Initiatives--The Memoranda of Understanding (MOU) in Port State Control

The sharing of information is crucial to the success of port state control regime. Without such information sharing, port state control may impose an undue burden on shipping activities when the same ships are inspected at every port. To facilitate information sharing, port states connect their control activities by establishing memoranda of understanding ("MOU's"). The first regional grouping was the 1982 Paris MOU – the Memorandum of Understanding on Port State Control in Implementing Agreements on Maritime Safety and Protection of the Marine Environment ("Paris MOU"),[27] which set a framework for subsequent Tokyo MOU, which Australia is a member.

The Tokyo MOU was establish for the Asian-Pacific region, which binds the maritime authorities of Australia, Canada, China, Fiji, Hong Kong, Indonesia, Japan, South Korea, Malaysia, New Zealand, Papua New Guinea, Philippines, Singapore, Solomon Islands, Thailand, Vanuatu, and Vietnam.

(2)Nothing in this section shall be construed as preempting any State or political subdivision thereof from imposing any requirement or liability with respect to the discharge of oil or hazardous substance into any waters within such State.

[24] (1983) 158 CLR 1.

[25] *World Heritage Properties Conservation Act* 1983.

[26] Convention Concerning the Protection of the World Cultural and Natural Heritage.

[27] The Paris MOU binds the maritime authorities of Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, U.K. and Northern Ireland. The Russian Federation became a member on January 1, 1996.

The MOU recognizes the need to avoid unhealthy competition between ports, and acknowledges the necessity for setting up a harmonized system of port state through exchange of information. The MOU requires each contracting member to inspect an annual total of 25 percent of foreign merchant ships calling at its ports. The MOU also provides guidelines for inspection and detention of substandard vessels for the purpose of insuring rectification of defects in the vessels. Since flag states bear the primary responsibilities to ensure compliance of international safety conventions for shipping, the basic premise is that where a vessel has a valid certificate issued by the flag state, it is prima facie evidence of compliance with relevant convention requirements.[28]

Accordingly, the initial task of the Port State Control Officer (PSCO),[29] on boarding a foreign vessel, is to exam its relevant certificates and documents.[30] If any of the certificates have expired or are invalid, there will be clear grounds for exercise the next level of control procedures.[31]

However, the "clear grounds" entails more than merely lacking valid documents issued by the flag state. Clear grounds are defined as:[32] "*Evidence that the ship, its equipment, or its crew does not correspond substantially with the requirements of the relevant conventions or that the master or crew members are not familiar with essential shipboard procedures relating to the safety of ships or the prevention of pollution*."[33] Paragraph 2.2.5 of IMO Resolution A.787(19) stipulated that if the PSCO believes that the ship or its crew do not substantially meet the requirements, the PSCO should proceed to a more detailed inspection. After the inspection, if a vessel has deficiencies that are hazardous to safety, health or the environment, the member shall not allow the ship to proceed to sea unless it first removes the hazardous defects.[34]

To prevent ship owners from running a detention, the MOU stipulates that other members shall refuse such ships to access their ports until the ship owner can provide evidence that the defects have been rectified. Exception is allowed where a ship needs to proceed to a repair port.

However, the exemption could be used as a way to avoid a detention order. For example, when BV withdrew the class of the Cypriot panamax bulk carrier San Marco, the vessel was detained by Vancouver port authorities. The vessel was allowed to proceed under tow, unmanned, for repairs in Mexico. But no repairs were undertaken. The vessel slipped her tow, took her crew back on board, and proceeded to load a full cargo of fertilizer. During this voyage, the vessel hit heavy weather off Cape Town and lost shell plating 14x7m.[35]

In terms of information sharing, the Tokyo MOU requires each member to publish quarterly data on detentions, with information about the name of the ship, its owner and operator, flag state, and classification society. In light of the publishing data, just like the Paris MOU, the Tokyo MOU recommends its members to avoid inspecting ships which have been inspected by other members within the previous six months unless "clear grounds" for inspection exist.[36]

Publication allows the shipbrokers to know what ships have been detained and why. It also allows the marine underwriters to better assess the risks of those substandard vessels. It lets the cargo

---

[28] See SOLAS Regulation I – 19(b).
[29] Port State Control Officer is the authorized person from the competent authority of a Party State to a relevant convention in carrying out port State control inspections. Paragraph1.6.6. of IMO Resolution A.787(19).
[30] IMO Resolution A.787(19) para 2.2.3.
[31] When a PSCO exerts port control activities on a vessel that could not provide valid documents issued by its flag state, the flag state would not be embarrassed because the flag state should welcome the intervention by the Port State authorities, as the vessel is in breach of its obligations to the Flag State.
[32] Paragraph 1.6. of IMO Resolution A.787(19).
[33] Ibid, paragraph 2.3 lists 10 examples of "Clear grounds".
[34] Paragraph 3.7 of Paris MOU Paris MOU.
[35] John Hare (1997), *Port State Control: Strong Medicine to Cure a Sick Industry*, 26 GAJICL 571.
[36] Paragraph 3.4 of Paris MOU provides that "Clear grounds" includes notification by another authority or complaint of the ship's master, crew or any person "with a legitimate interest in the safe operation of the ship."

owners know who the delinquents are, and shipper can avoid putting their cargoes onto the substandard ships.[37] With the ease of dissemination of information through the internet, the public can find detention lists on a monthly basis in Lloyds List (U.K., Australia, Canada and the U.S) and even on the internet.[38] The information sharing decreases the possibility of unseaworthy ships entering into unsuspecting ports.

8. Australian Safety Maritime Authority (AMSA)

Australia has designed one of the most comprehensive domestic legislation to carry its port state control program in the Asian-Pacific region. The Australian Safety Maritime Authority (AMSA) conducts port state control in Australia, and as a member of the Asia-Pacific MOU, Australia does more than comply with its 25% inspection target. In 1996, it inspected 2,901 vessels, of which 248 were detained. Australia has published data about delinquent flags, substandard classification societies, with details of detentions indexed by ship type, on the AMSA website on a monthly basis.[39]

In the mid-2008, the Council of Australian Governments (COAG) agrees to consider the establishment of a single national system for maritime safety regulation.[40] Currently across Australia, maritime safety is regulated by more than 50 pieces of legislation administered by eight independent maritime safety agencies. Under the proposed reform, the AMSA would responsible for regulating vessel design, construction, and equipment, vessel operation (e.g. safety management systems), and crew certification and manning.

The legislative basis of the AMSA's inspections is based on the Section 210 of the *Commonwealth Navigation Act 1912*.[41] Section 210 provides that if it appears to the AMSA that a ship is unseaworthy or substandard, the AMSA may order the ship to be provisionally detained. The AMSA must then issue a report as to whether the ship is unseaworthy or substandard. The AMSA will issue a report to the ship master on whether the ship be finally detained or be released unconditionally.

9. Australian Distinction between an "Unseaworthy" and a "Substandard" vessel

In IMO Resolution A.787(19),[42] the terms "unseaworthy" and "substandard" ships are used interchangeably. However, the two terms do not mean the same thing in Australian legal context.

In carriage of goods by sea, there is an obligation upon the carrier to provide a seaworthy ship.[43] For example, the Australian Insurance law[44] clearly stresses the importance of seaworthiness in contracts of marine insurance. Section 59 of the *Navigation Act* provides that in every contract of service between a ship owner/master and a seaman, there is an obligation upon the ship owner/master to exercise reasonable care to ensure that the vessel is in a seaworthy condition at the commencement of, and throughout, every voyage. Section 207 of the *Navigation Act* defines "seaworthy" as a fit state to

---

[37] John Hare, *Port State Control*, see super note 35, at 580.

[38] The public can find monthly detention lists in the following websites:

- The United Kingdom at Marine Safety Agency www.detr.gov.uk/msa/det97/det97.htm
- Australia at AMSA PSC Statistics www.amsa.gov.au/sp/shipdet/sdetlink.htm
- The United States at United States Coastguard www.uscg.mil/hq/g-m/psc/detained.htm

[39] AMSA Home Page, www.amsa.gov.au (last accessed in April, 16, 2009).

[40] National Approach to Maritime Safety Regulation, www.amsa.gov.au/namsr/ (last accessed in April, 16, 2009).

[41] Section 210 of the *Commonwealth Navigation Act, 1912* (Austl.) (detention of unseaworthy and substandard ships).

[42] IMO Resolution A.787(19) provides that "A ship whose hull, machinery, equipment, or operational safety is substantially below the standards required by the relevant convention or whose crew is not in conformance with the safe manning document".

[43] Schedule 1, Article 3 r1(a) of *Carriage of Goods by Sea Act 1991* (Cth).

[44] Section 45 of the *Marine Insurance Act 1909* (Cth).

encounter the ordinary perils of the sea. On the other hand, section 207A of the *Navigation Act* states that substandard has a different meaning:

- (1) A ship is substandard if the ship is seaworthy, but conditions on board the ship are clearly hazardous to safety or health.
- (2) In determining whether a ship is substandard, regard shall be had to such matters as are prescribed.[45]

Therefore, even a brand new vessel with all necessary equipment can be "substandard". The Australian High Court decision in *Great China Metal Industries Limited v Malaysia International Shipping Corporation*[46] opined that when evaluating seaworthiness in a carriage contract, the Court shall consider more than just the physical condition of the ship, but also the overall management of the vessel. Courts could find that a "brand new", but badly managed ship, as unseaworthy.

10. Power of Detention based on *Australian Navigation Act*

The *Navigation Act* gives AMSA inspectors the power to detain unseaworthy and substandard foreign ships on the following legal grounds:

- (1) If the ship is not manned with the minimum manning requirements;[47]
- (2) if the provisions and (potable) water are not of good quality;[48]
- (3) If the ship has incorrectly positioned load line markings;[49]
- (4) If the number of persons with appropriate radio operating certification does not comply with the requirements.[50]
- (5) If the ship carries particular cargo which is deemed to affect its safety.[51] This could include even fairly innocuous goods such as grain or slurry.[52]

If detailed inspection reveals that the actual condition on board does not correspond with the relevant certificate, Section 210 of the *Navigation Act* also authorizes the detention, even the vessel does possess valid certificates.[53] Section 210 even permits provisional detention without actual physical inspection if a ship appears unseaworthy or substandard from an external visual appraisal or from report of a PSC member. However, the power vested in section 210 is likely one of detention, not of arrest. Accordingly, the detention power exercised by AMSA under the *Navigation Act* is restricted to preventing substandard / unseaworthy vessels from departing from Australian ports, but do not extend to the right of arresting a substandard / unseaworthy vessel innocently passage through Australian maritime territory.

11. Detention Power of Foreign Ships based on *Australian Pollution Act*

The *Australian Pollution Act* grants a far more extensive power to AMSA to regulate substandard shipping in cases of actual or suspected marine pollution. The *Pollution Act*, unlike the *Navigation Act*, actually authorizes the AMSA to detain a foreign ship, if there is "clear grounds for believing that a pollution breach had occurred in the Australian territorial sea or Economic Exclusion Zone (EEZ) [54]" that is related to that foreign ship, even if the ship is in all respects compliant at the time.

[45] These are contained in *Marine Orders Part 11- Substandard ships- Issue 2*.
[46] (1998) HCA 65.
[47] Section 14(9)(a) of the *Commonwealth Navigation Act, 1912* (Austl.).
[48] Ibid., Section 120(2).
[49] Ibid., Section 227C.
[50] Ibid., Section 231D.
[51] Ibid., Section 254(2).
[52] Potential free surface movement that would reduce the available stability margins beyond acceptable limits.
[53] Section 210 of the *Commonwealth Navigation Act, 1912* (Austl.).
[54] On August 01,1994, Australia declared an EEZ extending 200 nautical miles from its coastline. Within its EEZ, Australia has sovereign rights to conserve and manage the living (e.g. fisheries and genetic material) natural resources. It also has jurisdiction over offshore marine scientific research and the protection and preservation of the marine environment. Australian Department of Agriculture, Fisheries and Forestry Home Page, www.daff.gov.au/fisheries/domestic/zone (last accessed in April, 16, 2009).

The *Pollution Act* also grants AMSA the power to detain and escort foreign vessels in the territorial sea and the EEZ into Australian port if the ship is suspected of causing the pollution. This provision is apparently contrary to the concept of "the right of innocent transit", but the Australian legal scholars submit that the legislation is valid under UNCLOS Part XII for the reason that the act of causing pollution renders the transit non-innocent. Subject to UNCLOS Article 228,[55] Australian authority may prosecute against a foreign ship for polluting Australian breaches, up to three years after the breach, with service on the agent of the ship as deemed to be served on the owner or master.[56]

12. Conclusion

Liability limitation law and the failure of flag state control make port state control a better legal alternative to prevent marine pollution by substandard vessels. Both UNCLOS and IMO made rules to facilitate the legal regime of port state control. Australia, through the *Tasmanian Dams Case*, places the enforcement of international law obligations a priority over conflicting domestic laws, which creates a relative effective legal framework on disallowing the substandard vessels to access the international waters once they entered the Australian ports. In addition, as a member of the Tokyo MOU, Australian publish the detention list of all substandard vessels with their flag states and classification societies, which let the public at large knows about who the miscreants are.

Australia even distinguishes an unseaworthy vessel from that of a substandard vessel, and the port state control authority can detain even a brand new foreign vessel if the management side deficiency makes it substandard vessel. The *Australian Pollution Act* grants more power to the port authorities than the *Navigation Act* in detaining and escorting pollution suspected foreign vessels even 200 nautical miles from the Australian coastline.

Acknowledgment

This research was supported under **Project Code: G-YH37** by The Hong Kong Polytechnic University.

[55] Section 29 of the *Australian Pollution Act*.
[56] Ibid s.29A.

Bankruptcy Courts' power on disposal of maritime assets

Owan Tang

Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hong Kong.

Abstract

This article investigates jurisdiction on US bankruptcy courts relates its power on disposal of a maritime asset. Global financial tsunami makes transportation companies short of liquidity, and the disposal of underutilized assets would be a sensible way for bankruptcy courts to bring cash inflow for both the reorganization or liquidation objectives.

However, when a shipping company files bankruptcy in the US, the assets treatments will involve two exclusive federal court systems: the admiralty and the bankruptcy court systems.

This article, through a study on relevant bankruptcy cases, investigates whether a bankruptcy court (a legislative court) can administer the debtor's maritime assets, which traditionally could only be dealt with by the admiralty court (a judicial branch of the government).

Key words: Bankruptcy Code, disposal of a maritime asset by bankruptcy court, maritime lien, admiralty jurisdiction, doctrine of Custodia Legis

1. Conflicts between Bankruptcy and Admiralty Jurisdiction

When a shipping company files bankruptcy in the US, the assets treatments will involve two exclusive federal jurisdictions: admiralty and bankruptcy.

In the US, by referring to the article of the US Constitution from which the court's authority stems, courts can be divided into Article III courts and Article I courts. The Article III courts are "constitutional courts," which were first created by the Judiciary Act of 1789. Article III courts constitute the judicial branch of the government (which is defined by Article III of the Constitution). Under the US Constitution, Article III protects the courts against influence by the other branches of government; for example, the salaries of judges from Article III courts may not be reduced during their tenure in office, and their appointments are for life, only subject to impeachment for bad behavior. Examples of Article III Courts are:

- Supreme Court of the United States
- United States courts of appeals
- United States district courts
- United States Court of International Trade

Article I courts, which are "legislative courts," which are regulatory agencies. Since Article III courts are the only courts with judicial power. Accordingly, the decisions of regulatory agencies remain subject to review by Article III courts. However, cases not requiring "judicial determination" may come before Article I courts. Article I judges are not subject to the Article III protections. For example, judges from Article I courts do not enjoy life tenure, and their salaries may be reduced by Congress. The existence of Article I courts has been controversial, and their power has been challenged before the US Supreme Court. The US Supreme Court has determined that Article I courts may exist, but that their power and their decisions are subject to ultimate review in an Article III courts. Examples of Article I Courts are:

- United States Tax Court
- United States bankruptcy courts
- Board of Patent Appeals and Interferences
- Trademark Trial and Appeal Board

The bankruptcy and admiralty jurisdictions serve different objectives, and the compliance of the two jurisdictions has generated a significant amount of confusion. The confusion centers on whether a bankruptcy court (Article I court) can administer the debtor's maritime assets.

In *Murray's Lessee v. Hoboken Land & Improvement Co*. (1856), the US Supreme Court ruled that cases involving admiralty inherently involves judicial determination, and must come before Article III courts. Other cases, such as bankruptcy cases, have been held not to involve judicial determination, and may therefore go before Article I courts. Then, whether a bankruptcy court (Article I court) can have judicial power to determine the maritime assets of a shipping company when it was filed for bankruptcy?

One line of cases held that federal bankruptcy courts have no jurisdiction to determine the validity of maritime liens or its enforcement against a debtor's ship. For example, in *Taylor v. Carryl*,[1] the US Supreme Court opined that the maritime lien for seamen wages is prior to all other claims on the vessel, and must be first paid. By the US Constitution, the only court that has jurisdiction over maritime lien, or authorized to enforce it, is the court of admiralty. It is the duty of the admiralty court to, and no court of common law can, enforce the maritime lien.

In *Moran v Sturges*,[2] the US Supreme Court held that admiralty court possesses exclusive power to enforce and execute maritime liens for in rem proceedings. In *Moran*, a vessel was attached by process from a court of common law. The US Supreme Court opined that the only interest this process could seize was a subordinate interest, which subject to the superior claims for seamen's wages. The court of common law could not know what the amount of those claims will be. The nature of the maritime claims must first be heard and decided in the court of admiralty.'[3]

In *re Interocean Trans. Co*.,[4] the federal district court decided that if a creditor used admiralty process to attach the assets of a shipping company, and such shipping company had already filed its bankruptcy petition before the admiralty attachment process, the creditor divested bankruptcy court of jurisdiction over the assets. Accordingly, it seems that the only US court that may sell a vessel free of maritime liens is the admiralty court.

Yet other line of cases upheld the jurisdiction of a bankruptcy court in administering maritime assets. This line of cases showed that when an individual lien holder submits himself voluntarily to the equitable jurisdiction of another court, his right to enforce the maritime lien would be extinguished. For example, in *Hudson v New York & Albany Transportation Co*.,[5] the issue was whether a federal district court, in administering an equity receivership, has the power to sell a vessel free of maritime liens. The federal district court held that a receivership court could adjudicate maritime liens if the maritime lien holder voluntarily submits to the court's jurisdiction. In *Hudson,* the Circuit Court of Appeals affirmed the general principle in *Moran v Sturges* by saying that: “It is undoubtedly true that proceedings against vessels in rem to enforce maritime liens are vested exclusively in the District Courts of the United States… and there are many other authorities… all holding that no court other than the admiralty court can exercise jurisdiction over maritime liens or divest or extinguish them.”[6]

---

[1] 20 How. 583 (1857).
[2] 154 U.S. 256 (1894), at pp. 277-78.
[3] Id., at p. 278.
[4] 232 F. 408 (1916), at p. 410.
[5] 180 F. 973 (1910), at p. 978.
[6] Id, at p. 975.

Then the Circuit Court of Appeals discussed the scenario of "consenting creditor"; if a maritime lien holder consents to a sale free of lien, a court of equity will have the right to make such decree. However, the appearance of a maritime lien holder in court to prove the amount of his claims cannot be construed as a consent.[7]

In *James Rees & Sons Co. v Pittsburgh & Cincinnati Packet Line*,[8] the court opined that by virtue of being a plaintiff in a receivership proceeding, the maritime lien holders who consented to the sale of boats, the purchasers would acquire the boats free of liens and encumbrances.

In re *Millenium Seacarriers, Inc.,*[9] a US Court of Appeals case decided in 2005, held that the maritime lien holders had consented to the sale, free of their liens, by appearing in the equity proceeding and placing the lien before the court for adjudication. The US Court of Appeals opined that when a maritime lien holder places his maritime lien claim for adjudication before the bankruptcy court by (1) filing his notice of objection, (2) remaining in the action, and by (3) litigating his lien actively through the adversary proceeding, the maritime lien holder assents to the equitable adjudication of the bankruptcy court under principles of admiralty law.[10]

In re *Millenium Seacarriers*, the maritime lien holder contends that the bankruptcy court could enjoin him from seeking subsequent attachment and enforcement proceedings in foreign admiralty courts, but the bankruptcy court could not extinguish his rights wholly. The US Court of Appeals rejects this contention by holding that maritime lien holder was not allowed to enforce his maritime lien in the foreign admiralty court; the maritime lien itself has been extinguished as a matter of admiralty law.

In order to harmonize the two lines of cases, the US Supreme Court has traditionally been adopted a judicial approach that if the bankruptcy petition was filed before the admiralty action, the bankruptcy court will obtain its jurisdiction over the debtor's vessel and all claims related to it. For example, in *The Philomena*,[11] the federal district court held that if the vessel had been seized under the admiralty process, before the bankruptcy proceeding begun, the admiralty court shall not surrender the vessel to the bankruptcy court.

In *The Bethulia*,[12] after the institution of the bankruptcy proceedings, but before the adjudication, the admiralty court took possession of the vessel. The receiver made a petition in bankruptcy court in seeking the proceeds of the sale of the vessel. The federal district court denied the receiver's petition in bankruptcy court, and decided that the sale of the vessel shall be proceeded with the admiralty case.

In *Casco*,[13] a maritime lien holder had rendered salvage services to dredge Casco. The dredging company which owned the dredge Casco was being instituted for involuntary proceedings in bankruptcy. The bankruptcy court appointed a receiver, and he at once took possession of the dredge. One month later, the maritime lien holder filed an in rem action against the dredge, which was in the custody of the receiver. The maritime lien holder made a motion to seek the dredge be arrested by the marshal, so that the salvage claim could be tried out in the admiralty court. The receiver opposed the motion. The federal district court did not let the case proceed in admiralty because all actions against the dredge had been previously stayed. *Casco* represents a scenario where the admiralty court is urged to seize vessels in the custody of bankruptcy court.

---

[7] Id.
[8] 237 F. 555 (1916).
[9] 419 F.3d 83 (2005).
[10] Id., at p. 103
[11] 200 F. 859 (1911)
[12] Id.
[13] 230 F. 929 (1916).

However, the US courts did not adopt such a judicial approach consistently. For example, in re *Waldeck-Deal Dredging Co*.,[14] the Circuit Court of Appeals made its decision without regarding to the time of the bankruptcy petition and admiralty action. In *Waldeck-Deal,* a Florida dredging corporation contracted with the US federal government to dredge a section of the Waterway. The work was carried out in North Carolina. Subsequently, an involuntary petition in bankruptcy was filed against the dredging corporation in Florida, and receiver was appointed. The receiver took possession the dredging equipments in North Carolina. Then seamen and supplies providers of the dredge made claims for maritime liens for the wages due and supplies furnished to the dredge. The seamen and supplies providers seek to enforce the maritime liens against the dredge in the possession of the bankruptcy receiver. The trustee moved to dismiss on the ground that Florida court had exclusive jurisdiction on the liens claimed. The court denied the trustee's motion, and the trustee appealed. The Circuit Court of Appeals held that a bankruptcy court could adjudicate maritime liens regardless of when the bankruptcy case was brought.

Even the courts have acknowledged that the law on whether a bankruptcy court could adjudicate maritime lien is unsettled. In *Empire Stevedoring Co. v. Oceanic Adjusters, Ltd*.,[15] the US government was a shipper of goods on a stricken vessel. After delivered aid to the stricken vessel, the stevedoring company sued for recovery of general average contributions from government. The federal district court decided that the claim for stevedoring services representing valid general average expense. However, it did not have lien on general average fund, even company's claim was listed on general average statement. In regarding the jurisdiction on bankruptcy, the federal court pointed out that a bankruptcy court's power to adjudicate maritime liens is unsettled.

Therefore, a bankruptcy court's power to adjudicate maritime liens remained unsettled even during the 70s, over 76 years after the principle established in *Moran v Sturges*. For example, when J. Landers published his article *The Shipowner Becomes a Bankrupt* in 1972, he indicated that US courts are unsettled on whether a bankruptcy court is empowered to extinguish maritime liens by selling a vessel free and clear of liens and interests.[16]

2. The Doctrine of Custodia Legis

Before the Bankruptcy Reform Act of 1978 (the "1978 Bankruptcy Code"), courts employed the custodia legis doctrine to uphold the exclusive jurisdiction of an admiralty court to administer a vessel when the bankruptcy petition was filed after the commencement of an in rem admiralty proceeding. For example in *Wong Shing v. MV Mardina Trader*,[17] a Hong Kong registered vessel named the Mardina Trader was arrested and seized in the Canal Zone pursuant to an action in rem filed by various crew members for wages. On the same date of arrest, Wong Shing and other seamen filed a complaint in rem in US District Court against the vessel and against the vessel owner Mardina Trader Ltd (a Hong Kong corporation). Thereafter, a judgment was obtained and the vessel was ordered to be sold. Mardina Lines (a Panama corporation) owned 100% of the Mardina Trader Ltd. Subsequently, a trustee was appointed for the the benefit of Mardina Lines' creditors. The trustee immediately obtained a temporary restraining order to postpone the sale of the vessel. The District Judge ordered the judicial sale to proceed and directed the US Marshal to disregard the temporary restraining order. The vessel was sold to a resident of the Republic of Panama. The trustee made an objection to confirm the sale. In appeal, the Court of Appeals held that the federal district court has no jurisdiction over vessel, so that it could not issue the temporary restraining order. The Court of Appeals affirmed an admiralty court's jurisdiction in selling the maritime assets on the ground of custodia legis.

---

[14] 45 F.2d 951 (1930) at pp. 952-53.
[15] 315 F. Supp. 921 (1970) at p. 925.
[16] J. Landers, *The Shipowner Becomes a Bankrupt*, 39 U. Chi. L. Rev 490 (1972) at pp. 506-07.
[17] 564 F.2d 1183 (1977), at p. 1188.

However, after the enactment of the 1978 Bankruptcy Code, if shipping company files a bankruptcy petition before the admiralty proceeding, section 362 will preclude the application of custodia legis.

3. The Bankruptcy Code

Even after 84 years of the *Moran v Sturges* case, the 1978 Bankruptcy Code still didn't clearly address the issue of whether bankruptcy courts possess jurisdiction over admiralty actions.

The 1978 Bankruptcy Code completely altered US bankruptcy law. It created a codified law on bankruptcy (Title 11 of the United States Code), and created bankruptcy courts which served as adjuncts to the US federal district courts. Under the previous law, the Bankruptcy Act of 1898, the federal district courts served as bankruptcy courts and appointed "referees" to conduct proceedings, so long as the district court chose not to withdraw a case from the referee.

The 1978 Bankruptcy Code eliminated the "referee" system. It allowed the US President to appoint bankruptcy judges for terms of 14 years (as opposed to the life tenure given to Article III judges), with the advice and consent of the Senate. Judges from bankruptcy courts could be removed by the judicial council of the circuit on grounds of incompetence, misconduct, neglect of duty, or physical or mental disability (as compared with Article III judges, who may only be impeached by Congress and are constitutionally forbidden from having their pay decreased while in office). Unlike Article III judges, their salaries were set by statute and subject to adjustment.

The 1978 Bankruptcy Code granted the bankruptcy courts jurisdiction over all "civil proceedings arising under Title 11 or arising in or related to cases under Title 11".

An important case for testing the validity of the 1978 Bankruptcy Code is the *Marathon* case.[18] After Northern Pipeline (Northern) filed a petition for reorganization under Chapter 11 under the 1978 Bankruptcy Code, it brought suit in the bankruptcy court against Marathon Pipe Line Co. (Marathon) for breach of contract. Marathon made a motion to dismiss the suit on the grounds that the 1978 Bankruptcy Act unconstitutionally conferred Article III powers on judges who lacked the career protections of Article III judges. The bankruptcy judge denied Marathon's motion. Marathon made an appeal to US federal district court, the district court agreed with Marathon's argument that the law was unconstitutional. Then the case moved to US Supreme Court, which held that Article III jurisdiction could not be conferred on non-Article III courts (i.e. courts without the independence and protection given to Article III judges).

The US Supreme Court stayed its judgment until October 4, 1982, in order to give the US Congress an opportunity to repair the constitutional flaws in the bankruptcy system. Congress dealt with the problem with the Bankruptcy Amendments and Federal Judgeship Act of 1984. This statute authorized the federal district courts to refer bankruptcy cases to the bankruptcy courts, but the bankruptcy court must submit proposed findings of fact and conclusions of law to the district court for de novo review.

An important case for bankruptcy for a shipping company in the context of 1978 Bankruptcy Code is the *Azioni* case,[19] which decided in 1983, just after the 1982 decision of *Marathon*. The *Azioni* case reviewed once again the issue of whether a bankruptcy court can constitutionally decide admiralty question, but in light of *Marathon*.

In *Azioni*, maritime lien holders filed an *in rem* attachment on a vessel named *Sorrento* in March, 1982, and a bankruptcy petition was filed in May, 1982.[20] The court applied the doctrine of *custodia legis* and demanded that all of the claims against the vessel *Sorrento* be heard and decided by the

[18] *Northern Pipeline Co. v. Marathon Pipe Line Co* ., 458 U.S. 50 (1982).
[19] *Ciel Y Cia S.A. v. Nereide Societa Di Navigazione Per Azioni*, 1983 A.M.C. 1192 (1983).
[20] 1983 A.M.C. 1192 (1983), at p. 1196.

federal district court sitting in admiralty. The court reasoned that the ship in question was within the jurisdiction of the district court at the time of the filing of the Bankruptcy petition. As such, the vessel, and the claims against it, should not have been before the Bankruptcy Court. In terms of procedure, the Marshal attached the *Sorrento* prior to the filing of the Bankruptcy petition, the federal district court sitting in admiralty would be the proper forum to dispose of all claims against the vessel. In summary, the court finds that after the expiration of *Marathon's* stay, the bankruptcy court can no longer constitutionally exercise jurisdiction over an admiralty case, and the debts related to the vessel *Sorrento* were predominantly admiralty in nature; since admiralty questions dominate the issues to be decided, it follows that the bankruptcy court does not have jurisdiction to decide the case.[21]

The *Azioni* decision is consistent with prior case law, which provides that "When a ship has been seized by the Marshal under *in rem* process before the filing of a petition in bankruptcy, the ship does not come into the control of the Bankruptcy Court. The action cannot, therefore, be enjoined and will proceed to final adjudication and a sale of the ship unless the Bankruptcy Trustee has procured its release under bond."[22]

Although section 362 of the 1978 Bankruptcy Code provides that the filing of a bankruptcy petition stays the commencement of an in rem proceeding to foreclose a ship mortgage.[23] Section 362 applies to any collection action against freights of a vessel. The section also applies to cargo in which the shipowner who possesses a lien for unpaid freight. The “determinations of the validity, extent, or priority of liens” are core matters of the bankruptcy court.[24] Judge Friendly observed in *re Penn Central Corp*.,[25]that “There appears to be no doubt that a bankruptcy court can be constitutionally vested with power to resort to its judgment in determining what constitutes satisfaction of the claims of creditors.” And F.R. Kennedy showed in his published article *Jurisdictional Problems Between Admiralty and Bankruptcy Courts*[26] that numerous precedents indicate the power of the bankruptcy courts in determining the validity of maritime liens.[27]

However, in *United States v. ZP Chandon*,[28] the federal district court held that that section 362 does not stay an action brought by a preferred maritime lien holder who enforces the claim for seaman's wages, which arise after the shipowner files a bankruptcy petition. The court reasoned that the US Congress did not intend the prohibition mentioned in section 362(a)(4) as against enforcing liens on estate property to include maritime liens.

In *re McLean Industries*[29], U.S. Lines (the debtor) commenced its Chapter 11 case based on the “1978 Bankruptcy Code”. The estate includes twelve exceptionally large vessels called “Econships”, which were designed to transport a large volume of goods packed in containers. The Econships were constructed under § 615 of the Merchant Marine Act of 1936, and the vessels, according to US Maritime Administration (“MarAd”), constitute 23% of the US flag commercial fleet capacity to carry containers and some 15% of the militarily useful deadweight tonnage of the privately owned US general cargo fleet. According to “MarAd”, the Econships are the largest, most fuel efficient, modern and competitive container ships under the U.S. flag. MarAd has declared these vessels to be essential for the national security interests because of their large sealift capacity. Two events happened after the commencement of the Chapter 11 proceedings: (1) secured creditors sought unconditional relief from automatic stay; and (2) foreign creditors commenced arrest proceedings against four United States flag vessels belonging to debtor. The New York Bankruptcy Court held that: (1) the US

[21] Id., at p. 1197.
[22] Gilmore & Black, The Law of Admiralty, (2nd Ed. 1975), at p. 807.
[23] 11 U.S.C.A. § 362 Automatic stay.
[24] 28 U.S.C. § 157(b)(2)(K).
[25] 384 F.Supp. 895 (1974), at p. 950.
[26] 59 Tulane L. Rev. 1183 (1985).
[27] Id, at pp.1199-1201.
[28] 889 F.2d 233 (1989), at p. 238.
[29] 76 B.R. 328 (1987), at p. 332.

Shipping Act prohibits transfer of any US flag vessel to noncitizen without the approval from the US Secretary of Transportation, and the bankruptcy court would give such prohibition an extraterritorial effect. This case shows that bankruptcy court has jurisdiction to adjudicate the validity of maritime liens or ship mortgages.

4. Policy in Favor of Vessel Mobility

Admiralty law has a strong policy of maintaining vessel mobility in liquidation scenario. Chapter 7 of the Bankruptcy Code is in consistent with such liquidation objectives of vessel mobility. When a shipowner defaults on its obligations to maritime lien holders or a ship mortgagee, this policy is fostered through sale of the vessel free and clear of such interests.

Given the similar liquidation objectives of admiralty and chapter 7 of the Bankruptcy Code, courts have consistently held that an admiralty court's jurisdiction over a debtor's vessel remains unaffected by the subsequent filing of a chapter 7 petition.

For example, in *Morgan Guaranty Trust Co. of NY v. Hellenic Lines*[30], US marshal in New York arrested the freight collected by Liner Company. The first portion of the freight was arrested before Liner Company filed Chapter 11 reorganization; the second portion was arrested after the Chapter 11 filing. US terminal operator of Virginia applied an order for the payment of the freight. At this point, the court believed that the rehabilitative goal of a reorganization proceeding, and the need for supervising the debtor's assets by a single court, would outweigh any competing admiralty concerns.[31] In other words, once a debtor files a petition for Chapter 11 reorganization, the doctrine of custodia legis would not be applicable.

Subsequently, the creditors of Liner Company voted to move the bankruptcy court for an order converting the Chapter 11 reorganization to a Chapter 7 liquidation. The bankruptcy judge then granted the motion and ordered a Chapter 7 liquidation. At this point of development, the court noticed that, in the absence of a reorganization proceeding, custodia legis is the appropriate rule. The court explained that: "If admiralty jurisdiction is based on an in rem action, it is painfully simple to tell whether a vessel will be administered in admiralty or bankruptcy. The first court to obtain jurisdiction over the assets administers it. Thus, if the marshal, pursuant to admiralty process, has attached the vessel first, the admiralty court administers the asset. If the bankruptcy petition is filed before the marshal reaches the vessel, the bankruptcy court administers the asset."[32]

On the other hand, when a liquidation is contemplated, the principle of comity permits the court that first obtained jurisdiction over an asset to supervise its liquidation.[33] After the case converted from Chapter 11 reorganization to Chapter 7 liquidation, the bankruptcy court would apply custodia legis, and Liner Company should pay all freights collected into the court registry.

The survival of the maritime lien depends on an adequate identification of the freights collected, i.e., a tracing of the freights back to the ships that earned them. An accurate voyage accounting of the freights within Liner Company's custody and control, which were subject to arrest in the admiralty in rem proceeding, is necessary for the protection of other maritime lien claimants.

5. Conclusion

After a lengthy academic discussion on some of the leading bankruptcy cases over various ways on admiralty assets treatments, I would like to present the conclusion by using a story style, so that the readers may get some general principles for practical application.

---

[30] 585 F. Supp. 1227 (1984).
[31] Id., at p. 1228.
[32] Id., at p. 1229.
[33] G. Gilmore & C. Black, The Law of Admiralty, § 9-92, at 807-08 (2d ed. 1975).

Creditor comes to see his lawyer and asks: "I have a maritime lien on a US fishing boat and I am thinking about trying to collect against the boat. What if the owner of the fishing boat files bankruptcy? Does that keep me from getting paid?"

Lawyer explains: "If you have a maritime lien against a vessel, you may file an admiralty action against the vessel in rem in the US District Court to enforce the lien. In the normal situation, neither the state courts nor the bankruptcy courts have jurisdiction to hear such a lien claim."

Lawyer further explains: "If the debtor (the owner of the fishing boat) is insolvent, he may file a petition in the US Bankruptcy Court for protection of himself and his property from any debt collection efforts by the creditors. In the normal situation, neither the state courts nor the US District Courts can proceed with actions against the bankrupt after filing of the bankruptcy petition."

Then Lawyer comes discuss about the relevant legal theories: "If there is a potential for conflict between the admiralty jurisdiction and the bankruptcy jurisdiction. The rule developed by the courts to resolve this conflict is known as custodia legis. Under the custodia legis rule, the court that first secures control of the vessel is the court which administers the vessel. Therefore, if a creditor files his admiralty action in the US district court, after the Judge issues a ship arrest warrant, the US Marshal will serve the process upon the vessel, if the creditor done all these steps before the debtor files a petition in the bankruptcy court, then the US district court will continue to administer the vessel under the admiralty jurisdiction."

Lawyer then gives an alternative scenario: "If the debtor files a petition in the bankruptcy court before in rem process is served on the vessel, then §362(a) of the 1978 Bankruptcy Code will automatically stop any further action to advance claims against the debtor and his vessel, and §362 will deprive all other courts of jurisdiction."

At the end of the day, Lawyer strongly reminds Creditor be prepared to "file early, and file often", and the creditor's rights may change if his debtor files his petition in bankruptcy court.

6. Acknowledgment

This research was supported under **Project Code: G-SAA8** by The Hong Kong Polytechnic University.

An activity based costing model for liner shipping pricing management

Shin-Chan Ting and Chung-Jen Yen

Department of Transportation and Navigation Sciences, National Taiwan Ocean University, Taiwan.

Abstract

Liner carriers have been conducting the incessant drive to cut costs through the deployment of larger ships. These post-Panamax ships have been deployed to east-west main trade routes, and many of similar type ships are under construction and delivered in a couple of years. The liner shipping industry with over capacity and lower price elasticity of demand is highly competitive with freight rates fluctuating wildly. In this competitive market, liner carriers could utilize revenue management systems to increase profits by using effective slot allocation and pricing. For the first step of pricing management, the variable costs of all origin-destination pairs and service combinations need to be accurately calculated to ensure the marginal contribution of each service. In this paper, a pricing decision support system is proposed and the all the variable cost items are included in the cost database. We applied the concept of bill of material (BOM) to build up the structure of service costs and an activity based costing model is proposed to identify the cost driven trigger of each variable cost item. Through this model, the variable cost of each shipment service can be calculated easily and accurately to support pricing decision making. We illustrate this model and database system with a case study of a Taiwan liner shipping company and the results show the applicability and better performances than the previous pricing decision used in practice.

Keywords: Liner shipping, Revenue management, Activity based costing (ABC), Pricing

1. Introduction

Since liner shipping is a capital-intensive industry, the liner companies must invest large sums on vessels and containers. With the current fiercely competitive market, freight rates cannot be increased, and it is costly to reposition empty containers due to trade imbalances. Liner companies have difficulty generating reasonable profits and even run deficits (Ting and Tzeng, 2004). The business of this industry has been being focused on cost-reduction, which in turn depends upon generating supply. Increasing the vessel capacity supply helps liner carriers' lower ceilings by forcing down per-unit costs. As shown in Figure 1, the post-Panamax containerships have been deployed to east-west main trade routes, and many of similar type ships are under construction and delivered in a couple of years, and it will increase the capacity rapidly. The problem is that for attracting more cargo to fill up these giant vessels, many trades are plagued with tremendous overcapacity, fierce competition and low rates. The result is a vicious circle: cutting costs; increasing space supply; building bigger ships; creating overcapacity; competing by reducing freight rates; suffering from low rates; cutting costs (Ting and Tzeng, 2004).

As for the demand, American Bureau of Shipping (ABS, 2007) is pessimistic to the development of the shipping industries. The projected annual trade volume growth of container trade is on the downside as shown in Figure 2. Additionally, rising delinquency rate for subprime loans has caused a sharply downfall in the value of related mortgage-based securities, consequent global stock markets correction, and the tighter financial conditions. Subprime mortgage will result in reducing the GDP, consumer expenditure and increasing more unemployed persons. Economist Intelligence Unit (2008) predicts that global GDP growth will be just 0.9% in 2009 and it will cause a large amount of reduction in the trade volume. The above factors will make demand-supply imbalance in the liner shipping market, and the demand of trade volume will fall short far from the supply of capacity in the future.

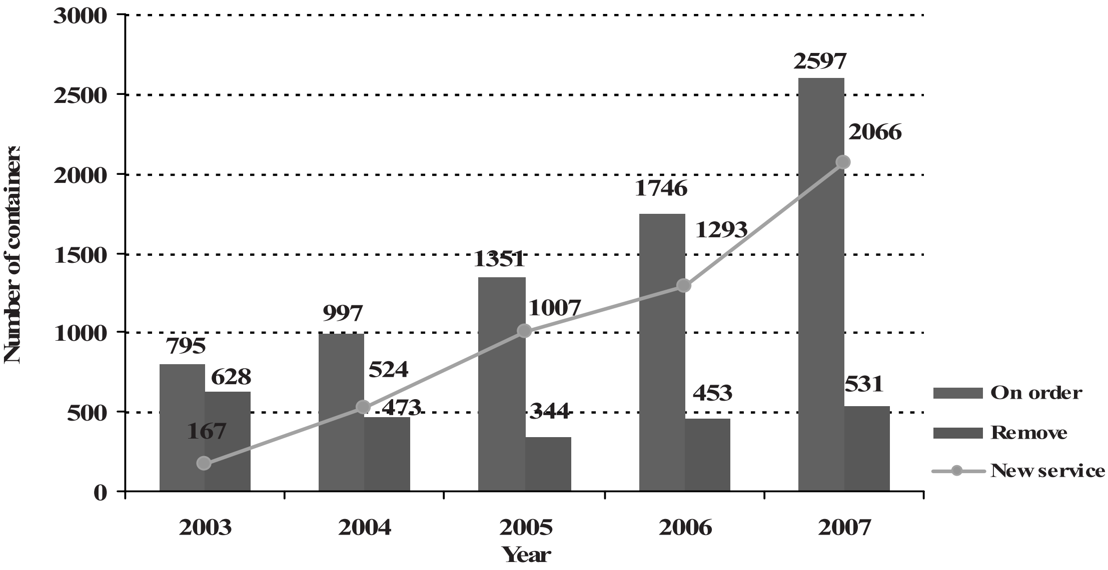


Figure 1: Actual Deliveries Run Out of Current Order Book
Source: ABS (2007)

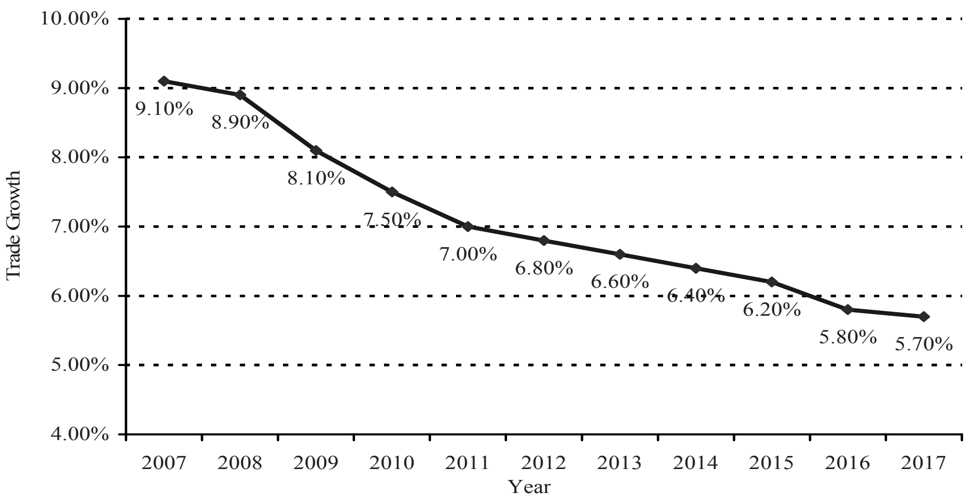


Figure 2: Projected Annual Trade Volume Growth
Source: ABS (2007)

On the other hand of operational costs, according to the shipping digest ROC (2007), the cost in salary and insurance premium will expand a rapid increasing rate from 2007 to 2011. The crew salary will be raised by 12% per annum growth and the insurance premium will increase by 12% per annum growth. In addition, the fuel price which is about 15% of total costs has fluctuated acutely on the last months, so it is difficult to control the total cost for shipping companies.

Due to low price elasticity of demand and low customer switching costs, overcapacity has been leading to price war. Agents, persons in charge of pricing and sales representatives lower the prices on the spot market, and to attract needed cargo tonnage. Many liner companies focus on short-term performance improvement by trying to control load factors. An increase in capacity utilization is usually viewed as a remedy for declining yields. A downward spiral of lower and lower yields is triggered by lowering prices to generate more demand. Clearly, pricing and revenue are directly linked: revenue equals price * lifts, which means that price determines revenue. Assuming that we are acting in a very simple market model, there are principally two ways to react in the market: either we change the price and cope with the reaction in terms of more or less demand by adjusting the capacity availability; or we influence the capacity availability and have to assess the necessary reaction in terms of prices. Most carriers simply use the low-rate policy to assure space utilization. This resulted in the space supply increase and lower rates (Ting and Tzeng, 2004).

From a practical point of view, the sales process is that salespersons get the reserve price (i.e. the lowest price ceiling) offered by pricing department of head quarter or branch offices to seduce cargo. Under the pressure of demand-supply imbalance, salespersons sell the space most likely with the reserve price. Although this price is not less than the reserve price that carriers can accept, the space could be sold with the higher price to gain more revenue. This possible revenue loss is the consumer surplus shown as in Figure 3(a). On the contrary, differential pricing is to offer a lower price to customers with a low willingness to pay and a higher price to those with a high willingness to pay (Phillips, 2005). Compared with the general pricing method, it does not only make more revenue and profit for carriers but reduce the customer surplus as in Figure 3(b), which can achieve much more economic effectiveness in the market.

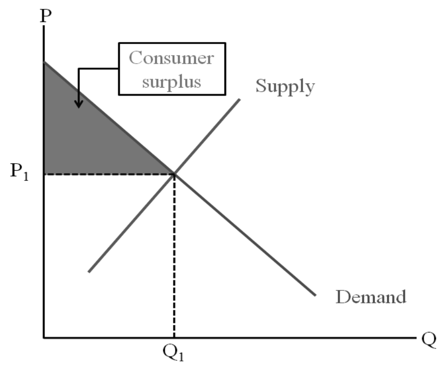


Figure 3: (a) Uniform Pricing

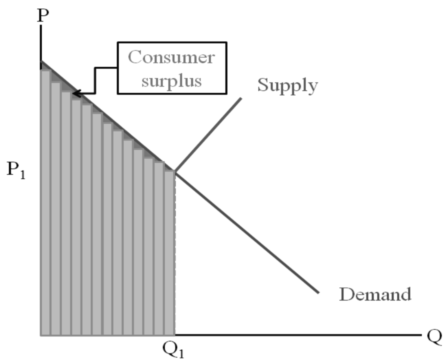


Figure 3: (b) Differential Pricing

In this competitive market, liner carriers could utilize revenue management systems to increase profits by using effective slot allocation and pricing. For the first step of pricing management, the variable costs of all origin-destination pairs and service combinations need to be accurately calculated to ensure the marginal contribution of each service. In this paper, a pricing decision support system is proposed and the all the variable cost items are included in the cost database. We applied the concept of bill of material (BOM) to build up the structure of service costs and an activity based costing model is proposed to identify the cost driven trigger of each variable cost item. Through this model, the variable cost of each shipment service can be calculated easily and accurately to support pricing decision making. The rest of this paper proceeds as follows. Section 2 reviews some related research on pricing in liner shipping and activity based costing (ABC) model. Section 3 clarifies cost items.

Section 4 develops an ABC model for liner shipping pricing and followed by concluding remarks in Section 5.

2. Related Studies

Ting and Tzeng (2004) proposed a conceptual liner shipping revenue management (LSRM) model. The LSRM is concerned with the integrated operation of long-term customer management, cost management, route planning and ship scheduling, as well as short-term cargo demand forecasting, container inventory control, slot allocation, pricing and dynamic space control. There are two major components: (a) long-term planning, which can assist with longer term customer management, cost management, market monitoring, service route planning and ship scheduling; and (b) short-term operations, which can assist with voyage revenue optimization in terms of demand forecasting, slot allocation, pricing, container inventory control and dynamic space control. Ideally, such a system should be integrated with freight revenue, cost, container inventory database and accounting systems.

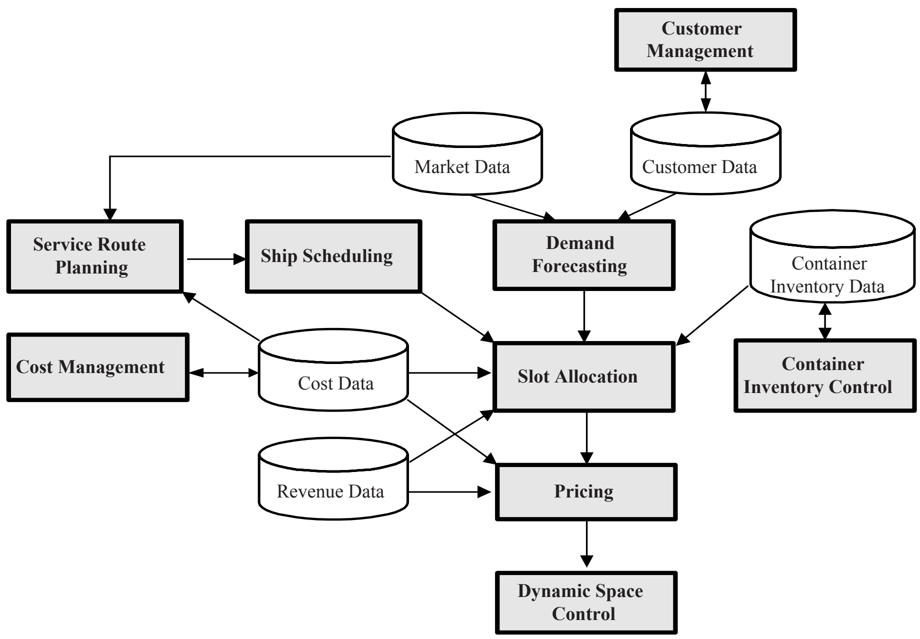


Figure 4: Conceptual Model for Liner Shipping Revenue Management System
Source: Ting and Tzeng (2004)

As for cost management, there must be a powerful database recording every item of costs including fixed and variable costs. Variable costs, in particular, should be tracked with detailed records of every shipment including truck, feeder and railway costs, container handling costs, terminal and depot stowage costs, commission, tally costs and cargo claim costs. The variable costs of all service point pairs are needed to accurately calculate the freight's marginal contribution. For pricing, based on information regarding costs, local market sales, demand pattern (e.g. distribution, time, volume, delivery condition), this subsystem provides tactical pricing decision support to make the space sell at the right price, to the right customer and at the right loading port, as well as to maximize the contribution and utilization of the vessel capacity.

Cooper (1988) initiated the methodology of activity based costing (ABC), which is a costing model that identifies the various activities in an organization and assigns the cost of each activity resource to all products and services according to the actual consumption by each. A detailed definition of the activity based costing was defined by Themido et al. (2000): ABC began as an alternative to traditional methods in the early 1980s when many companies started to realize the adverse consequences of accounting systems that could generate incorrect costing. The main reasons for this happening were the changes experienced in organizations' cost structures. Overheads and indirect costs increased and often became more important than direct costs. ABC attempts to identify all the various activities needed to provide a product or service and allocate costs accordingly, and it is shown as Figure 5.

Emblemsvåg (2007) linked ABC to economic profit to grow the bottom-line. The main effect of the economic profit analysis in this particular case is to make top management aware of the fact that the company has to make structural changes to survive in the long run, and he found that ABC and EP together are really powerful eye-openers and primers for change. Possibly the greatest benefit of ABC and EP is that management can act on facts.

To analyze the effects of ABC on traditional operation model in inventory system, Ping (2008) developed a new inventory model based on activity based model. In the new model, based on ABC perspectives he analyzed the cost drivers other than conventional ones specified in the inventory system, and then revise the inventory cost functions by discussing the allocation of replenishment cost, inventory carrying cost and shortage cost according to these cost drivers. Compared with traditional model, ABC model can reduce the cost in this example. Company can obtain more accurate cost information from ABC, and it also can help managers make right decisions.

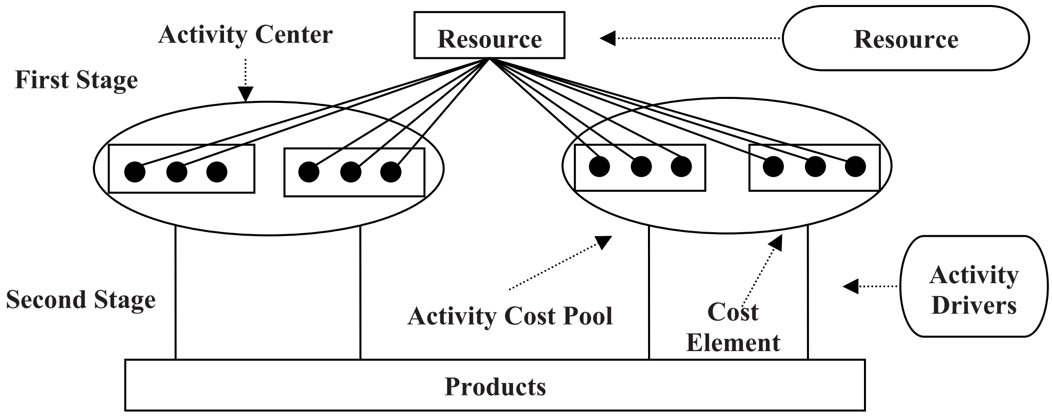


Figure 5: The Activity Based Costing (ABC) Model

Source: Turney (1992)

3. Cost Restructure for Liner Shipping Pricing

The costs in the liner shipping can be categorized into fixed costs and variable costs. Fixed costs are not relevant to the quantity of containers transported on the voyage, but variable costs vary with the quantity and required services of containers transported. Parts of variable costs could happen while the customers require the additional services, and these can be deemed as the additional variable costs. Carriers could collect surcharges for providing the additional services to balance the additional variable costs. In this section, we list the costs in the different types as follows:

3.1. Fixed Costs

Fixed costs include crew salaries and bonuses, stores, vessel insurance premium, vessel maintenance charges, port charges, bunker, administration and management fees, and so forth. Since marginal contribution (i.e. revenue minus variable cost) of each shipment is the major concern to pricing, and this ABC model is proposed to calculate the marginal costs for pricing. Thus the fixed costs can be left out of this model.

3.2. Variable Costs

Variable costs are directly related to the volume of freight, which includes several major items: feeder costs, trailer/railway costs, container handling costs, tally costs, and terminal stowage costs.

3.3. Variable Costs of Additional Services

Variable costs of additional services include the items relevant to specific services or occasions by collecting re-handling surcharge, BAF (bunker adjustment factor), CAF (currency adjustment factor), Suez Canal surcharge, Panama Canal surcharge, over-gauge surcharge, peak season surcharges, war risk surcharge, IMCO (Inter-government Maritime Consultation Organization) surcharge, optional port surcharge, port congestion surcharge, transshipment surcharge and so forth.

3.4. Empty Container Repositioning Costs

Empty container repositioning costs are mainly due to the trade imbalance and the demand of container is different around the world, especially in the transpacific and Asia/Europe trades as shown in Figure 6. There is no doubt that empty container repositioning is the biggest problem for carrier; they need to spend a lot of money and time transporting empty containers from the surplus area to the demanding area; hence, it could result in some opportunity costs for container carrying, handling and storage.

On the other hand, under the pressure of demand-supply imbalance, liner shipping business is more and more difficult, so salespersons want to seduce the cargo wherever the destination port is; this situation may increase empty containers in the surplus area, and then carriers will pay a lot of costs for repositioning, and it may turn the profitable freight into loss. For this reason, empty container repositioning costs must be taken into account.

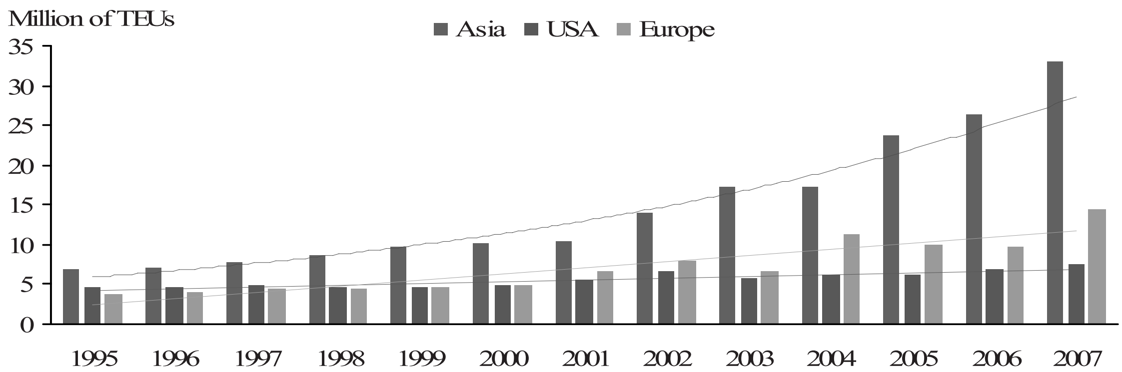


Figure 6: Global Trade Imbalance in the Container Flow

Source: UNCTAD (1997-2008)

4. The ABC Model for Liner Shipping

4.1. The Concept of Categorized Costs in Liner Shipping

We utilize the concept of BOM and ABC to analyze the cost in liner shipping. In the manufacturing industry, engineers use BOM to set up the structure of the product and to calculate the unit cost of producing this product by adding up all the costs of the needed components in the BOM. A BOM contains a listing of all the assemblies, subassemblies, parts, and raw materials that are needed to produce one unit of a finished product. The listing in the BOM is hierarchical in nature with the top level representing the finished product, which may be a sub-assembly or a completed item (Stevenson, 2005). For example, a product structure tree for a bench, shown as Figure 7, the end item is shown at the top of the tree, and it consists of the leg assembly, seat, and back assembly. A BOM presenting the product structure tree is useful to determine the quantities of each ingredient needed to produce a desired number of end items.

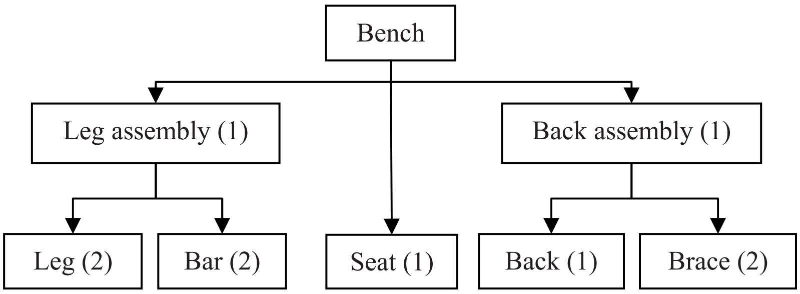


Figure 7: The product structure tree for a bench

We applied the concept of BOM to build up the structure of service costs and an activity based costing model is proposed to identify the cost driven trigger of each variable cost item, shown in Table 1.

Table 1: Cost Items and Costing Triggers

No.	Cost Items	Costing Attributes	Costing Triggers for Pricing
1	Crew costs	Fixed cost	Not to be taken into account
2	Vessel costs	Fixed cost	Not to be taken into account
3	Port charges	Fixed cost	Not to be taken into account
4	Bunker costs	Fixed cost	Not to be taken into account
5	Terminal handling costs	Variable cost	To be taken into account
6	Terminal storage costs	Variable cost	To be taken into account
7	Tally costs	Variable cost	To be taken into account
8	Trailer costs	Variable cost	To be driven by service routing requirement
9	Railway costs	Variable cost	To be driven by service routing requirement
10	Feeder costs	Variable cost	To be driven by service routing requirement
11	Re-handling costs	Re-handling surcharge	To be driven by the requirement that the customer want to change the sequence of discharge
12	Optional port costs	Optional port surcharge,	To be driven by the requirement that the customer want to change the port of discharge
13	Over-gauge handling costs	Over-gauge surcharge	To be driven by the shipment that is over-gauged or needs heavy-lift equipments
14	Transshipment costs	Transshipment surcharge	To be driven by the requirement that the customer want to transship the container
15	Hazardous cargo handling costs	IMCO surcharges	To be driven by the shipment that contains hazardous cargo
16	Suez Canal fees	Suez Canal surcharge	To be driven by routing through Suez Canal
17	Panama Canal fees	Panama Canal surcharge	To be driven by routing through Panama Canal
18	Price increase of bunker	BAF (bunker adjustment factor)	To be driven by increasing price of bunker
19	Currency devaluation costs	CAF (currency adjustment factor)	To be driven by devaluation of US Dollars
20	Port congestion costs	Port congestion surcharge	To be driven by calling at congestion ports
21	War risk costs	War risk surcharge	To be driven by calling at ports in war areas
22	Empty container repositioning costs	Container repositioning surcharge	To be driven by delivering containers at points where there are surplus containers

4.2. Costing Database for Pricing Decision Support

We manipulate SQL server 2005, which is a relational database management system (RDBMS) produced by Microsoft, to program a costing database for pricing decision support. The database is divided into five data sheets including costs of destination, costs of origins, ocean freight, customer information, order Information. The items have been listed in Table 2. The user just enter the information of the order including customer's number, customer's name, the number of customer's category, customer's category, and the distance between customer and container freight station, and then the user can check up the marginal cost of the container by the query code as shown in Figure 7.

Table 2: Costing Database Structure

Costs of Destination	Costs of Origins	Ocean Freight	Customer Information	Order Information
Number of destination	Number of origins	Number of destination	Customer number	Number of order
Name of destination	Name of origins	Name of destination	Customer name	Name of destination
Transportation cost per kilometer	Transportation cost per kilometer	Number of origins	Number of customer category	Name of origins
Stevedoring charge	Stevedoring charge	Name of origins	Customer category	Customer name
Warehousing charge	Terminal handling charge	Number of leg	Distance between customer and CY	Number of containers
	Warehousing charge	Name of leg		Name of leg
		Cost of leg		Free time

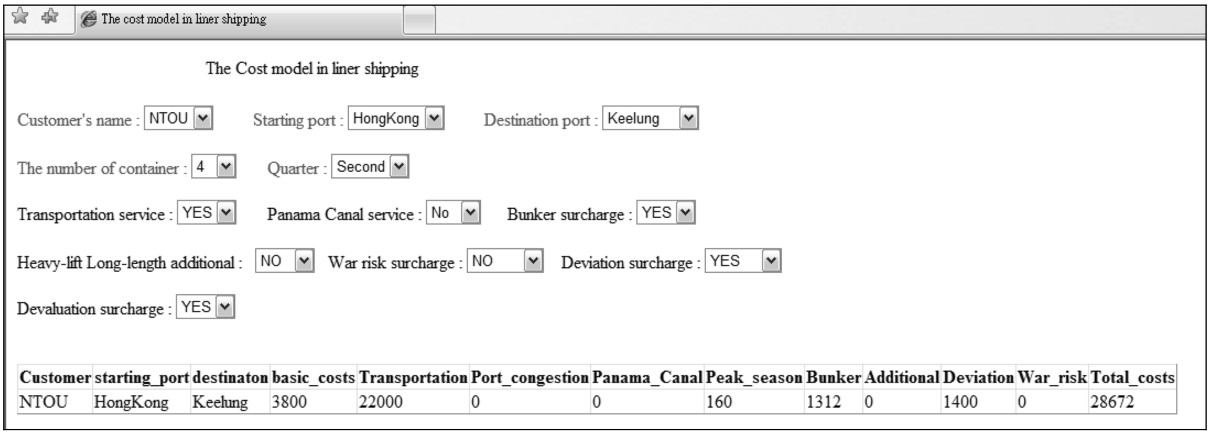


Figure 8: The Interface for Calculating Marginal Costs

5. Conclusions

In this paper, we applied the concept of BOM to build up the structure of service costs, which can help salespersons realize the combination of each service. Through this activity based costing model, the variable cost of each shipment service can be calculated easily. In addition, we also consider the costs of additional services in order to let salespersons offer the reasonable all-included freight rate to customers. Moreover, the costs of empty container repositioning are considered in this model, because this cost becomes a heavy burden to liner shipping companies; we consider that whether the destination port and starting port are the demand area or not, and each port has different weighted coefficient to decide the cost of empty container repositioning. Through including empty container repositioning cost, potential revenue loss can be avoided.

Practically, salespersons do not know the total costs of the shipments, so they may sell the space with the reserve price in order to get the order; perhaps the space can be sold with higher price to make more revenue. In this activity based costing model, it changed the traditional pricing process, because salespersons can check up the marginal costs of each business, and it can assist them to decide the reasonable price. We illustrate this model and database system with a case study of a Taiwan liner shipping company and the results show the applicability and better performances than the previous pricing decision used in practice.

References

American Bureau of Shipping (ABS) (2007), Wiernicki warns of shipbuilding over-supply in 2011.

American Bureau of Shipping (ABS) (2007), Annual review.

Cooper, R. (1988), The rise of activity-based costing - part one: What is an activity-based cost system? , Cost Management (summer): 45-54.

Emblemsvåg, J. (2007), Using activity-based costing and economic profit to grow the bottom-line, Business Strategy Series 8(6): 418-425.

Economist Intelligence Unit (2008), 29 countries to contract in 2009 says Economist Intelligence Unit, 12 December.

Phillips, R.L. (2005), Pricing and Revenue Optimization, Stanford Business Books: New York.

Ping, Y.T. (2008), The effects of activity-based costing on traditional inventory model, The Journal of American Academy of Business 13(2): 143-147.

Shipping digest ROC's website: http://www.shippingdigest.com.tw/news/news1125.htm, last accessed in November 2007.

Stevenson, W.J. (2005), Introduction to Operations Management, McGraw-Hill: Singapore.

Themido, I., Arantes, A., Fernandes, C. and Fuedes, A. (2000), Logistic costs case study- an ABC approach, Journal of the Operational Research Society 51(10):1148-1157.

Ting, S.C. and Tzeng, G.H. (2004), An optimal containership slot allocation for liner shipping revenue management, Maritime Policy and Management 31(3): 199-211.

Turney, P.B. (1992), What an activity based costing model look like, Journal of Cost Management 5: 54-60.

United Nation Conference on Trade and Development (UNCTAD) (1997-2008), Review of Maritime Transport 1997-2008.

Integrated berth and quay crane allocation to container vessels

Zheng Tong and Jiyin Liu

Business School, Loughborough University, Leicestershire LE11 3TU, United Kingdom.

Abstract

We study the integrated problem of berth allocation and quay crane allocation at container terminals with multiple berth areas. The objective is to minimize the maximum delay of vessel services. We formulate this problem into an integer programming model considering all the following practical constraints. Priority scheme, often used by terminal operators to strengthen their relationships with shipping lines, is included to ensure that vessels with priority are served as soon as they arrive. The remaining services carried over from previous planning horizon are considered in the current planning. Because the handling ability of the quay cranes in different berth areas may be different, berth-dependent processing times are considered in the model. Instead of using a pre-determined service time, the model determines the processing time of a vessel based on its workload and the number of quay cranes assigned to it. Preliminary numerical test results are provided to demonstrate the model effectiveness.

Keywords: Container terminal, Berth allocation, Quay crane allocation, Multiple berth areas, Service priority

1. Introduction

Transport by containers is by far the most pervasive and important global shipment methods. Accordingly, as the interface of the global container transportation, container terminals play a decisive role in global container logistics. How to maximize utilization of the limited handling capacity to satisfy the booming demand of container trade has attracted more and more attention from researchers. On the other hand, the increasing number of hub container terminals means more competition in both price and service. Efficient container handling at terminals is important in reducing handling costs and in keeping the shipping schedule.

Berth space is regarded as the most precious resource at container terminals since the expansion of berth space is costly, time-consuming and limited by the natural restrictions, such as water draft and land space. Berth allocation is essential for effective use of this expensive resource. It is also critical for overall efficiency of the whole handling process in the container terminal.

Quite a number of studies have been done on the berth allocation problem (BAP) during past decade. Based on different assumptions regarding the layout of berth area, these studies can be generally classified into two categories. The first category studies the discrete BAP in which the berth area is considered as partitioned sections and each berth section can only serve one vessel at a time. Studies on such problem include Imai et al. (1997), Imai et al. (2001) Nishimura et al. (2001) and Imai et al. (2003). Partitioning the berth area into discrete sections may lead to serious waste of berth resource and is not very common in practice.

Other studies considered a berth to be a continuous space, where vessels can be assigned to certain berth if its physical length is sufficient to accommodate them. This continuous berth allocation often results more effective use of berth space. Such studies include Lim (1998), Guan et al. (2002, 2004), Park and Kim (2002) Kim and Moon (2003) and Imai et al (2005).

Although, great efforts have been made to produce good BAP solutions, there are still several problems remain unsolved. One of the most challenging problems is how to determine the processing time for vessels. In practice, the processing time is determined mainly by two factors, the workload (amount of inbound and outbound containers) and the operation of quay cranes. In most previous research the processing time of a vessel is assumed to be fixed regardless of the number of quay cranes serving it, or to be proportionate to the ship size no matter how many of its containers will be discharged and loaded at this terminal. In some other research, the processing time of a vessel is assumed to be proportional to the travel distance between its berth position and the position of its containers in the stacking yard, and thus is irrelevant to the quay crane operation and the workload. To ensure more accurate estimation of processing times, it is necessary to include quay crane operation planning into BAP, resulting in an integrated berth and quay crane allocation problem.

Very limited work has been done on the integrated berth and quay crane allocation. Kim and Park (2003) developed a two-phase model to decide berth allocation and quay cranes assignment simultaneously. In the first phase, the berth position, berth time and number of quay cranes for each vessel were determined. Then the exact movement of each quay crane was scheduled in phase two. However, instead of identifying the real processing time of each vessel, they assumed that the processing time was inversely proportional to the number of cranes assigned to the vessel, which may not be accurate. Imai et al (2009) modelled the simultaneous berth and quay crane allocation problem and employed genetic algorithm to achieve an approximate solution to the problem. But the model is applicable only to container terminals with partitioned berth space, not those with continuous berth space. Meisel and Bierwirth (2009) addressed the combined problem of berth allocation and quay crane assignment. Their model included several practical constraints, such as the decrease of marginal productivity of additional quay crane assigned to vessels and the increase of processing time due to the deviation between the assigned berth position and the desired position of each vessel.

In addition to integrating berth allocation and quay crane allocation, in this paper we also consider various practical constraints such as multi-berth areas, priority scheme, berth-dependent processing time and remaining services.

Multi-berth areas and berth-dependent processing time

In practice, there are many container terminals consisting of several berth areas. Honghong International Terminals Ltd (HIT) is a typical one (see figure 1). A vessel cannot berth across two berth areas because they are not in a straight line or not connected. BAP models for one straight-line berth configuration may not be suitable to terminals with such multi-berth areas.

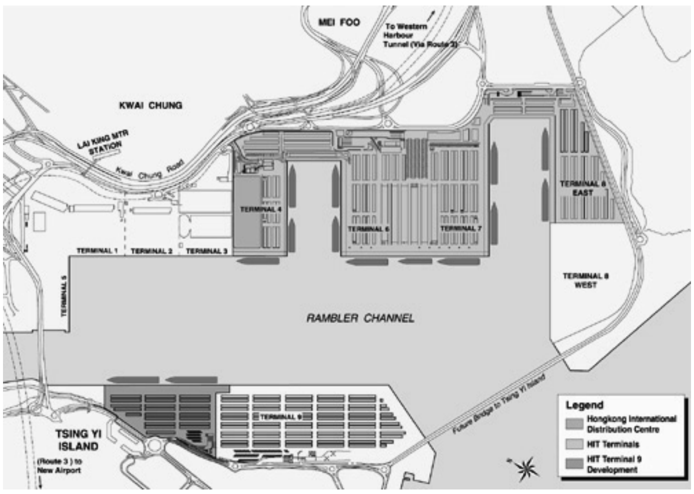


Figure 1: Layout of the Berth Area of HIT

Source: http://www.hit.com.hk/4fac/terminal.asp

At container terminals with multiple berth areas, different berth areas may be developed at different times, and the handling capacity and efficiency of its equipment varies due to the technical advance. Accordingly, the processing time of a vessel is berth-dependent.

Remaining services
Container vessels come to the terminal continually and get served. Operations planning at the terminal is normally done using a rolling horizon approach. At the beginning of a planning horizon, there are usually vessels previously berthed and still being processed. These vessels are referred to as remaining services and their unfinished work needs to be completed in the current planning horizon. Unlike the vessels to arrive or those waiting to be served, the remaining services have fixed berth positions and cannot be reallocated. Figure 2 shows a berth allocation result in the form of a time-space diagram. The black areas at the beginning of the planning horizon represent the remaining services carried over from previous planning horizon.

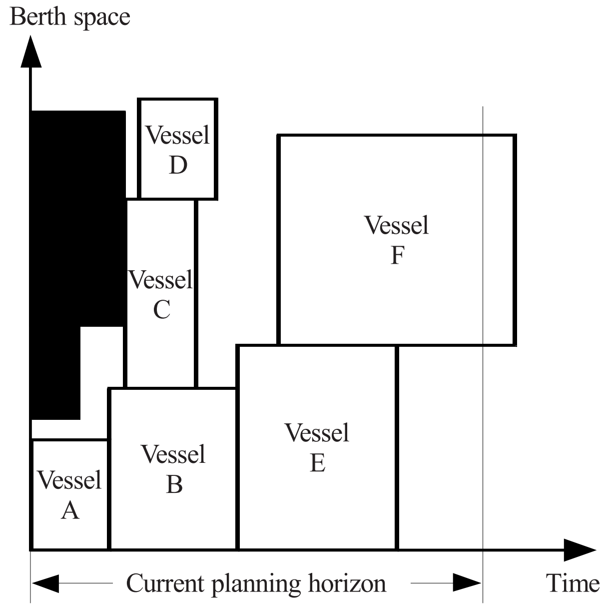


Figure 2: Remaining service at container terminal

Tang et al. (2009) explored the BAP at raw material docks of a steel company with multi-berth areas and also considered the remaining service in their work.

Service Priority
In practice, terminals give priority to certain vessels in berth allocation. These can be vessels belonging to partner shipping companies or vessels with tight shipping schedules. In our study, incoming vessels with priority are required to be served as soon as they arrive.

In this paper we formulated an integer programming model for the integrated berth and quay crane allocation problem considering all above mentioned features. In the rest of the paper, we first give a brief description of the problem setting in Section 2, and then present our model in Section 3. Section 4 reports preliminary computational results, before conclusions are drawn in Section 5.

2. Problem Description

We consider a terminal operator that runs several terminal areas in a container port. Each berth area is a straight berth line and equipped with a number of quay cranes. The handling capacities of cranes at different berth areas may be different. A berth area may serve one or more vessels simultaneously within its length limit. Vessels come over time to be served. The physical dimensions, the expected arrival time and the numbers of containers to be discharged and loaded for each vessel are known.

Among all the vessels some have priorities. A vessel may be allocated to any berth area that is capable to handle it. The processing time of a vessel depends on its workload and the berth area and the number of quay cranes allocated to it. For a given berth area and a given number of quay cranes allocated to a vessel, the processing time of the vessel can be calculated in advance using the vessel-level model by Liu et al. (2006). The problem is then to determine the berth and crane allocation to the vessels in the planning horizon (e.g., a week) so that the maximum service delay is minimized. Vessels with priority need to be served as soon as they arrive. The remaining services being processed at the beginning of the planning horizon have fixed berth positions. But the numbers of quay cranes assigned to them in the current planning horizon are to be determined.

3. Model Formulation

In this section, we present an integer programming formulation for the integrated problem of berth allocation and quay crane allocation. We first define the notation used before presenting the model.

Definitions of sets:

$\Omega = \{1, 2, 3\ldots\ldots N\}$: the set of incoming vessels within the planning horizon

$\Omega_r = \{N+1, N+2\ldots\ldots N+K\}$: the set of remaining services from previous planning horizon

Ω_p: the set of vessels with service priority, $\Omega_p \subseteq \Omega$

$\Phi = \{1,2\ldots\ldots\ldots N_b\}$: the set of berth areas.

Φ_i = the set of berth areas that can serve ship *i*, $\Phi_i \subseteq \Phi$

W: the set of all vessel pairs (i, j), $i, j \in \Omega \cup \Omega_r$, $i \neq j$

Parameters for berth allocation:

a_i = the expected arrival time for vessel *i*, $i \in \Omega$

p_{ic}^{B} = the processing time of vessel *i* if served by *c* cranes at berth area *B*

l_i = the length of vessel *i*, $i \in \Omega$

L_B^s = start position of berth area *B*, $B \in \Phi$

L_B^e = end position of berth area *B*, $B \in \Phi$

Parameters for quay crane assignment:

$\underline{C_i}$ = the minimum number of cranes that can be assigned to vessel *i*, $i \in \Omega$

$\overline{C_i}$ = the maximum number of cranes that can be assigned to vessel *i*, $i \in \Omega$

C_B = the total number of cranes at berth area *B*, $B \in \Phi$

Parameters for remaining service:

b_k = the start berth position of remaining service *k*, $k \in \Omega_r$

B_k = the berth area of remaining service *k*, $k \in \Omega_r$

l_k = the length of remaining service *k*, $k \in \Omega_r$

M: A very large positive number.

Continuous variables:

x_i = the start time of the service on vessel *i*, $i \in \Omega \cup \Omega_r$

s_i = the processing time of vessel *i*, $i \in \Omega \cup \Omega_r$

d_i = the departure time of vessel *i*, $i \in \Omega \cup \Omega_r$

s_k = the processing time of remaining service *k*, $k \in \Omega_r$

d_k = the departure time of remaining service *k*, $k \in \Omega_r$
y_i = the start berth position of vessel *i*, $i \in \Omega \cup \Omega_r$
f_i = the end berth position of vessel *i*, $i \in \Omega \cup \Omega_r$
u_i = the delay of the service on vessel *i*, $i \in \Omega \cup \Omega_r$
U = the maximum service delay for all vessels in planning horizon

Integer variables for berth allocation and crane assignments:

$$Z_i^B = \begin{cases} 1 & \text{if vessel } i \text{ is allocated to berth area } B \\ 0 & \text{otherwise} \end{cases}, \quad i \in \Omega, B \in \Phi$$

$$V_{ic} = \begin{cases} 1 & \text{if } c \text{ cranes are assigned to vessel } i \\ 0 & \text{otherwise} \end{cases}, \quad i \in \Omega \cup \Omega_r, \ \underline{C_i} \le c \le \overline{C_i}$$

C_i^B = the number of cranes assigned to vessel *i* at berth area *B*, $i \in \Omega \cup \Omega_r, B \in \Phi_i$
C_{ij}^B = the number of cranes that are serving vessel *j* at berth area *B* when vessel *i* arrives,
$(i, j) \in W, B \in \Phi_i$

Integer variables on relative position:

$$\sigma_{ij} = \begin{cases} 1 & \text{if completion time of vessel } i \text{ is earlier than the start time of vessel } j \\ 0 & \text{otherwise} \end{cases}$$

$$\delta_{ij} = \begin{cases} 1 & \text{if the ending berth position of vessel } i \text{ is below the starting berth position of vessel } j \\ 0 & \text{otherwise} \end{cases}$$

$$\beta_{ij} = \begin{cases} 1 & \text{if the start time of vessel } j \text{ is earlier than the start time of vessel } i \\ 0 & \text{otherwise} \end{cases}$$

$$\alpha_{ij} = \begin{cases} 1 & \text{if vessel } j \text{ is under service when vessel } i \text{ arrives} \\ 0 & \text{otherwise} \end{cases}$$

For the above four variables, $(i, j) \in W$.

The model:

Minimize U

Subject to

$u_i = x_i - a_i$	$i \in \Omega \cup \Omega_r$	(1)
$U \ge u_i$	$i \in \Omega \cup \Omega_r$	(2)
$s_i \ge p_{ic}^B + (V_{ic} + Z_i^B - 2)M$	$i \in \Omega, B \in \Phi, c = \underline{C_i}, \ldots, \overline{C_i}$	(3)
$s_i \le p_{ic}^B + (2 - V_{ic} - Z_i^B)M$	$i \in \Omega, B \in \Phi, c = \underline{C_i}, \ldots, \overline{C_i}$	(4)
$x_i - x_j \ge (\beta_{ij} - 1)M$	$(i, j) \in W$	(5)
$x_j - d_i \ge (\sigma_{ij} - 1)M$	$(i, j) \in W$	(6)
$y_j - f_i \ge (\delta_{ij} - 1)M$	$(i, j) \in W$	(7)
$C_i^B \ge \sum_{c=\underline{C_i}}^{\overline{C_i}} c V_{ic} + (Z_i^B - 1)M$	$i \in \Omega, \ B \in \Phi$	(8)

$$C_i^B \le \sum_{c=\underline{C_i}}^{\overline{C_i}} cV_{ic} - (Z_i^B - 1)M \qquad i \in \Omega,\ \ B \in \Phi \tag{9}$$

$$C_i^B = \sum_{c=\underline{C_i}}^{\overline{C_i}} cV_{ic} \qquad i \in \Omega_r,\ \ B = B_i \tag{10}$$

$$C_i^B = 0 \qquad i \in \Omega_r \quad B \ne B_i \tag{11}$$

$$\sum_{B=1}^{N_B} C_i^B = \sum_{c=\underline{C_i}}^{\overline{C_i}} cV_{ic} \qquad i \in \Omega,\ \ B \in \Phi \tag{12}$$

$$C_i^B \ge 0\,, C_{ij}^B \ge 0 \qquad i \in \Omega,\ \ B \in \Phi \tag{13}$$

$$\alpha_{ij} + \sigma_{ji} \le 1 \qquad (i, j) \in W \tag{14}$$

$$\alpha_{ij} \le \beta_{ij} \qquad (i, j) \in W \tag{15}$$

$$\beta_{ij} - \alpha_{ij} - \sigma_{ji} \le 0 \qquad (i, j) \in W \tag{16}$$

$$C_{ij}^B \le \alpha_{ij} M \qquad (i, j) \in W\,,\ \ B \in \Phi \tag{17}$$

$$C_{ij}^B - C_j^B \le 0 \qquad (i, j) \in W\,,\ \ B \in \Phi \tag{18}$$

$$C_j^B - C_{ij}^B \le (1 - \alpha_{ij})M \qquad (i, j) \in W\,,\ \ B \in \Phi \tag{19}$$

$$\sum_{j=1, j \ne i}^{N} C_{ij}^B + C_i^B \le C_B \qquad (i, j) \in W\,, B \in \Phi \tag{20}$$

$$\sum_{c=\underline{C_i}}^{\overline{C_i}} V_{ic} = 1 \qquad i \in \Omega \cup \Omega_r \tag{21}$$

$$d_i = x_i + s_i \qquad i \in \Omega \cup \Omega_r \tag{22}$$

$$f_i = y_i + l_i \qquad i \in \Omega \cup \Omega_r \tag{23}$$

$$x_i = a_i \qquad i \in \Omega_p \tag{24}$$

$$x_i = 0 \qquad i \in \Omega_r \tag{25}$$

$$y_k = b_k \qquad k \in \Omega_r \tag{26}$$

$$\delta_{ij} + \delta_{ji} + \sigma_{ij} + \sigma_{ji} \ge 1 \qquad (i, j) \in W \tag{27}$$

$$\delta_{ij} + \delta_{ji} \le 1 \qquad (i, j) \in W \tag{28}$$

$$\sigma_{ij} + \sigma_{ji} \le 1 \qquad (i, j) \in W \tag{29}$$

$$\beta_{ij} + \beta_{ji} = 1 \qquad i, j \in W \tag{30}$$

$$\sigma_{ik} = 0 \qquad i \in \Omega, k \in \Omega_r \tag{31}$$

$$\sum_{B=1}^{N_B} Z_i^B = 1 \qquad i \in \Omega \cup \Omega_r, B \in \Phi \tag{32}$$

$$Z_i^B = 0 \qquad i \in \Omega \cup \Omega_r, B \in \Phi \setminus \Phi_i \tag{33}$$

$$x_i \ge a_i \qquad i \in \Omega \cup \Omega_r \tag{34}$$

$$y_i \ge (Z_i^B - 1)M + L_B^s \qquad i \in \Omega, B \in \Phi \tag{35}$$

$$y_i \le (1 - Z_i^B)M + L_B^e - l_i \qquad i \in \Omega, B \in \Phi \tag{36}$$

$$\delta_{ij} \in \{0,1\},\ \ \sigma_{ij} \in \{0,1\},\ \ \beta_{ij} \in \{0,1\},\ \ \alpha_{ij} \in \{0,1\},\ \ (i, j) \in W \tag{37}$$

$$Z_i^B \in \{0,1\} \qquad i \in \Omega \cup \Omega_r, B \in \Phi \tag{38}$$

$$V_{ic} \in \{0,1\} \qquad i \in \Omega \cup \Omega_r,\ \underline{C_i} \le c \le \overline{C_i} \qquad (39)$$

The objective of this model is to minimize the maximum delay of service for the incoming vessels. Constraints (1) and (2) define the objective. Constraints (3) and (4) ensure that the processing time for each vessel is determined by the number of cranes assigned to it and by its berth area. The constraints identify the appropriate processing time from a matrix of pre-calculated processing times p_{ic}^{B}. Table 1 shows an example of such matrix for a vessel. The processing times in the matrix can be generated by the vessel-level quay crane assignment model by Liu et al (2006), which gives the minimum processing time of a vessel for a given number of quay cranes assigned to it.

Table 1: Processing times (hours) of a vessel with different berth and quay crane allocations

No of Cranes	Berth area 1	Berth area 2	Berth area 3	Berth area 4
1	18	18	18	×
2	10	10	10	×
3	8	8	8	×
4	6	6	6	×
5	4	4	4	×

×: This berth area cannot serve this vessel due to equipment limitation.

Constraints (5)-(7) identify the relative position of each pair of vessels in the time-space diagram. Constraints (8)-(13) define the allocation of quay cranes assigned to each vessel. Constraints (14)-(16) force α_{ij} equal to 1 when vessel *j* is under service when vessel *i* arrives. Constraints (17)-(20) ensure that, at any time, the total number of cranes assigned to vessels at a berth area must not exceed the total number of cranes at this berth area. Constraints (21) guarantee that each vessel is assigned certain number of cranes within the allowed range. Constraints (22) and (23) define the departure time and the end berth position for each vessel. Meanwhile they ensure the continuity of the service and berth space. Constraints (24) introduce the priority scheme so that vessels with priority will be served as soon as they arrive. Constraints (25) and (26) set the berth position and start service time of remaining services carried over from previous planning horizon. Constraints (27)-(31) together ensure that there is no overlapping among the vessels in the time-space diagram. Constraints (32) guarantee that each vessel can only be allocated to one berth area. Constraints (33) exclude circumstances that berth areas with insufficient equipment are assigned to vessels. Constraints (34) enforce that the handling process of a vessel can only start after the arrival of the vessel. Constraints (35) and (36) set the boundaries for berth areas and define the decision variable Z_i^B for berth area assignment. Constraints (37)-(39) define the binary decision variables.

4. Results of Preliminary Experiment

We have done preliminary experiment to solve the integrated berth and quay crane allocation problem using the model presented above, to illustrate the feasibility of the model and to observe the computation time requirement. The models were formulated and then solved using the Xpress software package on a PC with Intel Core2 CPU and 1Gb of memory.

Six problem cases with different number of incoming vessels, ranging from 8 to 20, were used in the experiment. All cases were solved to optimality. Figure 3 shows the computation time used to solve the problems. From the figure we can see that the computation time is acceptable for problem size up to 20 vessels. However, the computation time increases dramatically when there are more than 16 vessels. This observation is consistent with the structure of the model. The complexity of the integer programming model is due to the integer variables. The number of integer variables in the model is in the order of the number of vessels squared. Therefore, as the number of vessels increases, the computation time required to optimally solve the model increases dramatically.

If the number of vessels further increases beyond 20, it would be difficult to get an optimal solution within reasonable computation time on the same PC. This suggests the necessity of developing heuristics to generate approximate solution in relatively short time. However the formulation could be useful in understanding the relationships among different decisions in the integrated problem and hence providing hints in develop effective heuristics. The model solution can also serve as benchmarks for testing heuristics on problems with manageable size.

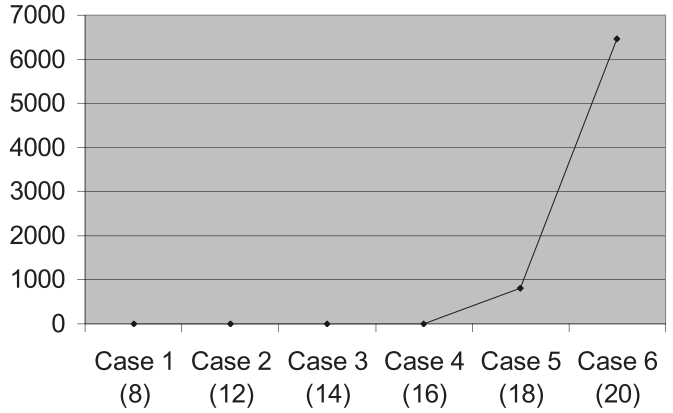


Figure 3: Computation time (seconds) for different cases

5. Conclusions

In this paper we studied the integrated problem of berth allocation and quay crane allocation at container terminals, and formulated this problem into an integer programming model. The model is applicable to container terminals with multi-berth areas, and considers practical constraints including priority scheme, remaining service carried over from previous planning horizon, and berth-dependent processing times. Especially the model gives more practical and accurate consideration of the processing times. Instead of taking the processing time of each vessel as a pre-determined fixed value, the model determines the exact processing time based on the workload and number of quay cranes assigned to it. Preliminary experiment indicated that the model could solve problems with up to 20 vessels. It could also be helpful in developing and testing heuristic methods for larger problems.

6. Acknowledgements

The research of the second author was partially supported by the National Natural Science Foundation of China (Grant No. 70728001).

References

Guan, Y.P. and Cheung, R.K. (2004), The berth allocation problem: models and solution methods, OR Spectrum 26(1): 75-92.

Guan, Y.P., Xiao, W.Q., Cheung, R.K. and Li, C.L. (2002), A multiprocessor task scheduling model for berth allocation: heuristic and worst-case analysis, Operations Research Letters 30(5): 343-350.

Imai, A., Nagaiwa, K. and Tat, C.W. (1997), Efficient planning of berth allocation for container terminals in Asia, Journal of Advanced Transportation 31(1): 75-94.

Imai, A., Chen, C.H., Nishimura, E. and Papadimitriou, S. (2008), The simultaneous berth and quay crane allocation problem. Transportation Research Part E-Logistics and Transportation Review 44(5): 900-920.

Imai, A., Nishimura, E. and Papadimitriou, S. (2003), Berth allocation with service priority, Transportation Research Part B-Methodological, 37(5): 437-457.

Imai, A., Nishimura, E. and Papadimitriou, S. (2001), The dynamic berth allocation problem for a container port, Transportation Research Part B-Methodological 35(4): 401-417.

Imai, A., Sun, X., Nishimura, E. and Papadimitriou, S. (2005), Berth allocation in a container port: using a continuous location space approach, Transportation Research Part B-Methodological 39(3): 199-221.

Kim, K.H. and Moon, K.C. (2003), Berth scheduling by simulated annealing, Transportation Research Part B-Methodological 37(6): 541-560.

Lim, A. (1998), The Berth planning problem, Operations Research Letters 22(2-3): 105-110.

Liu, J., Wan, Y.-w. and Wang, L. (2006), Quay crane scheduling at container terminals to minimize the maximum relative tardiness of containership departures, Naval Research Logistics 53(1): 60-74.

Meisel, F. and Bierwirth, C. (2009), Heuristics for the integration of crane productivity in the berth allocation problem, Transportation Research Part E-Logistics and Transportation Review 45(1): 196-209.

Nishimura, E., Imai, A. and Papadimitriou, S. (2001), Berth allocation planning in the public berth system by genetic algorithms, European Journal of Operational Research 131(2): 282-292.

Park, K.T. and Kim, K.H. (2002) Berth scheduling for container terminals by using a sub-gradient optimization technique, Journal of the Operational Research Society 53(9): 1054-1062.

Park, Y.M. and Kim, K.H. (2003), A scheduling method for Berth and Quay cranes, OR Spectrum 25(1): 1-23.

Tang L., Li S.H. and Liu J. (2009), Dynamically scheduling ships to multiple continuous berth spaces in an iron and steel complex, International Transactions in Operational Research 16(1): 87-107.

The challenge of trade liberalization in logistics services: the case of Indonesia

Jose Tongzon

Graduate School of Logistics, Inha University, Incheon, South Korea 402751.
Phone: +8232 860 8234
Fax: +8232 860 8226
E-mail: jtongzon@inha.ac.kr

Abstract

There has been a general movement towards liberalization of international trade in logistics services in the multilateral, regional and bilateral contexts as nations have realized the economic benefits of trade liberalization under the auspices of WTO, regional and bilateral free trade agreements. The current global financial and economic crisis gripping both developed and developing countries has further raised the issue of the nature, extent and the pace of trade liberalization in logistics services particularly for developing countries whose logistics services sector is not yet as efficient and as competitive as their counterparts in the developed world.

In this light, this paper intends to assess the economic implications of trade liberalization in logistics services from the perspective of Indonesia – a typical developing country with a logistics sector that is still at its infancy stage. Specifically, it tries to examine the competitiveness and efficiency of Indonesia's logistics services sector, its current regulatory and policy framework for its logistics services sector and the policy implications of trade liberalization in logistics services.

Keywords: Trade liberalization, logistics services, Indonesia, policy implications

1. Introduction

There has been a general movement towards liberalization of international trade in logistics services in the multilateral, regional and bilateral contexts as nations have realized the economic benefits from trade liberalization under the auspices of the World Trade Organization (WTO), regional and bilateral free trade agreements. Trade liberalization in logistics services is seen as part of a strategy to improve exports and achieve economic development. Indonesia in particular has pursued a multi-track approach to pursue its export interests. As an active member of the WTO, Indonesia has committed to the most-favoured-nation principle and participates in negotiations under the framework of General Agreement on Trade in Services. As a founding member of ASEAN, Indonesia has committed to the ASEAN vision of an economically integrated Southeast Asian region and the free flow of services by 2015. Indonesia has also made a number of commitments under the Indonesia-Japan Economic Partnership Agreement.

However, the current global financial and economic crisis gripping both developed and developing countries has further raised the issue of the nature, extent and the pace of trade liberalization in logistics services particularly for developing countries whose logistics services sector is not yet as efficient and as competitive as their counterparts in the developed world. In this light, the objective of this paper is to assess the economic implications of trade liberalization in logistics services from the perspective of Indonesia – a typical developing country with a logistics sector that is still at its infancy stage. Specifically, it tries to examine the competitiveness of Indonesia's logistics services sector, its current regulatory and policy framework for its logistics services sector and the policy implications of trade liberalization in logistics services.

The scope for logistics adopted in this paper is based on the recommendation by the World Trade Organization (WTO) which is also employed by the ASEAN countries in their current effort towards the integration of their logistics services. It consists of three major categories: core freight logistics services, related freight logistics services and non-core freight logistics services. For more detailed description of these categories and their respective codes, see Appendix 1.

2. Indonesia's Logistics Services Sector

Like other ASEAN countries, the services sector is increasingly becoming important to the Indonesian economy. This is trend is main driven by Indonesia's continued economic growth and rising demand for services which is generally more responsive to the growth in income than the demand for agriculture and manufactures. Its services sector has grown significantly over the past 36 years so that by 2006 it accounted for almost 50 percent of Indonesia's total output and employment, respectively (Harjono and McGuire, 2006). Indonesia's logistics services are one major component of this services sector playing a significant role in its rising economic importance.

Further, its services sector and the logistics services sub-sector in particular are less export-oriented than its manufactures and rely so much on Indonesia's large domestic market. As of 2006, the services sector accounted for only 0.45 percent of Indonesia's overall exports and 1.19 percent of its overall imports (Harjono and McGuire, 2006). Indonesia is a net importer of services and therefore contributes to the outflow of Indonesia's foreign exchange with a negative implication for its balance of payments position.

Since there are no trade data on logistics as a whole, trade data on transportation which is available from UNCTAD's database can be used as a proxy for logistics services. These data are not decomposed into countries but are only aggregates in terms of exports and imports. However, these data should give us some indication of Indonesia's logistics trade performance compared to other older countries in ASEAN for which data are available.

Table 1: Exports and Imports of Transportation Services in Selected ASEAN countries (million USD)

Year	Indonesia		Malaysia		Philippines		Singapore		Thailand	
	Exports	Imports	Exports	Imports	Exports	Imports	Exports	Imports	Exports	Imports
1990	70	2,795	1,198	2,531	246	980	2,225	3,513	1,327	3,576
1991	81	3,187	1,389	3,023	242	1,062	2,517	3,794	1,484	4,185
1992	89	3,574	1,593	3,181	273	1,199	2,766	3,582	1,526	4,539
1993	44	3,667	1,613	3,433	218	1,425	3,173	4,015	1,964	5,005
1994	..	3,913	1,871	4,356	233	1,457	4,042	4,797	1,842	5,862
1995	..	4,861	2,466	5,609	274	2,051	8,296	9,290	2,455	7,780
1996	..	5,244	2,822	5,433	358	2,287	8,935	10,107	2,618	7,845
1997	..	5,400	2,861	5,549	357	2,673	8,780	10,377	2,417	6,890
1998	..	3,730	2,271	4,088	324	1,983	9,064	9,574	2,671	4,604
1999	..	3,274	2,492	4,720	307	2,011	10,690	11,094	3,017	5,305
2000	..	4,016	2,802	5,890	464	2,052	11,889	12,814	3,250	6,760
2001	..	3,877	2,748	5,736	590	2,019	11,531	12,383	3,057	6,830
2002	1,058	5,150	2,855	5,892	877	2,303	12,015	10,905	3,265	7,121
2003	856	4,824	2,767	6,260	951	2,419	13,557	13,308	3,503	8,484
2004	2,279	5,474	3,163	7,842	1,001	3,095	16,923	17,815	4,350	10,830
2005	2,841	7,030	4,056	8,396	1,041	3,125	17,904	19,876	4,626	13,999

Source: UNCTAD Handbook of Statistics 2006-2007

As shown in Table 1, the trade performance of Indonesia's transport sector over the period of 1990 to 2005 (the latest year for which data are available) has not been contributing to Indonesia's current account balance. The discrepancy between its export and import of transportation services has been quite substantial over the years. Export of transportation services has not been performing as well as the countries of Singapore, Malaysia and Thailand.

Indonesia's logistics sector has only accounted for a small proportion of its GDP which indicated the nascent nature of its logistics sector. Compared to other four ASEAN countries, its logistic services sector is probably the smallest sector, except for the Philippines.

3. Indonesia's Logistics Policy: Its Objective and Implementation

Indonesia's logistics policy is supposed to complement its overall trade objective of promoting Indonesia's export performance by increasing its export of logistics services and by managing the flow of logistics services such that they contribute to the productivity and international competitiveness of the Indonesian economy.[1] Achieving this objective in logistics services trade will necessarily move Indonesia closer to its strategic vision: to realize international trade as a core driver for increasing national competitiveness and the welfare of the Indonesian people.[2]

In an effort to open up more markets for its exports and further liberalize the international trading environment, Indonesia has pursued a number of approaches which may be categorized into three major forms: multilateral under the auspices of the World Trade Organization (WTO), regional within the framework of the ASEAN Economic Community (AEC) and bilateral in the context of Indonesia's individual relationships with its individual trading partners.

As an active member of the WTO, Indonesia is committed to the most-favoured-nation (MFN) principle and has participated in a number of negotiations under the General Agreement on Trade in Services (GATS) where countries have to make bilateral offers and requests in the context of the member countries' limitations on market access (MA) and national treatment (NT). As a founding member of ASEAN, Indonesia has committed itself to the ASEAN vision of an economically integrated Southeast Asian region and in particular has officially supported the objectives under the ASEAN Roadmap for the Integration of Logistics Services. In bilateral contexts Indonesia has also upheld certain commitments under the Japan-Indonesia Economic Partnership Agreement (JIEPA) which came into force in 2007.

Table 2 summarizes Indonesia's international trade in logistics services commitments under the WTO, ASEAN and JIEPA.

[1] National Medium Term Development Plan 2004 – 2009, Government of Indonesia, 2003

[2] Strategic Plan 2004 – 2009, Ministry of Trade, 2003

Table 2: Indonesia: International trade in logistics services commitments

WTO	ASEAN	JIEPA
General: Indonesia is committed to the liberalization of logistics services across the board to reap the benefits of international trade. Key specific commitments: • Undertake substantial and meaningful market access (MA) and national treatment (NT) commitments as well as other commitments as may be needed to ensure the effectiveness of liberalization commitments; • In respect of core services essential to logistics operations, MA and NT commitments with right of establishment should be provided so as to enable commercially meaningful liberalization; where limitations and restrictions are to be scheduled, they may be time-limited to be phased out; • Ensure that access is provided on reasonable and non-discriminatory basis; • Ensure that procedures and formalities are not unnecessarily burdensome.	General: Indonesia is committed to the implementation of the ASEAN Roadmap for the integration of logistics services by 2013. Key specific commitments: • Achieve substantial liberalization of logistics services in the following sectors: maritime cargo-handling services (CPC 741), storage/warehousing (CPC 742), transport agency (CPC 748), other auxiliary services (CPC 749), courier (CPC 7512), packaging services (CPC 876), maritime transport services (CPC 7212), air transport (CPC 732), rail transport (CPC 7112) and road transport services (CPC 7213); • Enhance competitiveness of logistics services providers through trade and logistics facilitation; • Expand capability of logistics service providers; • Enhance human resource development; • Enhance multi-modal transport infrastructure and investment.	General: Indonesia is committed to the implementation of the terms under the JIEPA agreement. Key specific commitments: • Undertake substantial market access and national treatment commitments as specified in the Agreement; • Ensure that Indonesia's commitments are consistent with the economic objectives and priorities of JIEPA; • Ensure that Indonesia's commitments are accompanied by capacity building and other programs aimed at strengthening Indonesia's international competitiveness;

Source: Tongzon (2009)

Despite its declared objectives, the implementation of its trade liberalization commitments in the area of logistics services has been slow and cautious. Indonesia's logistics services are still highly shielded from foreign competition and this sheltered environment has contributed greatly to their relative inefficiency and low international competitiveness. The liberalization process in logistics services has been very slow particularly in the maritime transport and distribution sub-sectors due to their political sensitivities. Specifically, the current regulations on foreign equity ownership and conditions for foreign establishments are still quite limited and discriminatory in favour of domestic enterprises. The lack of having a clear definition and scope for logistics services has also made it difficult for Indonesia to negotiate for more market access and national treatment.

Although there are potential benefits that can be derived from international trade in logistics services, there are perceived short-term political and adjustment costs which need to be managed effectively. Reforming those logistics sectors where Indonesia has a competitive disadvantage entails a strong opposition from those firms that are negatively affected by foreign competition. In addition, reform in Indonesia's logistics sector does not guarantee an increased inflow of foreign investment. Indonesia's overall infrastructure needs to be further improved and its overall climate for foreign investment must be attractive enough for foreign investors in logistics. After all, the private sector will have to do their own calculations in deciding whether they want to invest in Indonesia or not.

4. Economic Implications of Trade Liberalization in Logistics Services

The likelihood that Indonesia is going to accelerate the process of liberalizing its logistics sector largely hinges on how they perceive the likely impacts of this liberalization. In particularly, it is likely that the perceived economic benefits and adjustment costs will affect the pace and the nature, extent and pace of trade liberalization in logistics services.

4.1. Economic Benefits

Trade liberalization in logistics services can benefit Indonesia in two main ways. First, it can support Indonesia's international trade policy objective. Indonesia's international trade policy objective is to promote Indonesia's overall export performance, which can be partly achieved by exporting those specific logistics services where Indonesia is currently and potentially internationally competitive. Second, trade liberalization in logistics services can improve Indonesia's international competitiveness in the export of goods. Logistics costs are a major part of the overall cost of exporting and have become even more important as tariffs and other non-tariff barriers to international trade have progressively been reduced multilaterally and regionally.

To increase Indonesian exporters' ability to respond quickly to fast-changing market demands and compete in an increasingly competitive environment, Indonesian exporters need to have access to the most efficient and cost-effective logistics services available. Given its current fragmented and developing logistics sector, developing links with foreign best-practice logistics providers could be a shorter way to improve Indonesia's export competitiveness. For example, PT Monang Sianipar Abadi (MSA) Kargo has strengthened its presence in the global logistics market by forming a partnership with a US logistics company. Domestic logistics company PT Fajar Insan Nusantara (FIN) Logistics has tapped into the overseas market by affiliating with Helmann Worldwide and NOL (The Jakarta Post, 14 August 2008).

Further, importing logistics services can improve the efficiency and reduce the cost of domestic logistics services by bringing more competition into Indonesia's logistics sector. For example, the recent new shipping law which allows more private sector participation in the operation and ownership of their ports will not only improve the efficiency of their ports but will also contribute to the enhancement of Indonesia's export competitiveness.[3]

Although this may mean substantial adjustment costs for the domestic logistics sector, these effects could be more than offset by lower costs and greater efficiency for logistics users in the long run. Liberalizing the logistics services sector in a measured and gradual way over time for foreign logistics service providers can further help cushion Indonesia from the short-run adjustment costs while getting more foreign private participation in the management and operation of its key transport infrastructure assets.

A number of countries in the region have already done this with a great deal of commercial success. Singapore, Thailand and Malaysia, for example, have been able to provide value for money services to their exporters with even greater speed, quality and reliability. They have maintained and further improved their overall efficiency by reforming their ports and other logistics service sectors and by constantly looking for ways through which competition can be increased.

[3] The new shipping law was enacted and implemented in 2008.

4.2. Areas of International Competitiveness

To turn the above potential economic benefits into reality requires that Indonesia focuses its exports on those areas where Indonesia is internationally competitive. Due to the lack of export and import data on the various components of logistics services, it is not possible to address this issue based on the standard economic approach. The standard approach to quantify Indonesia's competitive advantage makes use of world market shares and revealed comparative indices.

In the absence of trade data on logistics services by sub-sector, the paper adopts a qualitative and hands-on approach. First, it is assumed that a country has a competitive advantage in a certain type of logistics service if she is relatively endowed with those resources which are used relatively more intensively in the production of that type of logistics service. Indonesia, being endowed with relatively abundant and cheap but low-skilled labour, should have competitiveness in the production of labour-intensive logistics services. Second, the results of any previous studies on logistics services dealing with Indonesia's comparative advantage can provide a useful basis for assessment. Third, a survey of selected experts in Indonesia's logistics sector (selected from relevant government agencies in Indonesia and representatives of the private sector and private researchers) will provide another basis for forming our own assessments.

Indonesia's competitive advantage using the above methodology should be found in those logistics services which use relatively more low or semi-skilled labour, which Indonesia is relatively abundant with. However, given their limited access to capital and state-of-the-art technology constrained by less developed infrastructure and institutional quality, Indonesia's logistics services are relatively inefficient and deal mostly with low-valued and basic logistics operations. It is therefore difficult for Indonesia to export these types of services under other modes of supply. Further, there are also certain regulations that have limited the market access for logistics services in these countries. Thus, in the short run Indonesia's capacity and competitive advantage only lies in exporting its low and semi-skilled labour under Mode 4 in certain sectors that require these types of labour. Indonesia has already been an exporter of seamen for foreign commercial ships. There are currently 300,000 Indonesian nationals working in Malaysia as domestic helpers (The Jakarta Post, 4 August 2008). There is no reason why Indonesia cannot become a major exporter of low and semi-skilled workers for the ports, airports and other logistics services sub-sectors in other labour-scarce countries in the region and beyond.

These sub-sectors include the following: cargo-handling services, storage and warehousing services, transport agency services, maritime transport services, internal waterways transport services, road transport services, air transport services, retailing services, other supporting services and packaging services. Except for cargo-handling, maritime, air transport and packaging services where Indonesia has a competitive advantage under other modes of supply, Indonesia's competitive advantage is under Mode 4.

Indonesia potentially holds a competitive advantage in exporting the above services under other modes of supply, provided that Indonesia can implement the required institutional reforms, develop its transport infrastructure (for example, expansion and upgrading its road and rail networks, seaports and airports) and address other constraints facing its logistics industry. For example, packaging services is an area where Indonesia has a potential competitive advantage due to its well-established packaging industry, which boasts revenues of Rp 20 trillion (USD 2.18 billion) a year. Provided that there is a stable source of polypropylene (PP) which is used in plastic-based packaging material, this industry is likely to face a bright prospect (The Jakarta Post, 28 July 2008).

4.3. Economic Costs

The economic costs of trade liberalization will be short-term adjustment costs in terms of employment losses in those logistics sectors where Indonesia has competitive disadvantage. These types of logistics services are usually employing a more capital or skilled labour-intensive technology which Indonesia is relatively short of. Specifically, these are the high-valued logistics services such as supply chain consulting, inventory management, logistics design, forecasting and planning, database and data computing services. These would also include the supply of logistics services under mode 3 (i.e. in the form of commercial presence).

However, if appropriate measures are in place to bring about an effective transfer of technology, this should lead to the enhancement of Indonesia's capacity in these areas of high-valued logistic services which can contribute to Indonesia's logistics services exports.

It would therefore benefit in the long run the Indonesian trading community and the wider public to import these types of logistics services. The logistics component has increasingly become an important component in the entire international trading process. This is particularly true for Indonesia where the logistics costs account for a sizeable portion of total cost (World Bank, 2007). This implies that reform of the logistics services sector, especially in the rail and road transport sub-sectors and the auxiliary services for these two modes of transport, should go a long way in bringing down the cost of transport and improve the efficiency of delivery across the logistics chain.

In the case of cargo-handling and storage services, the insufficient number of consolidators has often resulted in a situation where goods either travel less than truckload or less than container load from origin to destination. The absence of adequate warehousing and specialized storage facilities has required Indonesian truckers to time their deliveries to vessel arrival, causing congestion, longer waiting times and delay costs (CARANA Corporation, 2004).

In the case of freight transport, there are a large number of inefficient ports inhibiting cargo aggregation leading to higher transportation and cargo handling costs. Consequently, attract frequent and competitive direct liner services has been difficult as cargo is dispersed over a large number of smaller ports and carried by its fleet of smaller and older vessels. Indonesia has over 140 operational ports. Of these, 43 ports are feeder ports which act as transshipment points to the major Indonesian gateways such as Jakarta or Surabaya. Lack of adequate road and air networks with limited size and capacity to handle the increasing flow of goods has further constrained the flow of cargoes throughout the logistics chain.

Table 3 lists the categories of logistics services that can be reformed to enhance Indonesia's international competitiveness under mode 3 and their rationale.

Table 3: Areas of competitive disadvantage and the rationale for reforms

Service sector	*Target service sector and mode of supply*	*Rationale*
I. Core Freight Logistics	• Cargo-handling services (CPC 741) - mode 3 • Storage and warehousing services (CPC 742) - mode 3 • Transport agency services (CPC 748) - mode 3 • Other auxiliary services (CPC 749) - mode 3	• Opening Indonesia's cargo-handling services to foreign and private competition will improve competition in the port sector and thus enhance Indonesia's competitiveness in the trade of goods. • Reform in this sector will lead to more modern and adequate storage and warehousing facilities in Indonesia which has been constrained by lack of adequate and modern storage and warehousing. • More foreign transport agencies and auxiliary services will improve the adequacy and efficiency of transport and reduce the cost of exporting as foreign transport agencies will bring with them the latest technology and best practices.
II. Related Freight Logistics	• Railway transport services (CPC 7112) - mode 3 • Road transport services (CPC 7123) - mode 3 • Air transport services (CPC 732) - mode 3	• Reform in rail, road and air transport services can provide greater capacity and efficiency for the delivery of Indonesia's traded goods to its export markets. • These in turn will further facilitate the movement of goods with access to the latest technology and best practice management.
	• Technical testing and analysis services (CPC 8676) - mode 3	• This will make Indonesia's manufactured exports and export of services more internationally competitive by having access to state-of-the-art testing technologies and standards
III. Non-core Freight Logistics	• Data processing services (CPC 843) - mode 3 • Database services (CPC 844) - mode 3 • Management consulting and related services (CPC 865) - mode 3	• This will stimulate more automation in the delivery of logistics services and facilitate the adoption of electronic data interchange among the various parties in the logistics chain. . • Moving Indonesia to a higher valued logistics exports through technology transfer.

Source: Tongzon (2009)

5. Policy implications

Given the potential gains that can be derived from international trade in logistics services and the high costs of policy reversal towards protectionism, Indonesia cannot afford to adopt a protectionist trade policy to cope with an increasingly competitive trading environment. Indonesia should therefore continue to pursue liberalization to become more integrated into the global economy and thus reap the benefits of international trade in logistics services.

The standard acceptable argument calls for a gradual liberalization process while substantial and organized effort is made to improve Indonesia's international competitiveness in the logistics services sector. The idea is to make its logistics services internationally competitive before full liberalization is fully implemented so that greater benefits only come with a minimum of injury to Indonesia's logistics services industry. This particular developmental approach if adopted will have implications for the type of negotiating strategy that should be adopted. This implies that liberalization should be pursued only in those areas where Indonesia has a competitive advantage while protecting those areas where Indonesia has a competitive disadvantage. This selective protectionism can only be lifted once these sectors have become more efficient and more able to compete in a liberalized trading environment.

Although this approach is politically acceptable and economically sensible, it is hard to implement this in practice as it requires a clear and effective timetable for liberalization, strong political will from the government and cooperation from the private sector. A clear timetable for this gradual liberalization is necessary and should be disseminated to the private sector so that the private sector will know exactly how long protectionism will stay, and thus can plan accordingly to improve their efficiency and performance. It also avoids the danger, as in the past, of the private sector's complacency on the assumption that government protectionism will remain forever.

It should be noted that liberalization in one sector or segment in the logistics chain may not have a full impact since logistics services providers deals with a bundle of logistics services and thus a liberalization of one aspect (say in cargo handling) without liberalizing the other aspects (say transport modes, storage and warehousing) may not really help facilitate the movement of goods from the point of production to the point of consumption. The offer therefore is to focus trade in services strategy on those segments or links in the logistics chain where we find the least restrictions on market access and national treatment. Prioritizing on these segments will entail least adjustment costs and should therefore be easier to implement. Although this approach does not address the weakest link in the chain - crucial to the improvement of the overall logistics performance, it does not ignore the political sensitivities of trade liberalization. However, the government should make its commitment to liberalization known to the private sector with a clear short and longer-term liberalization plan so that the private sector is assured of the full potential impact of trade liberalization in logistics services.

On the export side, the basic principle in international trade prescribes that Indonesia should focus on those logistics services where Indonesia has a comparative advantage. In addition, these priority sectors must also generate substantial economic gains in terms of their impacts on employment, contributions to Indonesia's economic growth and strategic importance as perceived by the Indonesian government. They must also be the world' fast growing logistics sectors so that Indonesia can take advantage of the fast growing export markets (Harjono and McGuire, 2006).

6. Conclusion

There is no doubt that logistics plays a crucial role in the economic development process of developing countries. Apart from facilitating international trade, it contributes to the enhancement of a nation's international competitiveness by adding customer value in terms of delivering the goods at the right time, at the right place and in the most cost-effective way. This is more so in the case of Indonesia. The importance of international trade in logistics to Indonesia is further based on the fact that logistics costs in Indonesia are considered to be quite high by international standards affecting Indonesia's international competitiveness. Having access to the most efficient logistics services is therefore crucial to Indonesia's export success. On the other hand, logistics services export is an important component of Indonesia's international trade objective. Reforms in the trade in logistics services will therefore have beneficial effects on Indonesia's international competitiveness and export performance.

In this context Indonesia has to engage in international trade in logistics services. In pursuing international trade, Indonesia has to focus its exports on those logistics services where Indonesia has major and potential competitive advantage in terms of resource endowments and market access. In the case of imports, it is worthwhile for Indonesia to focus on those services where Indonesia has a competitive disadvantage but which can enhance Indonesia's international competitiveness. In prioritizing these services however the political sensitivities and strategic interests have to be taken into account.

References

Harjono, A. S. and McGuire, G. (2006),Where Should Indonesia Focus its Trade in Services Strategy? Unpublished paper. Technical Assistance Management Facility (TAMF), Jakarta, Indonesia.

CARANA Corporation (2004), Impact of Transport and Logistics on Indonesia's Trade Competitiveness. May 2004. Produced for review by the United States Agency for International Development as part of the Trade Enhancement for the Services Sector (TESS) Project.

Indonesian Ministry of Trade (2003), National Medium Term Development Plan 2004–2009. Jakarta: Government of Indonesia.

Indonesian Ministry of Trade (2003), Strategic Plan 2004-2009.

The Jakarta Post, "Packaging industry hit by raw material shortage", 28 July 2008, p.13.

The Jakarta Post, "RI seeks more pay for maids", 4 August 2008, p.2.

The Jakarta Post, "Synergy in the map of express service cargo business", 14 August 2008, p. 18.

Tongzon, J.L. (2009), Indonesia's Interests in International Trade in Services Negotiations: Logistics Services. Report produced for the Government of Indonesia, Jakarta, Indonesia.

UNCTAD (2008), UNCTAD Handbook of Statistics (UNCTAD CD), Geneva.

World Bank (2007), Connecting to Compete: Trade Logistics in the Global Economy – The Logistics Performance Index and Its Indicators. Washington, DC: The World Bank. Accessible from http://www.worldbank.org/lpi.

Appendix 1: Scope for Logistics Services

Categories	**Codes**
I. Core Freight Logistics Services	
Cargo handling services	CPC 741
Container handling services	CPC 7411
Other cargo handling	CPC 7419
Storage and warehousing services	CPC 742
Transport agency services	CPC 748
Other auxiliary services	CPC 749
II. Related Freight Logistics Services	
(1) Freight transport services	
Maritime Transport Services	CPC 7212
Internal Waterways Transport Services	CPC 7222
Air Transport Services	
. Air freight transport	CPC 732
. Rental of aircraft with crew	CPC 734
Rail Transport Services	
. Freight transport	CPC 7112
Road Transport Services	
. Freight transport	CPC 7123
. Rental of commercial vehicles with operator	CPC 7124
- without operator	CPC 83102
(2) Other related logistics services	
Technical testing and analysis services	CPC 8676
Courier Services	CPC 7512
Commission Agents' Services	CPC 621
Wholesale Trade Services	CPC 622
Retailing Services	
. Food retailing services	CPC 631
. Non-food retailing services	CPC 632
. Sale of motor vehicles	CPC 6111
. Sale of parts and accessories of motor vehicles	CPC 6113
. Sales of motorcycles and snow mobiles & related parts & accessories	CPC 6121
Other supporting services not covered (CPC 743, 7113, 744 excluding 7441, and 746).	
III. Non-core Freight Logistics Services	
Packaging services	CPC 876
Leasing or rental services concerning vessels without crew	CPC 83103
Leasing or rental services concerning aircraft without operator	CPC 83104
Computer and related services	
. Data processing services	CPC 843
. Database services	CPC 844
Management consulting and related services	CPC 865

Notes: CPC 743: Supporting services for railway transport; CPC 7113: Pushing or towing services; CPC 744: Supporting services for road transport; CPC 7442: Highway, bridge & tunnel operation services; CPC 7443: Parking services; CPC 7449: Other supporting services for road transport; CPC 7441: Bus station services; CPC 746: Supporting services for air transport.

Ship standardisation on the Yangtze River

Albert W Veenstra

Rotterdam School of Management, Erasmus University Rotterdam, 3000 DR Rotterdam, The Netherlands.

Abstract

We present an analysis of the feasibility of the Chinese inland shipping standardisation program on the Yangtze River. This program is implemented around the Three Gorges Dam in the upper reach of the Yangtze River. For this section of the river, there are specific standardisation targets for the years 2010, 2015 and 2020. This paper aims to investigate if and to what extent the targets set for standardisation of container ships on the main fairway of the Yangtze River will actually be met in the future.

We conclude that the standardisation in the upper reach of the river will succeed according to the targets if ship building occurs strictly in the standards. Some of the intended effects of standardisation, such as increasing the average ship size and reducing fleet diversity, will not materialise.

1. Introduction

The Yangtze River is an important connection between the booming East with the developing Central and Western parts of the People Republic of China (PR China). It links important cities such as Shanghai, Nanjing, Wuhan and Chongqing along its navigable section that stretches 2400 km.

The Chinese government has been stimulating the development of especially container shipping in the Yangtze River for several years. The Yangtze Golden Waterway, as the development program is often dubbed, is an integral part of the 11th Five Year Plan (2006-2010). The aim is to develop the river into a modern, high performance waterway by the year 2020 [5]. Important components of the Yangtze River development plan are the deepening of various river sections, the development of modern port facilities, and the construction of the Three Gorges Dam. The Dam will not only generate electrical power but should also stabilise the water level in a large section of the river from Wuhan to Chongqing, and thus improve the navigability in this part of the river.

A bottleneck to the river's development is the structure of the container ship fleet that is operating on the main fairway and the many tributaries. The lack of standards and the surplus of old, technically outdated and small ships currently prevents making 'full use of the advantages of the water transport system' (PR China's late vice-prime minister Huang Ju in his speech to provincial governors at the annual Ministry of Communications (MOC) work meeting in 2005).

MOC has initiated a national ship standardisation program for inland vessels about thirty years ago. The execution of this program for the Yangtze River has been delegated to the Chang Jiang River Administration of Navigational Affairs (CJHY). The goal of standardisation of vessels is to improve the performance of the fleet, the operation of ship locks and the utilisation of the main fairways in rivers. This program consists of sets of standard dimensions for various inland ship types and various river systems, targets for the standardisation program for 2010, 2015 and 2020 and additional governance policies, such as guidelines for the development of new standards (see stand2).

This paper aims to investigate if and to what extent the targets set for standardisation of container ships on the main fairway of the Yangtze River will actually be met in the future and how the standardisation program will change the structure of the fleet. This investigation is based on the available statistics on Yangtze river and Chinese merchant vessels and a dynamic fleet simulation approach. This is the first time that such an analysis is presented for the Yangtze River and for inland shipping. The data processing and fine tuning of the fleet simulation model we present here are also new. The analysis in this paper aims to contribute to the understanding of Chinese governance of the Yangtze River Transport System, and the structure of the shipping industry on the river.

2. Ship Standardisation Overview

The Chinese government, particularly the Chang Jiang (or Yangtze) Administration of Navigational Affairs (CJHY), which is part of the Ministry of Communications (MOC), has been trying to standardise inland vessels for over thirty years [2]. The National Program for Ship Standardisation [2], distinguishes three phases: 1975-1980, which was the 'initial stage', in which standardization was gradually introduced; 1980-2000, the 'development phase', in which ship models were selected (some 200 out of 2000 different ship types and sizes models) and optimized; and 2001-the present. This third phase is characterized by the aim to make 'a comprehensive push forward', through the development of specific regulation that should remove substandard tonnage from the market and promote the construction of new ships according to the standards. This regulation specifically targets very small and old fashioned ships made of wood or concrete, and ships with obsolete propulsion systems.

In 2003, a pilot project was launched for the Beijing-Hangzhou Grand Canal, where ships can now only be built according to the standard drawings that are provided by the Grand Canal Adminstration of Navigational Affairs.

Two other river systems, the Heilongjiang system and the Pearl River system also have full standardisation programs (see, for instance, stand3), but the main focus of the Chinese Authorities seems to be on the Yangtze River. A document on standardisation for the Yangtze River around the Three Gorges Dam (where it is called Chuan Jiang) has been issued [6], and a document for the lower part of the river is still 'under discussion', despite various planned release dates in 2006 and 2007. The lower part of the river is also accessible for short sea and ocean vessels. The river port of Nanjing, for instance, can receive vessels up to 50,000 tons deadweight, and all river ports below Nanjing can and do receive foreign vessels. Standardisation in the Yangtze River Delta beyond Nanjing would therefore be difficult to enforce, and the expected benefits in terms of infrastructure utilisation would probably not emerge as a result of standardisation (but rather as a result of the higher level of traffic).

The standards are developed for passenger, ro-ro, general cargo, chemical, oil, container vessels and (non-self propelled) barges. We report the standards for container vessels.

Table 1: Container ship standards upper reach of the Yangtze River

Standard name	*LOA*	*BOA*	Draught	TEU	Speed
50	62-64	10.8	2.0-2.4	45-55	≥ 20
60	67-70	13	2.0-2.6	60-70	≥ 20
100	72-75	13	2.6-3.0	90-110	≥ 20
150	85-90	13.6	2.8-3.2	120-157	≥ 20
200-I	85-90	14.8	2.8-3.2	135-170	≥ 20
200-II	85-90	16.2	3.0-3.5	150-200	≥ 20
250	105-112	16.2	3.5-4.0	240-260	≥ 20
300	105-112	17.2	3.5-4.0	260-310	≥ 20

'LOA' stands for length overall (meters), 'BOA' stands for breadth overall (in meters), Draught is in meters, TEU stands for twenty foot equivalent unit carrying capacity, speed is measured in kilometers per hour.

Note that the standards in the table are labelled with a round number that is somewhere in the TEU range. This highlights the importance of the TEU capacity as one of the main characteristics in the standards. Furthermore, the standards overlap in most of the categories, except when the four size categories are taken together. In terms of TEU, the standards labelled 150, 200-I and 200-II all have the TEU range 150-157 in common. It is only the width variable that prevents some ships fitting into two or three categories.

The importance of the standards for the Three Gorges Dam area can be illustrated with the following information on the productivity of the lock system in the Dam. According to Hou, there is a substantial gap between the design productivity of the lock system, and the actual performance. One of the main reasons is that, in the design, a much higher degree of standardisation of ships was assumed than is actually the case nowadays (see table 2). Partly as a result of this, the capacity of the lock system in the Dam is not nearly what it should be.

Table 2: Dimensions of the Three Gorges Dam shiplock system

Variable	*Design parameters*	*Actual*
Vessel number per lock lift	4x3000 ton vessels	8-9 vessels of different types
Width of lowest lock head	34 m	29 m
Water level in operation	175 m	135-139 m
Time between two lock lifts	59.7 min	110 min
Nr. of lock lifts per day	22.1	12

Source: Hou

'M' stands for meter, 'min' stands for minute.

From table 2, one can observe that the problem is mainly the diversity of ships in the lock chambers, leading to loss of space and inefficient entry and exit of the lock chambers. The CJHY hope that with the reinforcement of the standards, and with the active barring of non-standard ships from traversing the locks, the performance of the locks in the dam will be enhanced. The Dam is thus the main element in enforcing the standards and the main focus for the benefits of standardisation. In other parts of the river, the only other way of enforcing the standards is to prevent non-standard ships from entering ports, which is not the jurisdiction of the CJHY, but of local Port Administration Bureaus.

The final element in the standardisation program for inland vessels is a set of quite specific targets for the standardisation effort. In a recent interview [4], these targets are specified by river system and ship type, and for the years 2010, 2015 and 2020. The last year coincides with the planning horizon for the Yangtze River modernization program in the current five year plan.

The targets are specified for the upper reach of the Yangtze River main fairway with the added remark that owners in the lower reaches of the river will build ships according to market demand and should seek 'guidance from the government'. There are separate targets for the main Yangtze tributaries. Table 3 contains the targets for container ships.

Table 3: Targets Yangtze River

	Year		
	2010	*2015*	*2020*
Percentage ships in standard	75%	85%	95%
Average size (tons)	1000	1200	1500

Source: Zhang (2007)

Size of ships is measured in net cargo tons. This is comparable to the deadweight tons measure of size.

We will investigate the feasibility of these targets for the Yangtze River as a whole, and for the upper reach (Chuan Jiang and Three Gorges Dam area) separately as well as the impact of the standardisation program on the structure of the fleet.

3. Approach and Data Availability

This section introduces the simulation approach that we use to estimate the fleet growth over time and the construction of the dataset that was used to perform the calculations.

3.1. Approach

We suggest a simple dynamic fleet simulation that can be based on the following stock equation:

$$K_t = K_{t-1} + B_t - S_t, \tag{1}$$

where K_t is the fleet at time t, B_t represent newbuildings of ships that are delivered in period t and S_t are scrappings in period t. This equation features in many shipping industry models, see, for instance, the annex in Stopford97. It is also the basis for the supply side of the model in engelen.

We divide the fleet K into size classes based on the standards, in such a way that the size classes do not overlap. The standards may thus sometimes be made up of several size classes, depending on which dimension (length, width, draught, size in TEU) is looked at. The gaps between the standards may also contain ships, so we include these gaps as separate classes as well. We define F as the index set of all size classes.

Usually, newbuildings and scrappings are defined as a percentage of the fleet in the previous period. In our case, we get:

$$B_{t,i} = \begin{cases} b_i \cdot K_{t-1,i} & \forall i \in I \subset F \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

and

$$S_t = K_{t-1}^{35}, \tag{3}$$

where I is the index set of the size classes that make up the standards, b_i represents a percentage that may vary per size category. K^{35} represents the number of ships in the fleet that are over 35 years old. Equation (2) implies that there will only be newbuilding in the classes that coincide with the standards. Scrapping will be across the board. For scrapping, we assume mandatory scrapping for all vessels of 35 years old, regardless if they are in a standard or not [3].

A flat newbuilding rate as in (2) will lead to an explosive fleet growth, especially with the current level of growth on the Yangtze River (20-25% per year). We therefore assume that the growth rate will level off as follows:

$$B_{t,i} = b_i \cdot (1 - r_i)^{t-2007} \cdot K_{t-1,i}, \tag{4}$$

where r_i is the rate of decline for size class i, and $t = 2007K\ 2020$.

On the basis of this simulation approach, we can calculate the percentage of the fleet that fits in the standard categories at t=2010, t=2015 and t=2020. In addition, we can also calculate the average ship size in TEU, average age and a spread or concentration indicator for the fleet.

3.2. Data Availability

This simulation requires the following data:

1. Ship characteristics for the Yangtze container ship fleet: size in TEU,
2. An age profile of the Yangtze River container ship fleet,
3. Assumptions on newbuilding (i.e. addition of new ships) and scrapping (i.e. removal of old ships from the fleet).
4. A division in ships that operate in the upper reach, and those that do not.

Unfortunately, the data on the Yangtze River fleet are not generally available in the detail that we require. General data on the total number of ships by ship type are available with a considerable time lag from the Yangtze River Statistical Yearbook [7]. These data do not give any detail on individual ship dimensions or age.

We therefore take the following approach. We collected information on active container ships on the Yangtze from published service schedules. This provides information on the size in TEU on around 300 container ships in 2006, and on the part of the river in which they operate. Industry experts (see the reports in ndrc) estimate that in 2006, there should be around 350 ships in total, while in 2005, there were 277 [7]. We therefore obtained a sample of about 85%, which represents most of the ships that operate on the trunk line between Chongqing and Shanghai. These are relatively bigger and newer vessels. Our sample is therefore probably somewhat biased in size and age.

The assumption we make is not unreasonable. Several Chinese shipping companies are active in ocean or coastal trades as well as river trades (COSCO, Changjiang Shipping, China Shipping), and developed their fleets in these various trades more or less according to the same timetable. The river trades are closely related to the coastal and ocean trades, because they are part of the same logistics chains that carry cargo from inland locations to overseas destinations and vice versa. veenzhang finds that most container traffic on the Yangtze River is to and from Shanghai. One would therefore expect a considerable similarity between the ocean going, coastal and river fleets in terms of age profile.

The information in the service schedules does not allow us to collect the ships' age. However, we do have detailed information from the China Classification Society (CCS) on ocean going, coastal and some inland ships from the China Register of Ships [1]. We now make the important assumption that the age profile of container and general cargo vessels in this register is similar to the age profile of the Yangtze River container fleet. This set contains 316 ships (205 container vessels, and 111 general cargo vessels). Of these vessels, 174 (55%) have a container carrying capacity smaller than 500 TEU. These ships should therefore, in theory, be able to sail on the Yangtze River. The general cargo vessels that were included all had container carrying capacity. General cargo vessels in the Register without container carrying capacity were not included.

We compute the discrete empirical probability distribution of the age profile of the container fleet that is in the Chinese Register of CCS, and use this distribution as the basis for a customised random number generator that attaches an age to each vessel in our data set according to the same empirical distribution. The age profile is depicted in figure 1. This data set now contains 302 vessels with ship size in TEU and ship age in years. Notice the uneven distribution of ages, and the clear building sprees in the early to mid 1980s and 1990s. Also note that the scrapping age of 35 years is not restrictive.

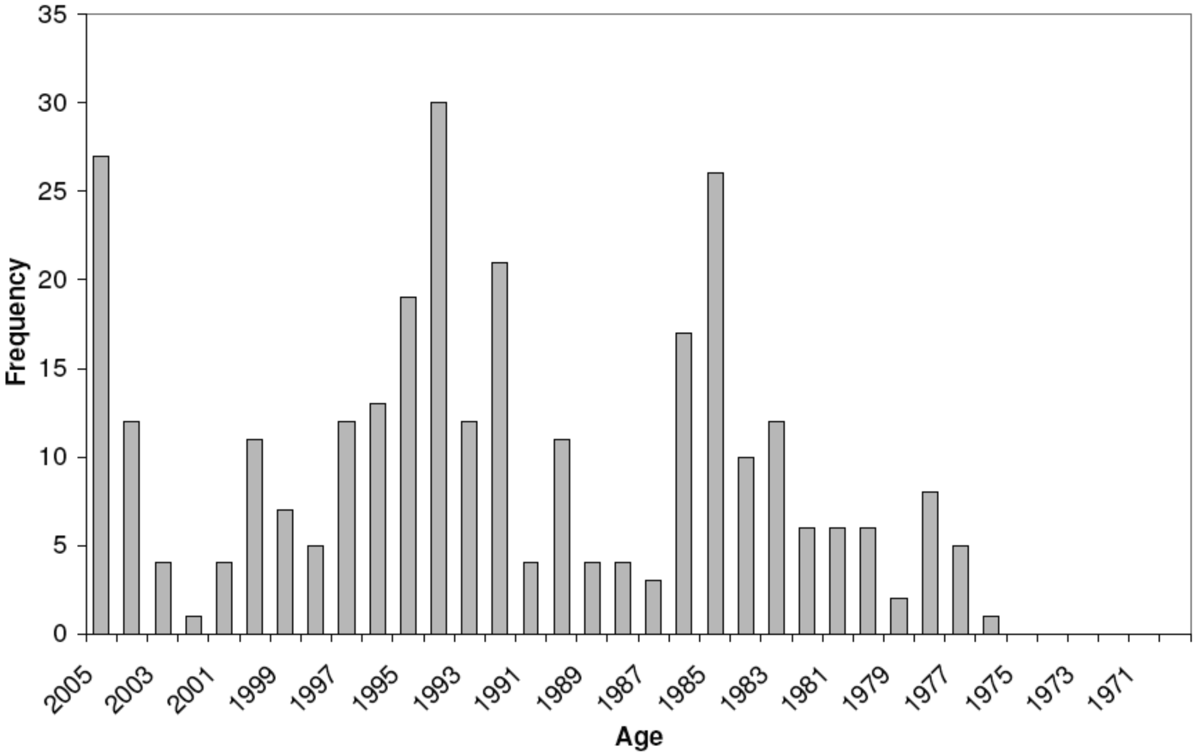


Figure 1: Age histogram

We make one further transformation in the data. Based on statements in the press [4], discussions with industry experts, reports [2], and field observations during various trips along the Yangtze River [9], it is commonly accepted that the small ships are relatively older than the bigger ships. This is not reflected in our data set. The correlation between ship size in TEU and age in the data set as constructed above is -0.02 (: 0). To make sure the data exhibit a positive correlation, we order the data set in ascending order by size, and switch all ages below 15 years of the ships in the bottom quartile of size (i.e. the smaller sizes) with all ages below 15 years in the top quartile of size. The age profile itself is invariant to this transformation, but the correlation between size and age now becomes 0.39.

The schedule data allow us to separate the 302 vessels in vessels operating in the upper reach and vessels that do not. The upper reach is commonly defined as the section between Luzhou and Yichang (the location of the Three Gorges Dam). It turns out that 119 vessels (39%) in our sample operate on the trades between Chongqing and neighbouring ports, and Shanghai. Some of these vessels' sizes are specific for the upper reach, and for the other sizes, we select ages at random from our full Yangtze River data set. This random selection means that, apart from the specific upper reach vessels, the upper reach fleet will get a similar age profile as the total river fleet.

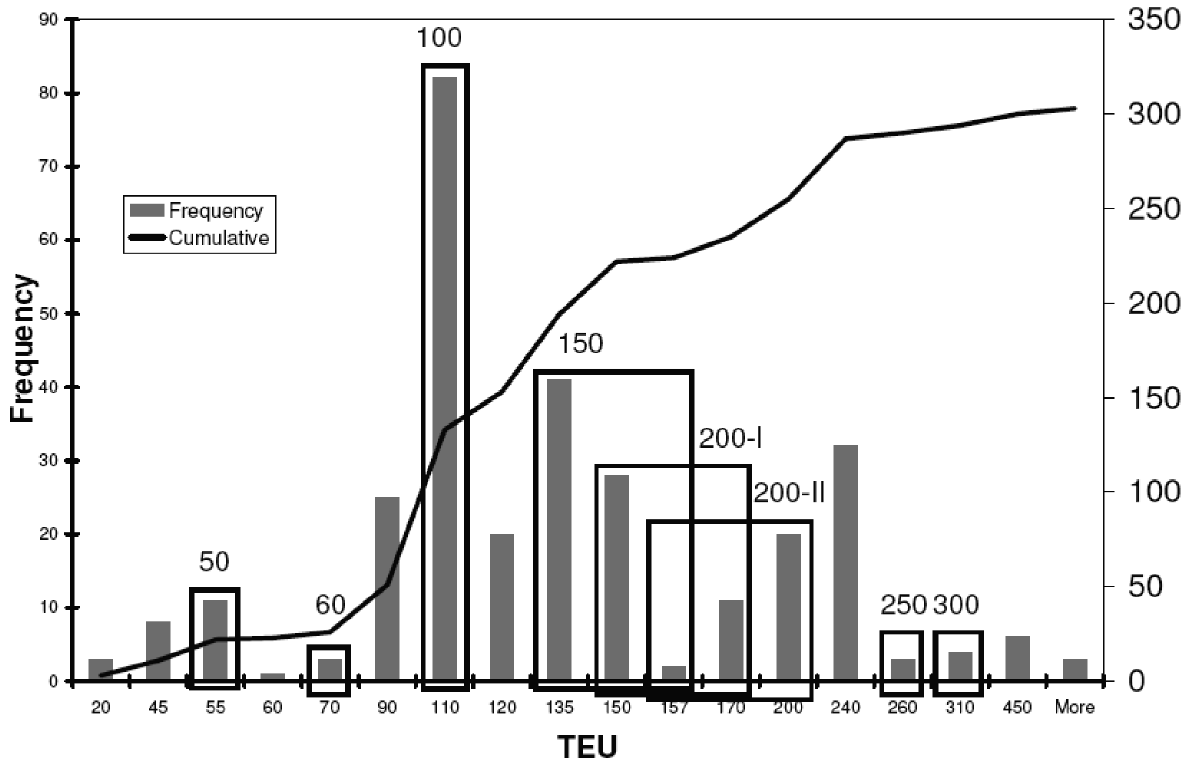


Figure 2: Size histogram (in TEU); entire river

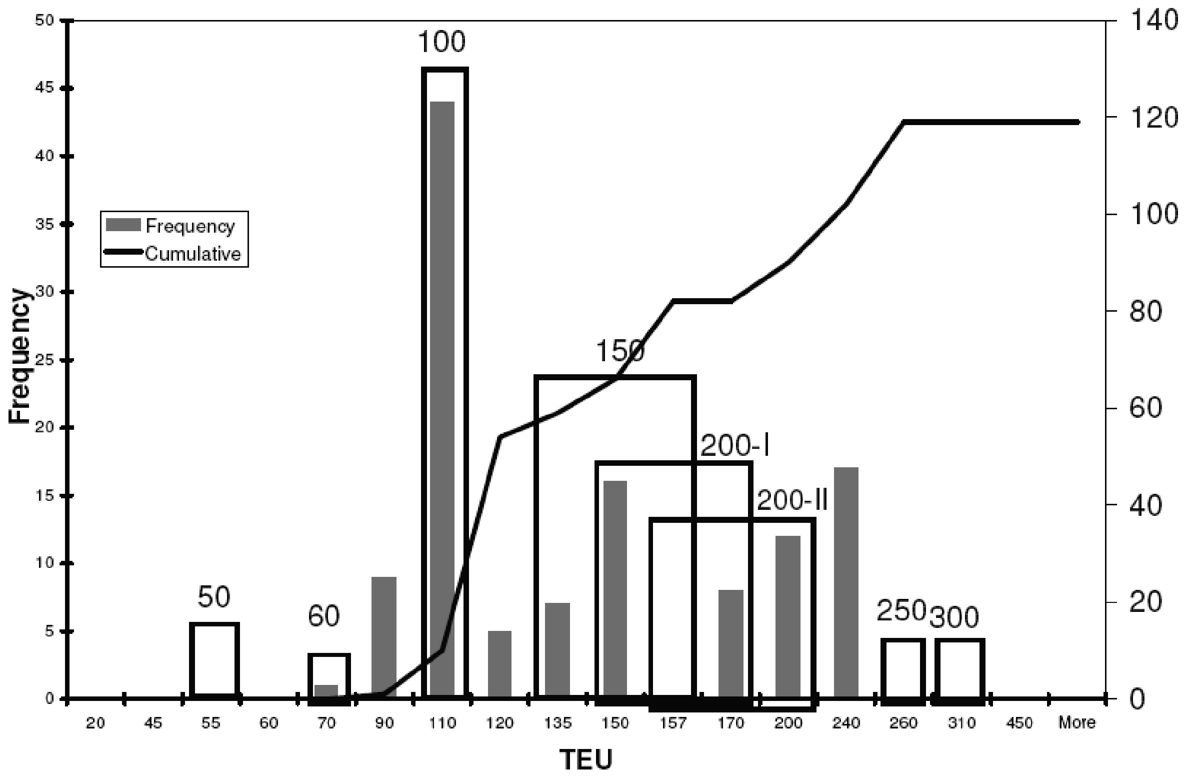


Figure 3: Size histogram (in TEU); upper reach

We report the size histograms for our data set in figures 2 and 3. We have drawn boxes around the size classes that make up the various standards. On the basis of this histogram, we find that for the river and the upper reach, 68% and 74% of the fleet already falls in one or more standards in 2006 (based on TEU). The upper reach percentage is slightly higher, which is due to some anticipating of the shipping companies of the standards coming into effect in the last few years.

Note that the larger ship size categories do not fall into a standard yet. These larger ships are being built, however, because they can be used to reach the new Yangshan Deepwater Port of Shanghai. We will therefore assume that all vessels larger than 310 TEU will fall in a newly developed standard in the future. This only applies to the river as a whole, since the size distribution for the upper reach is narrower than for the river as a whole. The smaller and larger vessels at the extremes of the distribution do not appear in the upper reach.

The data and figures presented sofar allow two general observations on the ship standardisation program: (1) the main aim to remove relatively old and small ships from the fleet does not really apply to container shipping on the Yangtze: there are some small ships, but there are also small ports and tributary rivers, and most of the ships are not old. There are no ships older the 35 years, and most ships (93%) are in fact 25 years of age or younger; and (2) the set of standards covers a large part of the existing fleet. Enforcing this set of standards will change the structure of the fleet only little, and will probably not reduce the diversity that is currently hampering the performance of the locks in the Three Gorges Dam (see table 2).

We have constructed our data set from available data. There are a few benchmarks to evaluate the degree of reality of our data set. The first is the average age of the fleet of container vessels that is reported by various government agencies. The second and third are average size of vessels in TUE and in tons. Given that our data set does not contain independent ton figures per ship, this third benchmark is only included as an illustration. The benchmarks only exist for the river as a whole. The following data is available:

Table 4: Benchmarks

		Our data set	
	Benchmark values	Entire river	Upper reach
Average ship size (TEU)	96	139	136
Average ship (tons)	1540	2230	2181
Average age	14	13.9	13.3

The average ship size benchmarks originate from the Yangtze River Statistical Yearbook (2004 data) [7]. The average age is from [2]. The tonnage figure for our data set was calculated on the basis of the same ton/teu ratio as in the benchmark.

The average TEU figures confirm that our data set is biased upward in size, compared to the entire container fleet on the Yangtze River. In terms of average age, the data set quite accurately reflects the average age of the container fleet on the Yangtze River. This provides support for our approach to use the entire Chinese container carrying fleet as the source for the age data. The figures also show that the structure of the fleet, apart from the slightly narrower size distribution, seems to be rather uniform across the river.

Note also that the average size of ships is already 1540 tons. Therefore, the tonnage targets in table 3 are automatically met.

4. Analysis

We follow the simulation approach using (1), (3) and (4) for the period 2007-2020. While our data set covers 302 ships, the total fleet in 2006 was around 350 ships. We choose not to extend our data set to reflect this last number, because the outcome of average size and %-fleet-in-standards are relative values. We make the assumption that the newbuilding percentage will be 20% for all size classes that are part of a standard and 0 elsewhere. In accordance with this, we also assume the decline percentage r is the same for all size classes.

To determine the decline in newbuilding, we consider the following. It seems reasonable to assume that the fleet growth on the Yangtze River will parallel the growth of the Port of Shanghai. This port is expected to grow to at least 45 mln TEU handled from the current 21 mln (2006) in the next decade or so. Currently, the inland traffic on the Yangtze River is about 10% of this [8], but this share could easily in the coming decade or so. This means that a growth rate of $45/21*2 = 4.3$ in 2007-2020 is a relatively conservative estimate for the Yangtze River container fleet. This means that in 2020 we expect to have a fleet of around 1300 ships. This amounts to a decline rate r = 6%. The newbuilding rate then reduces from 20% in 2007 to 8% in 2020.

The outcome is summarised in table 5.

Table 5: Summary of the simulation: base case

	Entire river					Upper reach			
	year					*year*			
	2007	2010	2015	2020		2007	2010	2015	2020
Total fleet	346	508	866	1289		137	203	351	527
Newbuilding	43	60	85	100		18	25	37	41
Scrapping	0	0	5	22		0	0	3	9
Average ship size (TEU)	139	140	142	144		135	132	130	130
% in standard	74	84	90	94.5		77.3	84.7	93.0	94.9
Average age	13.2			9.5		12.6			9.6
Fleet concentration	0.084			0.137		0.156			0.178

All figures are numbers of ships, unless otherwise indicated. The newbuilding and scrapping numbers are for the years mentioned. The fleet concentration is measured with the normalised Herfindahl index, $H = \frac{\sum_i^n s_i - 1/n}{1 - 1/n}$, where s_i is the share of the n^{th} ship size class in the total number of ships. $H = 0$ indicates a lack of concentration.

From table 5 it is clear that the standardisation targets are met in 2010 and 2015, but not in 2020 (if only just). If the simulation is run without the assumption that newbuilding will take place in the entire fleet, the percentage of the fleet that fits in the standards is constant at around 70%, but average age will also increase in a similar manner as in the table. The progression of the numbers in tabel 5 therefore illustrates that restricting newbuilding to the allowed standards is an effective but also a necessary strategy to achieve the standardisation targets, even if the ones for 2020 are just missed. Scrapping does not play a role for container shipping on the Yangtze, because the fleet is relatively young (given that containerised transport only emerged in China in the 1980s). On average the scrapping percentage in 2007-2020 is around 2%. Alternative calculations with age drawn from a uniform distribution results in very similar outcomes as reported in table 5. This also means that the specific age profile, in the container ship case, plays a negligible role.

Note that the average size of ships increases only very little for the entire river, and actually falls for the upper reach. The latter is the result of shape of the size distribution in the upper reach. This means that the envisaged growth of average ship size, which is an integral component of the standardisation program and is envisaged in the targets, will not materialise in the upper reach of the river. Furthermore, the concentration of the fleet increases but only marginally. The fleet composition therefore also changes very little.

The simulation has several parameters that can be changed: the scrapping policy (35 years), the newbuilding percentage, the decline of the newbuilding percentage over time and the assumption that the share in the modal split of inland shipping in the hinterland traffic to Shanghai will double. The sensitivity analysis will be performed for two tipping points: reaching the 2020 target, or failing the 2015 target. The following table summarises the required changes in the variables.

Table 6: Summary of the simulation: sensitivity

	Entire river				upper reach		
variable	value	reaches target	fails target		value	reaches target	fails target
		2020	2015			2020	2015
Scrapping	26 years	*			34 years	*	
Scrapping	*		×		*		×
b	22%	*			21%	*	
b	12%		*		12%		*
r	4.5%	*			5.5%	*	
r	20%		*		*		×
ms share	23%	*			23%	*	
ms share	10%		*		7%		*

'MS share stands for modal split share. '×' means that the target could not be failed or reached. The change in modal split share is achieved by simultaneously lowering the newbuilding rate and raising the newbuilding decline percentage (or vice versa). The results are reported for the combination 21%/5.5% (=share of 23%) and 14%/8%(= share remains stable at 10%) for the entire river, and 20.5%/5.5% (23%) and 11%/12% (7%) for the upper reach.

Table 6 shows that only slight changes are required to reach the target for 2020. However, very drastic changes are required to fail the target of 2015. In the cases where the 2015 target failed, the final fleet size was around 500-560, which means that the fleet did not match the expected growth of the Port of Shanghai. In other words, only in very particular circumstances will the Yangtze River standardisation effort fail the 2015 target, let alone the 2010 target.

From this sensitivity analysis, we can conclude that the degree to which the targets will be met is largely based on the overall growth of the fleet from now to 2020. If the fleet grows in line with the Port of Shanghai, and no newbuilding will be allowed outside the standards, the targets will be met. If the fleet will not match the growth of Shanghai, targets will not be met, but this will also be a sign of a more important systemic problem in the container transport system on the Yangtze River.

5. Concluding Remarks

We present an analysis of the feasibility of the Chinese inland shipping standardisation program on the Yangtze River. This program is under development, and currently only implemented around the Three Gorges Dam in the upper reach of the Yangtze River. For this section of the river, there are specific standardisation targets for the years 2010, 2015 and 2020.

Not all data required for the current analysis are in the public domain. For the purpose of the current analyse, we use a constructed data set that is an amalgam of collected data and information available from the China Classification Society. The data set does seem to be a relatively realistic representation of the size and age profile of the Yangtze River Container fleet on the river as a whole. The analysis shows that the particular shape of the age profile does not influence the results in a crucial way.

We present a deterministic fleet simulation approach that is in line with current thinking about market structures in shipping. This approach has three important outcomes. First it shows that adherence to the restriction on newbuilding outside the standards is crucial for reaching the targets. The Three

Gorges Dam provides an effective tool in achieving this in the upper reach of the river. The companies in this part of the river cannot risk being prevented from using the lock system in the Three Gorges Dam, which is an effective way to enforce the standardisation program. In the lower part of the river, such a mechanism does not exist, and therefore, standardisation in this part of the river will be meaningless.

Secondly, the model shows that the likelihood of reaching the targets in 2010 and 2015 is quite robust to changes in input values, and depends mainly on the growth of the fleet. Furthermore, we found that the average ship size did not rise much and that an active scrapping policy or a tighter scrapping policy had a negligible effect on the outcomes.

Thirdly, we find that the current diversity of ships, as measured by a concentration index, will change very little.

On the basis of these results, we can conclude that the standardisation in the upper reach of the river will succeed according to the targets, but that the intended effects, such as increasing average ship size, increasing the performance of the lock system in the Dam and improving port utilisation, will probably not be realised.

Further research could focus on measuring the impact of the ongoing standardisation effort on the performance of the fleet, the river container ports and the performance of the Three Gorges Dam lock system. The analysis could also be extended to other river systems and other ship types.

Acknowledgements

The author would like to thank Zhang, Mo (Delft Technical University) and Kevin Lu of the Waterborne Transport Institute for assistance with the collection of the ship schedules.

References

[1] CCS. (2006), Register of Ships. China Classification Society, Beijing, China.

[2] CJHY. (2003) National Programme for the Development of Inland Ship Type Standardization (in Chinese). Notice, 56, Chang Jiang Administration of Navigational Affairs.

[3] CJHY. (2007), Report on developments in Yangtze River Intermodal Transport. Private communication during various NDRC meetings in June 2007.

[4] CJHY. (2006), The current development of ship standardization - interview with Zhang, Shou Guo (in Chinese).

[5] Huang, Q. (2007), Maximizing the Potential of The Yangtze Golden Waterway (in Chinese). Thesis collection, Yangtze Water Resources Commission.

[6] MOCa. (2004), Chuan Jiang and Three Gorges Dam Main Ship Type Standards (in Chinese). Announcement, 30, Ministry of Communication of China.

[7] YRHS. (2006), *Yangtze River Statistical Yearbook (in Chinese)*. China Government Publishing Company, Beijing, China.

[8] Zhang, M. (2007), Container Shipping Development on the Yangtze River. Master's thesis, Delft Technical University.

[9] van der Zwam, H. and Kleingeld, E. and de Leijer, H. and van der Leur, J. and Nieuwaal, M. and Veenstra, A.W. Review of Intermodal Transport Developments in the Yangtze River corridor. Technical report, Dutch Ministry of Economic Affairs/Agency for International Business and Cooperation, 2007. this report was written as part of an inspection trip under supervision of the Chinese National Development and Reform Committee. It can be obtained from the author upon request.

Maritime safety and security: developments and challenges

Jin Wang

Liverpool Logistics, Offshore and Marine (LOOM) Centre, Liverpool John Moores University, Liverpool, L3 3AF, UK.

Abstract

The maritime industry is moving from a largely prescriptive safety regime to a risk-based goal-setting one. Formal Safety Assessment (FSA) is a systemic and structured approach adopted by the International Maritime Organization (IMO) to support its rule-making process in terms of improving shipping safety. The approach has shown its superiority in identifying hazards, analysing risks and providing cost effective solutions when assessing regulations and rules governing maritime safety. It is also possible for the approach to be used to support risk-based design and operation of large maritime systems such as ships and sea ports. Thus, the approach has attracted great attention in the maritime industry over the past decade. More recently, there has been a growing international recognition that security issues of maritime systems such as container line supply chains need to be reviewed on an urgent basis. In this paper, following a brief review of shipping risk assessment, a FSA framework is outlined with the latest developments highlighted. A threat-based security risk concept is introduced and discussed. Major challenges and problems encountered in the application of maritime safety and security assessment are identified and described. Recommendations on future work are finally given.

Keywords: Accident, decision making, formal safety assessment, security assessment, shipping

1. Introduction

The safety-related rules have predominately been prescriptive, quite often derived as a reaction to major incidents at sea in order to prevent similar accidents from occurring again. For example, the capsize of the *Herald of Free Enterprise* in 1987 raised serious questions on operation requirements. This eventually resulted in the adoption of the International Safety Management (ISM) Code for the Safety Operations of Ships and for Pollution Prevention. The *Exxon Valdes* accident in 1989 seriously damaged the environment by a large scale oil spill. It facilitated the implementation of the international convention on Oil Pollution Preparedness, Response and Co-operation (OPRC) in 1990. Double hull or mid-deck structural requirements for new and existing oil tankers were subsequently applied (Sekimizu, 1997). The *Scandinavian Star* disaster in 1990 and the catastrophic disaster of the *Estonia* in 1994 highlighted the role of human error in marine casualties, and as a result, the new Standards for Training, Certificates and Watchkeeping (STCW) for seafarers were subsequently introduced (Wang, 2006).

Since 1990s, there has been a move from a largely prescriptive and reactive safety scheme to a risk-based proactive regime in the maritime industry. The responsibility for safety is being placed on those in the industry to set out and justify their basis for managing the risks. Such a change should create new perspectives on risk-based decision making. It is believed that maritime safety may be significantly improved by introducing a formal "goal-setting" safety assessment approach so that the challenge of new technologies and their application to maritime industry may be dealt with properly.

In 1992, Lord Carver's report on the investigation of the capsize of the *Herald of Free Enterprise* raised the issue of a more scientific approach to the ship safety subject and recommended that emphasis should be given to a performance-based regulatory approach (House of Lords, 1992). The

UK Maritime & Coastguard Agency (MCA) quickly responded and in 1993 proposed to the International Maritime Organization (IMO) that formal safety assessment (FSA) should be applied to ships to ensure a strategic oversight of safety and pollution prevention. Since then, substantial work including the demonstration of FSA's practicability by trial applications has been done.

More recently security issues associated with threat-based risks have attracted much public concern. This is in recognition of the increasing vulnerability of a ship which is an exposure to serious disturbances arising from threats. The recently implemented International Shipboard and Port Facility Security (ISPS) Code (IMO, 2002c) has also stimulated research in this important area.

2. Maritime Safety

2.1. Marine Accidents and FSA Development

The international safety-related marine regulations are guided by lessons learned from serious marine accidents that have happened. There have been quite a few shocking accidents since 1980s. Typical ones include the capsize of the *Herald* of *Free Enterprise*, the Estonia accident and the Prestige Tragedy.

The capsize of the *Herald of Free Enterprise* on the 6th March 1987 was caused by a combination of adverse factors. Those which have been identified were the trim by the bow, the bow door being left open, the speed of the vessel just before capsizing and the location of the ship's centre of gravity. Their combined effect was to cause a quantity of water to enter G-deck and loss of the vessel's stability (Wang, 2002; Wang and Trbojevic, 2007). The findings of the inquiry clearly demonstrated the contributions of human actions and decisions to the accident. These ranged from weakness in the management of safety to human errors, caused by various factors including a heavy workload. The basic Ro-Ro ferry design was questioned, in particular the single compartment standard for G-deck. There were no watertight bulkheads at all on this deck to prevent the free surface effect along the full length of the vessel. The public inquiry into the accident of the *Herald of Free Enterprise* led by Lord Carver resulted in changes of marine safety related regulations, demonstrated by the adoption of the enhanced damage stability and watertight closure provisions in SOLAS'90, the introduction of the ISM Code, and the development of the FSA framework in the shipping industry.

The passenger ferry *Estonia* sank in the northern Baltic Sea with the loss of 852 lives on the 28th September, 1994. The cause of the accident was that the design and manufacture of the bow visor locks were inappropriately conducted, resulting in the locks being too weak. During bad weather conditions the locks were broken and the visor fell off and pulled open the inner bow ramp (Wang, 2006). Water flooded the main Ro-Ro deck and the vessel lost its stability and sank. *Estonia*, at her last voyage, was not seaworthy and she did not fulfill the SOLAS requirements. The crew members also made mistakes, which partially contributed to the loss of so many lives. The *Estonia* tragedy resulted in a surge of research into the phenomenon of Ro-Ro damage survivability and was instrumental in the adoption of the North European regional damage stability standard in SOLAS'95 and the Stockholm Agreement. These standards require the upgrading of every passenger Ro-Ro ship operating in Northern Europe (Channel, North Sea, Irish Sea, and Baltic Sea).

The *Prestige* tanker carrying 77,000 tonne of heavy oil seriously polluted the Spanish coast by oil spills on the 19th November 2002. It shocked the public and focused attention on tanker safety. Under new rules adopted by the EU, single hull tankers carrying heavy oil have already been banned from EU ports. The ban brings the EU in line with the United States, which restricted single-hull tankers carrying heavy oil from its waters three years after the 1989 *Exxon Valdez* disaster. The EU's ban came a year after the *Prestige* tanker accident.

2.2. Origin of FSA

Following the tragedy of *Herald of Free Enterprise*, a FSA framework was proposed to the IMO in order to ensure a strategic oversight of safety and pollution prevention. The FSA framework consists of the following five steps (IMO, 1997b):

1. Identification of hazards.

2. Assessment of risks associated with those hazards.

3. Ways of managing the risks estimated.

4. Cost benefit analysis of the risk control options (RCOs).

5. Decisions on which options to select.

The above framework was initially studied at the IMO Maritime Safety Committee (MSC) meeting number 62 in May 1993. At the 65th meeting of the MSC in May 1995, strong support was received from the member countries, and a decision was taken to make the FSA a high-priority item on the MSC's agenda. Accordingly, the UK decided to embark on a major series of research projects to further develop an appropriate framework and to conduct a trial application on the selected subject of high-speed passenger catamaran ferries. The framework produced was delivered to MSC number 66 in May 1996, with the trial application programmed for delivery to MSC number 68 in May 1997. An international FSA working group was formulated at MSC number 66 and MSC number 67 where draft international guidelines were generated. The IMO eventually approved the application of the FSA for supporting the rule-making process in 2002 (Wang, 2006). The application of the framework in ship design and operation may:

1. Provide improved performance of the current fleet and then be able to measure the performance change so ensuring that new ships are of good design.

2. Ensure that experience from the operational field is used in the current fleet, and that any lessons learnt are incorporated into new ships.

3. Provide a mechanism for predicting and controlling the most likely scenarios that could result in incidents.

Ship safety is also significantly driven by classification. Current classification society rules have evolved over many years and have been mainly developed on an empirical basis (Card et al., 2004). The basis of the rules is not always transparent to the users. There have been many calls from the maritime industry for classification societies to adopt an approach that would lead to the development of rules that are more easily understood and based on clearly identifiable scientific principles (Card et al., 2004). Recently, there have been positive developments on this aspect led by some leading classification societies such as the ABS, DNV and Lloyds Register (LR). For example, to meet the expectations of the marine industry and to make use of the best standards practice, ABS, DNV and LR have jointly developed a new set of classification rules for oil tankers that would provide, through transparency, a better understanding of the design principles underpinning the rules (Wang, 2006).

In recent years, many test cases of FSA have been conducted to improve and enhance the application of ship safety. Those include a trial study on high-speed craft (IMO, 1997a), a trial study on bulk carrier (IMO, 1998a, 2002a and 2002b), a trial study on passenger Ro-Ro vessels with dangerous goods (IMO, 1998b), its application to fishing vessels (Pillay, 2001), its application to offshore support vessels (Sii, 2001), its application to cruise ships (Lois et al., 2004), its application to ports (Trbojevic, 2002), its application to containerships (Wang and Foinikis, 2001), and its application to liner shipping (Yang, 2007). It should be pointed out that the above only constitutes an incomplete list of investigations selected. Researchers worldwide have investigated FSA and its application to maritime systems such as ballast tanks, helicopter landing areas, oil tankers, floating, processing, storage and offloading vessels, etc. There are many advanced risk modelling and decision making techniques developed/being developed, that may be applied to facilitate maritime risk-based design and operation (Wang, 2006).

2.3. FSA Framework

FSA is a new approach to maritime safety which involves using the techniques of risk and cost-benefit assessment to assist in the process of decision making. It has been noted that many leading classification societies have been moving towards a risk-based regime. It is believed that the framework of FSA can facilitate such a move. This may be mainly due to the fact that FSA involves much more scientific aspects than previous conventions. According to the benefits identified by UK MCA (MCA, 1993), adopting FSA as a regulatory tool will effectively address all aspects of safety in an integrated way, present a pro-active approach, enabling hazards that have not yet given rise to accidents to be properly considered and provide a rational basis for addressing new risks posed by ever-changing marine technology.

Step 1 of FSA (hazard identification) identifies and generates a selected list of hazards specific to the problem under review. Hazard identification is concerned with using the “brainstorming” technique involving trained and experienced personnel to determine the hazards. The accident categories include contact or collision, explosion, external hazards, fire, flooding, grounding or stranding, hazardous substance related failure, loss of hull integrity, machinery failure and loading and unloading related failure. Various scientific safety assessment approaches such as HAZard and Operability (HAZOP) study, can be applied in this step. Once the hazards are identified with respect to each of above accident categories, significant risks can be selected using screening techniques for further analysis.

Step 2 of FSA (risk estimation) aims at estimating risks and factors influencing the level of safety. The assessment of risks involves studying how hazardous events or states develop and interact to cause an accident. Shipping consists of a sequence of distinct phases between which the status of ship functions changes. The major phases include design, construction and commissioning, entering port, berthing, unberthing and leaving port, loading and unloading, dry docking and decommissioning and disposal. A ship consists of a set of systems such as machinery, control system, electrical system, communication system, navigation system, piping and pumping system and pressure plant. A serious failure of a system may cause disastrous consequences. Risk estimation may be carried out with respect to each phase of shipping and each such system. The occurrence likelihood of each failure event and its possible consequences can be assessed using a variety of safety assessment techniques.

Step 3 of FSA (identification of RCOs) aims at proposing effective and practical RCOs. High risk areas can be identified from the information produced in Step 2 and then the identification of risk control measures (RCMs) can be initiated. In general, RCMs have a range of the attributes including those relating to the fundamental type of risk reduction (i.e. preventative or mitigating), those relating to the type of action required and therefore to the costs of the action (i.e. engineering or procedural), and those relating to the confidence that can be placed in the measure (i.e. active or passive, single or redundant).

Step 4 of FSA (cost benefit analysis) aims at identifying benefits from reduced risks and costs associated with the implementation of each RCO for comparisons. To conduct cost-benefit assessment, it is required to set a base case that can be used as a reference for comparisons. A base case is the baseline for analysis reflecting the existing situation and what actually happens rather than what is supposed to happen. The costs and benefits associated with each option can be estimated.

Step 5 of FSA (decision making) aims at making decisions and giving recommendations for safety improvement taking into consideration the findings obtained. The information generated in the previous steps can be used to assist in the choice of cost-effective and equitable changes and to select the best RCO.

3. Maritime Security

The tragedy of September 11, 2001 had stimulated the concern in the international shipping society that a terrorism organization capable of the suicide hijackings of airlines could readily adapt these capabilities to major shipping targets. Under such an environment, a series of anti-terrorism maritime security measures and strategies has been urgently developed and applied to realistic shipping operations. The concept of maritime security assessment has been proposed and widely accepted to be the basis of developing appropriate security measures in this process. The recently implemented ISPS Code (IMO, 2002c) requires security assessment for various ship and port facility security plans. However, apart from its Section 8 in Parts A and B, the Code does not prescribe a generally accepted methodology to carry out such assessment. Although Section 8 in Part B provides a number of issues to be considered when a security assessment is carried out, an obvious problem involved is that Part B is not mandatory and this may leave maritime stakeholders to choose and define their own "suitable" methodologies and guidelines for individual maritime security assessment. For example, the American Bureau of Shipping (ABS) or Lloyd's Register favours the risk assessment guidelines provided by the United States Coast Guard (USCG), while Det Norske Veritas (DNV) and Germanischer Lloyd (GL) have developed their own guidelines based on checklists which have a close relationship to the ISPS Code. The USCG guidelines do not include any statements about likelihood of security threats, whereas the DNV-GL approach allows for a consideration of likely threats only (Schroder et al., 2006). While the users of the DNV-GL approach have to update their security assessment frequently depending on the latest security information available, the USCG approach requires the development of mitigation strategies and clear identification of the best option(s) from costly risk control measures. Furthermore, US DOHS (Department of Homeland Security) led the Federal effort to develop a comprehensive National Strategy for Maritime Security, to best blend public and private maritime security activities on a global scale into an integrated effort that addresses all maritime threats in 2005 (DOHS, 2005). In addition to this strategy, eight supporting plans have been developed to address the specific threats and challenges of the maritime environment. In line with this strategy, both qualitative and quantitative security assessment and decision support tools such as On-shore and Off-shore Security Assessment (Emerson and Nadeau, 2003), Maritime Security Risk Analysis Model (MSRAM) and MSRAM-PLUS (Adler and Fuller, 2007) have been presented to enrich the concept of layered defense and to quantify maritime threats, vulnerabilities and consequences of attacks. However, the above studies that have contributed to maritime security decision making also trigger a serious concern of security analysts in dealing with the difficulties of precisely analyzing the likelihood of the threats encountered.

Safety and security assessment is a process of analyzing both threats and hazards and making the respective decisions. Over the past several years there has been a growing international recognition that safety and security issues of maritime systems such as Container Line Supply Chains (CLSCs) need to be reviewed on an urgent basis. Both public and political authorities have become aware of the situation where the evolution of liner shipping from the original general-cargo liner service to the current CLSCs, together with the increasing dependency of the world economy on them, has dramatically increased its risk stake and categories. The definitions of the risks that exist in the CLSCs have changed and broadened forever. This has stimulated research in vulnerability and risk studies in the supply chain context.

CLSCs, with many complex physical and information flows, have contributed themselves to economic prosperity and also rendered themselves uniquely vulnerable to many possible undesirable events ranging from delay of cargo delivery to environmental pollution and from terrorist attacks to damage of economic stability. Security is becoming one of the most important criteria for measuring the performance of the design, control and management of maritime systems. Depending on the nature of these processes, the most predictable relationships among different security and risk variables may emerge at a variety of spatial, temporal or functional scales. Therefore, current knowledge might be better represented if each relationship were described at or between the dynamic and interactive levels of detail at which the key safety and security variables could be identified, rather than at a static and steady scale that is identical for all processes.

It is inevitable that information is typically obtained incrementally in safety and security assessment of CLSCs. The inherent uncertainty in such information can be caused by imperfect understanding of the domain of a CLSC, incomplete knowledge of the state of the domain at the time when a given task is to be performed, randomness in the mechanisms governing the behavior of the domain, or a combination of them. It is necessary to model the assessment domain so that the measure of each event becomes more reliable in light of new information received.

4. Use of Advances in Technology for Facilitating Maritime Safety and Security Assessment

Lack of reliable failure data and lack of confidence in safety and security assessment have been two major challenges in safety and security assessment of various maritime activities (Wang, 2001). Under this circumstance, one of major concerns of maritime safety and security analysts is to explore and exploit flexible and advanced risk modelling and decision making approaches using uncertainty treatment methods for producing detailed guidelines to facilitate the practical application of safety and security assessment.

In recent years, many research activities have taken place to tackle the uncertainties in maritime safety and security assessment. In the context of FSA and its application, research has been carried out using the methods based on fuzzy possibilistic, Bayesian probabilistic, Dempster-Shafer theories and their combination, depending on the availability and completeness of failure data, the level of the analysis required and the degree of complexity of the interrelationships of risk factors. Many relevant studies and their results have been observed and reported, including:

1. Approximate reasoning approach for dealing with risk analysis problems associated with a high level of uncertainty (Wang et al., 1995, 1996). It has been extended to include evidential reasoning techniques, fuzzy set modelling methods, and the Dempster-Shafer method for risk modelling and decision making (Sii, 2001; Pillay and Wang, 2003).

2. Belief fuzzy rule-based evidential reasoning approach for risk estimation (Liu et al., 2004; Yang et al., 2009). It has been extended to develop belief fuzzy link-based evidential reasoning method for safety-based multiple criteria decision making and to combine with a Bayesian reasoning mechanism for the generation of belief fuzzy rule-based Bayesian reasoning approach for both risk/security modelling and decision making (Yang et al., 2007).

3. Application of Bayesian network (BN) approach. It includes the use of traditional BN and fuzzy BN approaches for risk diagnosis and prediction (Eleye-Datubo et al., 2006; Eleye-Datubo, 2006) and the development of the hybrid of BN, fuzzy sets and entropy theory to handle multiple dynamic risk attribute decision making (Yang, 2007).

The above is only a partial list that has been investigated by some selected researchers, and there are more techniques (i.e. Monte Carlo simulation, artificial neural network, fuzzy fault tree analysis, fuzzy TOPSIS and analytical hierarchy process, etc.) in general engineering and technology that have been/may be applied to facilitate the uncertainty treatment of risk modelling and decision making in FSA of ship design and operation.

Security assessment research has also been conducted actively over the past several years. The following research developments have been noted:

1. Use of subjective judgments and approximate synthesis techniques for modelling vulnerability and security based decision making (Yang, 2007).

2. Use of BN for security estimates of CLSC systems (Yang, 2007).

3. Use of fuzzy rule base, artificial neural networks, analytical hierarchy process, multiple criteria decision making and Six Sigma method for security studies of port operations (Ung, 2007).

Apart from the uncertainty treatment, several research topics that have been investigated include the introduction of individual ship safety cases, the collection and analysis of failure data, the definition of risk acceptance criteria, the consideration of human error, the standardization of subjective judgment through the employment of effective expert assessment procedures, and the exploration of appropriate computing software. It is believed that through further research into such topic areas, flexible and effective measures will be developed and applied towards the establishment of a safer shipping community.

5. Discussions and Recommendations

Although showing much attractiveness, maritime safety and security assessment still has a large space for improvement on their applications. Before their full potential is realised, there are areas where further studies are necessary.

1. FSA for ships should develop the ability to interact with regulatory bodies responsible for port and land-based operations, especially with the fast development of container-based multi-model transportation. Sharing the relevant data of non-compliance with established safety and quality standards for the relevant industries would eliminate a considerable percentage of the uncertainty created in this direction.

2. Acceptable criteria need to be developed to determine if risks are acceptable, unacceptable or need to be reduced to an as low as reasonably practicable (ALARP) level. Large variations exist in risk acceptance criteria in different industries by individual administrations. When quantitative risk assessment (QRA) is performed, quantitative risk acceptance criteria are normally required. However, it may not always be appropriate to use numerical risk acceptance criteria as inflexible rules given uncertainties involved in risk assessment. In general, there are no quantitative risk acceptance criteria in FSA at the moment although the MCA's trial applications have used QRA to a certain extent (Wang, 2002).

3. It becomes apparent that FSA's success largely depends on two essential conditions. The first condition is the development of a safety culture at all levels of the industry's infrastructure, from company managers to vessel operators. The second one is the development of further guidance on how human factors would be integrated into the FSA framework in a feasible manner. To a large extent, human error can be identified, dealt with and reduced by the application of effective human reliability methods to the FSA framework and adequate training of crews.

4. In order to overcome the problems related to the availability and reliability of failure data, international co-operation and co-ordination are required with the intention that a new global database will be established, controlled and updated by an international regulatory body (i.e. the IMO). Such a database should be easily accessible by both administrations and analysts/researchers providing reliable data with defined parameters upon which the incoming information has been processed. As far as expert judgments used to compensate the lack of failure data are concerned (a) scientific procedures that standardize a human being's performance (b) analytical techniques that enable the reasonable aggregation of the estimations from multiple experts and (c) logical approaches that maximally reduce their own limitation to the presentation of experts' judgments, should always be employed to support effective, consistent and informative decision making thus avoiding making costly inappropriate decisions.

5. Cost benefit analysis can be unexpectedly complicated, especially in cases where human life values are involved. Furthermore, the use of different values on different nationalities would have an adverse and undesirable effect on both international relations and working conditions onboard ships. A feasible solution to this problem would, once more, involve an international agreement on a reliable method of estimating the current value of human life. The international regulatory bodies should not only be responsible for the initial deliberations, but also for the constant follow up of the international economic, political and social trends which influence that value.

6. Maritime security study as a new area of research faces many challenges, Firstly there is a significant lack of statistical data and the level of uncertainty in the available data is usually unacceptably high given that threats have their own unique characteristics. Secondly, many maritime systems such as CLSCs involve cargo flow and information flow with many nodes in the supply chain process. Thirdly there are no clearly stated acceptance criteria for security levels and significant subjective judgments are necessary. A systematic security assessment methodology for international shipping is highly necessary.

6. Conclusion

The FSA philosophy has been approved by the IMO for reviewing the current safety and environmental protection regulations and justifying a new element proposal to the IMO by an individual administration. This paper has described the current FSA methodology in a five steps framework. Maritime safety and security assessment has also been described together with the current status of its development and challenges. Both safety and security studies are attracting great interests from many maritime stakeholders. Further research on both risk and security studies is highly necessary. Use of advances in technology may significantly facilitate the application of safety and security assessment through the development of flexible and novel tools.

7. Acknowledgements

This research is funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under Grant References: GR/F024436/01 and GR/F024606/01.

References

Adler, R.M. and Fuller, J. (2007), An integrated framework for assessing and mitigating risks to maritime critical infrastructure, Processing IEEE Conference on Technologies for Homeland Security, May 16-17, Woburn, USA.

Card, J., Haugland, B.K. and Pomeroy, V. (2004), Developing the next generation of classification rules for oil tankers, Proceedings of the International Conference on Design and Operation of Double Hull Tankers, the Royal Institution of Naval Architects, London.

DOHS (2005), The national strategy for maritime security, USA Government Publication, available at (http://www.whitehouse.gov/homeland/maritime-security.html).

Eleye-Datubo, A.G. (2006), Integrating Risk-based assessment modelling of safety-critical marine and offshore applications, PhD Thesis, Liverpool John Moores University, UK.

Eleye-Datubo, A. G., Wall, A., Saajedi, A. and Wang, J. (2006), Enabling a powerful marine and offshore decision-support solution through Bayesian network technique, Risk Analysis 26(3): 695-721.

Emerson, S.D. and Nadeau, J. (2003), A coastal perspective on security, Journal of Hazardous Materials 104(1): 4-11.

House of Lords (1992), Safety aspects of ship design and technology (Carver Report), Select Committee on Science and Technology, 2nd Report, HL Paper 30-I, February.

IMO (1997a), Formal Safety assessment: trial application to high speed passenger catamaran vessels, Final Report, DE 41/INF.7, submitted by International Maritime Organization (IMO) UK, IMO Sub-Committee on Ship Design and Equipment, 41st Session, Agenda Item 5, London, UK.

IMO (1997b), Interim guidelines for the application of formal safety assessment to the IMO rule-making process, IMO/MSC Circular 829, London, 17th November.

IMO (1998a), Bulk carrier safety, proposal for a formal safety assessment of bulk carriers, MSC70/4/Add1, submitted to the IMO by the UK MCA.

IMO (1998b), Trial application of formal safety assessment to dangerous goods on passenger/ro-ro vessels, MSC69/INF.24, submitted by International Maritime Organization (IMO), Finland.

IMO (2002a), International collaborative FSA study on bulk carriers - Step 2 of FSA (risk analysis) WP 11 — develop risk contribution tree components, MSC 75/INF.22, submitted by France to IMO, IMO, London, UK.

IMO (2002b), International collaborative FSA study - FSA Step 3 (risk control options), MSC 76/INF.8, submitted to the IMO by UK, IMO, London, UK.

IMO (2002c), International Ship and Port Facility Security Code, IMO Publication, SOLSA/CONF. 5/34, Dec. 17, London, UK.

Liu, J., Yang, J.B., Wang, J., Sii, H.S. and Wang, Y.M. (2004), Fuzzy rule-based evidential reasoning approach for safety analysis, International Journal of General Systems 23(2-3): 183-204.

Lois, P., Wang, J., Wall, A. and Ruxton, T. (2004), Formal safety assessment of cruise ships, Tourism Management 25: 93-109.

MCA (1993), Formal safety assessment, Submitted by the UK to IMO Maritime Safety Committee, Marine Safety Agency, MSC66/14.

Pillay, A. (2001), Formal safety assessment of fishing vessels, PhD Thesis, School of Engineering, Liverpool John Moores University, UK.

Pillay, A. and Wang, J. (2003), A risk ranking approach incorporating fuzzy set theory and grey theory, Engineering Reliability & System Safety 79(1): 61-67.

Schroder, J.U., Mejia Jr, M.Q., Mukherjee, P.K., Manolis, F.M. and Dreessen, S. (2006), Potential consequences of imprecise security assessment", IAMU Journal of Maritime Affairs 4(2): 31-38.

Sekimizu, K. (1997), Current work at IMO on formal safety assessment, Proceeding of Marine Risk Assessment: A Better Way to Manage Your Business, The Institute of Marine Engineers, 8-9 April, London.

Sii, H.S. (2001), Marine and offshore safety assessment, PhD Thesis, Staffordshire University/ Liverpool John Moores University, UK.

Trbojevic, V.M. (2002), Risk-based methodology for safety improvement, Journal of Hazards Materials 71: 467-480.

Wang, J. (2001), Current status of future aspects of formal safety assessment of ships, Safety Science 38: 19-30.

Wang, J. (2002), Offshore safety case approach and formal safety assessment of ships, Journal of Safety Research 33(1): 81-115.

Wang, J. (2006), Maritime risk assessment and its current status, Quality and Reliability Engineering International 22(1): 3-19.

Wang, J. and Foinikis, P. (2001), Formal safety assessment of containerships, Marine Policy 25(2):

143-157.

Wang, J. and Trbojevic, V.M. (2007), Design for Safety of Large Marine and Offshore Engineering Products, Institute of Marine Engineering, Science and Technology (IMarEST), London, UK.

Wang, J., Yang, J.B. and Sen, P. (1995), Safety analysis and synthesis using fuzzy set modelling and evidential reasoning, Reliability Engineering & System Safety 47(3): 103-118.

Wang, J., Yang, J.B. and Sen, P. (1996), Multi-person and multi-attribute design evaluations using evidential reasoning based on subjective safety and cost analyses, Reliability Engineering & System Safety 52(2): 113-128.

Yang, J.B., Liu, J., Wang, J., Sii, H.S. and Wang, Y.M. (2005), Belief rule-base inference methodology using the evidential reasoning approach – RIMER, IEEE Transactions on Systems, Man and Cybernetics – Part A: Systems and Humans 36(2): 266-285.

Yang, Z.L. (2007), Risk assessment and decision making of container supply chains, PhD Thesis, Liverpool John Moores University, UK.

Yang, Z.L., Bonsall, S. and Wang, J. (2009), Use of Hybrid Multiple Uncertain Attribute Decision Making Techniques in Safety Management", Expert System with Applications 36(2): 1569-1586.

The capacity utilization of a container shipping line

Wei-Ming Wu

Graduate Institute of Business and Management, National Kaohsiung First University of Science and Technology, Kaohsiung City 811, Taiwan
E-mail: wwu@ccms.nkfust.edu.tw

Abstract

Under-utilization of fleet capacity has been a long-lasting issue in container shipping industry. In general, capacity utilization ratio is defined as the ratio of actual output to potential output. However, there are still some difficulties unsolved as measuring the CU ratio for a shipping line. In practice, an engineering approach which is based on the installed slot capacity delivered during a year is the general way to represent the potential output. Since the installed capacity output lacks of theoretical support, this paper follows an economics approach by deriving the optimal short-run average cost to find the potential output for a shipping line. In addition, this paper will also develop an output index called TUE-mile index to reflect the influence of transportation distance on measuring the actual output for a shipping line. Once the potential and actual outputs are well defined and measured, the derived CU ratio could be utilized to investigate the fleet utilization for a container shipping line.

Key Words: capacity utilization, container shipping line, potential output

1. Introduction

Capacity utilization (hereafter, CU) has attained significant attention in the literature, as it is an important indicator of economic performance, related to short run output, investment, and employment. Thus, changes in economic environment will alter short run decisions of the firm and affect the observed CU. The CU measures have also been used extensively in helping to investigate the extent of excess capacity installed by an industry or firm. In turn, the studies focused on knowing the role of excess capacity in determining a firm's investment and pricing decisions for maintaining market power have been intensively discussed (Spence 1977; Dixit 1980; Masson and Shaanan 1986).

During the past two decades, the major shipping lines have all devoted to introduce larger containerships into markets. The ultra-large containerships can be deployed efficiently on the major trade routes, provided they are fully loaded. Due to the unpredictable business cycle and slot overcapacity created by shipping lines collectively delivering large containerships into some trade routes, the freight rates and slot utilization have been greatly reduced. As a result, the scale enlargements in vessel and fleet size have not really brought in the expected benefit for the carriers. The persistent over capacity existing in the container shipping industry appears to be puzzling in view of the fact that shipping lines have continuously deployed large containerships into major trade routes. It seems a little odd on observing the coexistence of fleet and vessel size enlargement and the under-utilized fleet capacity prevailing in container shipping market.

Excess capacity is a long-standing problem existing in the container shipping market. Although the issue of capacity appears frequently in shipping studies, we still lack an operationally and economically meaningful approach to properly define or measure the capacity for a shipping line. In order to know the real utilization of fleet capacity for a carrier, conceptually, a method to measure the capacity output should be derived firstly. Meanwhile, it is a quite common practice for a carrier to operate several service routes by deploying different number of ships with different size in a market segment. For example, two shipping lines allocate eight containerships with same size on the trans-Pacific routes. Operationally, the total slot capacities provided by these two carriers are identical. However, the optimal capacity offered by each ship line might not be identical because the sailing frequency and port calls are usually different between two these service routes. Clearly, to develop a method to measure the actual and capacity output becomes a critical task on evaluating the capacity utilization for a shipping line. The main aim of this study is to provide a theoretical-based method to measure the CU ratio for a shipping line.

2. Optimal Capacity: The Concept

Although the term, CU, appears frequently in the industrial organizational studies, there is little consensus as to the proper way of defining or measuring CU. In general, CU is defined as the ratio of actual output to some measure of potential output; however, there is considerable ambiguity in defining the potential (or capacity) output. Some studies employ an engineering approach, in which potential output represents the maximum possible output from a given set of inputs and technology. A principal criticism for such a CU measure is that the crucial like between underlying economic theory and the constructed measure of CU is weak (Berndt and Morrsion, 1981; Nelson1989; Azeez 2001). Academically, Cassels (1937) has pointed out that potential output is conditioned in most cases by economic circumstances and must be interpreted as being the optimum output from the economic point of view. Berndt and Morrsion (1981) defined capacity output as the minimum point on the short-run average cost function. Coelli et at. (2002) suggests a measure where capacity output is given by the point that maximizes short-run profits. In contrast, a CU ratio based on an engineering approach will be difficult to interpret economically. As allowing for the optimization framework, furthermore, both actual and potential outputs may be different and change over time due to the variations of, say, factor prices and/or product prices. Following this, the concept of capacity and CU are inherently short-run notions, conditional on the firm's stock of quasi-fixed inputs and in put prices. Therefore, it measures the output level at which the short-run average cost is minimized.

3. The Model

Assume a shipping line possesses a production function as:

$$Q = f(L,\ F,\ M,\ K,\ T) \qquad (1)$$

where L, F, M and K represent the quantities of labor, fuel, intermediate materials inputs, and stock of capital invested, respectively, and T is an index of technology which is included to express the influence of deploying large containerships and measured by average size among the containership fleet deployed by a carrier. Following the theory of duality, there exists a total cost function dual to the production function (1) such that

$$TC = f(P,\ Q,\ K,\ T) \qquad (2)$$

where P is the price vector of factor inputs.

The total cost function (2) assumes that all inputs are variable. However, in the short run, the shipping line may not be able to alter the level of its capital stock. If the shipping line does not minimize cost with respect to all inputs, the total cost function (2) does not exist (Caves, et. al., 1981, Berndt and Morrison, 1981). Under such a condition, the firm minimizes the cost of a subset of inputs (variable factors) conditional on the levels of the remaining inputs (quasi-fixed factors), then there exists a variable cost function as follows:

$$VC = f(Q,\ P,\ K,\ T) \qquad (3)$$

This variable cost function represents the minimum variable cost of producing a given output, Q, conditional on a given set of input prices, the stock of capital, and a technological parameter.

The short-run total cost ($SRTC$) can be defined as:

$$SRTC = VC + rK \qquad (4)$$

where r is the price of capital. Subsequently, the short-run average cost ($SRAC$) is defined as:

$$SRAC = (VC/Q) + (rK/Q) \qquad (5)$$

In regards to the specification of the functional form, Viton (1981) has utilized the translog variable cost function to estimate the relevant level of capacity output to investigate the cost efficiency of firms in the transportation industry. With reference to Christensen et al. (1975) and Christensen and Greene (1976), the translog variable cost function of equation (3) may be specified as:

$$\begin{aligned}\ln VC = {} & \alpha_0 + \sum_f \delta_f D_f + \sum_i \alpha_i \ln P_i + 0.5 \sum_i \sum_j \alpha_{ij} \ln P_i \ln P_j \\ & + \beta_Q \ln Q + 0.5 \beta_{QQ} (\ln Q)^2 + \sum_i \beta_{Qi} \ln Q \ln P_i + \gamma_K \ln K \\ & + 0.5 \gamma_{KK} (\ln K)^2 + \sum_i \gamma_{Ki} \ln K \ln P_i + \gamma_{KQ} \ln K \ln Q + \delta_T T \\ & + 0.5 \delta_{TT} T^2 + \sum_i \delta_{Ti} T \ln P_i + \delta_{TK} T \ln K + \delta_{TQ} T \ln Q \end{aligned} \tag{6}$$

where $i, j = F, L, M$, and D_f appears in this variable cost equation to represent the dummy variables that are set to account for the immeasurable attributes of a shipping line's operation that remain constant over time. Given the translog variable cost as shown in equation (6), the following restrictions on the parameters should be imposed:

$$\begin{aligned} & \sum_i \alpha_i = 1 \qquad \sum_i \alpha_{ij} = \sum_j \alpha_{ij} = 0 \qquad \sum_i \beta_{Qi} = 0 \\ & \sum_i \delta_{Ti} = 0 \qquad \sum_i \gamma_{Ki} = 0 \end{aligned} \tag{7}$$

Following economics theory, conditions in equation (7) are imposed to ensure that the variable cost is homogeneous of degree one with respect to factor prices and that the Hessian matrix, $\partial VC^2 / \partial P_i P_j$, is symmetrical.

In estimating the parameters shown in equation (6), econometrically, the appropriate approach is Zellner's seemingly unrelated regression (SUR) technique by treating the cost function and the associated cost share functions as a multivariate regression system to estimate the parameters of the cost function (Christensen and Greene, 1976; Caves et al., 1981; Nelson, 1989; Azeez, 2001). By applying Shephard's Lemma, the cost share equations for variable inputs are obtained by logarithmically differentiating equation (6) with respect to the variable input prices, yielding:

$$S_i = \frac{\partial \ln VC}{\partial \ln P_i} = \alpha_i + \sum_j \alpha_{ij} \ln P_j + \beta_{Qi} \ln Q + \gamma_{Ki} \ln K + \delta_{Ti} T \tag{8}$$

In order to measure CU ratio, it is necessary to find the optimal output. If Q_m is defined as the output that minimize $SRAC$, the $\partial SRAC / \partial Q_m = 0$, which in term of (5) implies that

$$\frac{1}{Q_m} \frac{\partial VC}{\partial Q_m} - \frac{VC}{Q_m^2} - \frac{rK}{Q_m^2} = 0 \tag{9}$$

where $\partial VC / \partial Q_m = (\partial \ln VC / \partial \ln Q_m) \times (VC / Q_m)$, and

$$\frac{\partial \ln VC}{\partial \ln Q_m} = \beta_Q + \beta_{QQ} \ln Q_m + \sum_i \beta_{Qi} \ln P_i + \beta_{KQ} \ln K + \delta_{TQ} T \tag{10}$$

Multiplying (10) by (VC / Q_m) and substituting the resulting expression into (9) yields an equilibrium condition for Q_m:

$$\ln Q_m = \left(1 + \frac{rK}{VC} - \beta_Q - \sum_i \beta_{Qi} \ln P_i - \beta_{KQ} \ln K - \delta_{TQ} T \right) \times \frac{1}{\beta_{QQ}} \tag{11}$$

Since Q_m and $\ln Q_m$ appear in (11) simultaneously, it is not possible to obtain an analytical or closed model solution for Q_m using equation (11). For this reason numerical or iterative computational procedure must be employed to solve for Q_m. Then the estimates of CU ratios could be computed by dividing the actual output, Q by the optimal output, Q_m. By definition, a CU ratio could be less, equal or larger than one to indicate underutilization, full-utilization or over-utilization of existing capacity, respectively.

4. Data and Variables

The data used in this study was taken from a sample of three major container shipping lines in Taiwan over the period spanning 1992-2006. The variable cost is defined as the total operating cost. The price of labor is denoted by the average compensation per employee. The fuel price is obtained by dividing the total fuel expenditure by the amount of fuel consumed. Due to the absence of fuel consumption reporting in a company's financial statement, this empirical study has created some regression equations to estimate the fuel consumption for the three shipping lines. These regression equations are developed to incorporate the relationships between actual fuel price, fuel market price, fuel consumption, sailing distance, and ocean freight rate.

Once we have ascertained the fuel and labor costs, the evaluation of the cost of intermediate materials input can be completed easily by subtracting the labor and fuel costs from the total operating cost. Since the items of intermediate materials are fairly diversified, it is not an easy task to find a common unit to compute the unit cost of intermediate materials input. In shipping practice, the cost of intermediate materials is closely related to activities in the operations of moving containers. Thus, the cost of intermediate materials adjusts in proportion to the amount of containers shipped. Accordingly, the unit cost of intermediate materials input per TEU shipped is computed by dividing the cost of intermediate materials by the total TEUs of containers carried in a year.

The total slot capacity is set as the proxy variable of stock capital in this study. Since the containership fleet operated by a shipping line may include owned and charted-in vessels, the total slot capacity provided by the owned and chartered-in containerships is utilized to measure the capital stock invested by a shipping line. As a result, the unit cost of capital is the corresponding capital cost per slot capacity installed. On calculating the unit cost of capital, this study initially estimates the total cost of capital by subtracting the accumulated depreciation from the book value of owned containerships. In turn, the derived net value is multiplied by the average annual interest rate to measure the opportunity cost of operating the owned ships. Given that the depletion of the ship has to be incurred during the ship operation, the annual depreciation expense is used to approximate the cost of depletion of operating the owned ships. In addition to the opportunity cost and depreciation expense of holding owned containerships, the hire expenditure paid for chartering vessels in has also been included in the calculation of the total cost of capital. Finally, the unit cost of capital may be computed by dividing the total cost of capital by the slot capacity actually installed.

In order to allow for the influence of the sailing distance on the measure of the output level, an index called the TEU-mile index has been constructed in this study. The index will be multiplied with the total TEUs of containers shipped to find the amount of TEU-mile shipped by a carrier during a year. The derived fuel consumption per TEU carried could be employed to develop the TUE-mile index. The rationale for applying fuel consumption per TEU carried to represent the TEU-mile index is to reflect the close relationship between fuel consumption and sailing distance.

The data required for computing the three categories of inputs and fuel consumption are all collected from the carriers' annual financial statements. In this study, the financial statements of the three container shipping lines are required by government to report in public and available at the website of Taiwan Stock Exchange Incorporation (http://newmops.tse.com.tw/). The data regarding the total slot capacity of the owned and chartered containerships for each shipping line is collected from the relevant issue of the *Containerisation International Yearbook*.

5. Empirical Results

This empirical study is based on a system of four equations, one variable cost equation and three cost share equations. Since the cost share equations must sum to unity by definition, one equation is deleted to avoid a singular covariance matrix. Estimation is carried out utilizing SUR model, the results of which is equivalent to maximum likelihood estimates, and thus invariant with respect to the deleted equation. The estimated coefficients of the variable cost function, along with their t ratios, are presented in Table 1. The R-square values for the variable cost equation, fuel share equation and intermediate material share equation are respectively, 0.98, 0.9, and 0.63. Of the 30 parameters in variable cost equation, there are 20 parameters are significant at 10% level. To test for the validity of the various parameter restrictions, as shown in equation (7), the likelihood ratio (λ) is estimated. It has been shown that under the null hypothesis (-2log(λ)) has an asymptotic distribution that is chi-square distribution with the number of relevant degrees of freedom equal to the number of restrictions tested. The result presents a significant χ^2 value (38.4) to indicate that the imposition of the restriction among the parameters might not be necessary in this empirical study.

Table 1: Estimated coefficients of translog variable cost function

Variable	Est. Coef.	t-ratio	Variable	Est. Coef.	t-ratio	Variable	Est. Coef.	t-ratio
α_0	40.3	3.8*	α_{LM}	0.0002	-0.01	γ_{KF}	-0.04	-5.42*
δ_1	-0.36	-3.57*	α_{MF}	-0.05	-8.97*	γ_{KM}	-0.03	-1.48
δ_2	-0.31	-3.57*	β_Q	-5.83	-2.41*	γ_{KQ}	-0.49	-1.87*
α_L	-0.008	-0.03	β_{QQ}	1.13	3.42*	δ_T	-1.5	-0.63
α_F	-0.2	-3.26*	β_{QL}	-0.08	-2.9*	δ_{TT}	-0.013	-0.23
α_M	1.2	6.14*	β_{QF}	0.05	7.44*	δ_{TL}	0.02	0.7
α_{LL}	0.02	1.01	β_{QM}	-0.02	-0.86	δ_{TF}	0.03	3.7*
α_{FF}	0.07	14.56*	γ_K	3.3	1.57	δ_{TM}	-0.05	-1.97*
α_{MM}	0.05	2.98*	γ_{KK}	-0.34	-0.99	δ_{TK}	1.07	3.52*
α_{LF}	-0.02	-3.55*	γ_{KL}	-0.08	3.04*	δ_{TQ}	-0.66	-2.52*

Notes: * indicates that the estimated coefficient is statistically significant at the 10% level.

The estimated CU ratios are plotted in Figure 1. All the CU ratios in Figure 1 are less than one. This empirical result indicates that the long-lasting over capacity prevailing in container shipping industry has been verified. Thus, it also provides an evidence to support that the three container shipping lines have all delivered too much fleet capacity from the viewpoint of economic optimization. Since the CU ratios in this paper is a short-run measure and based on economic optimization framework to investigate the actual utilization of installed capacity conditional on the given business environment, the long-lasting under-utilization hints at the possibility of carriers not operating optimally. As referring to the real situation that the shipping market does indeed have gone through an over tonnage situation during the last decade, implicitly, the over capacity installed by the three shipping lines implies that the fleet capacity expansion has become a strategic tool for a container shipping line to survive in market.

In order to reduce the operating cost per TEU shipped and strengthen competitiveness, in practice, most of container shipping lines are devoted to put bigger and more containerships into the market. Due to the impressively cost efficiency on large containership, it has pushed no shipping lines dare to be left behind on expanding the scale of containership fleet. And, such a tendency has put lots of pressures on market participants to move quickly and aggressively. When most of shipping lines have collectively assigned the newer and bigger containerships into markets, as a result, the common actions have pushed the over tonnage situation to go far beyond a foreseeable demand of market. The uncontrolled expansion of fleet capacity, in the meantime, has incurred intense price wars in the market. Eventually, most of shipping lines have not reaped any benefit on deploying big ships because the dropped freight rates have outweighed the cost reductions. Additionally, the huge capital investment has also prevented a shipping line from individually operating a major service route with a frequent schedule. To respond to these problems, most of shipping lines have sought to operate cooperatively in the form of strategic alliance. Since the midst of 1990s, the shipping alliance has gradually become a popular competitive strategy for a shipping line to improve the under-utilized fleet capacity and strengthen its position in the market place.

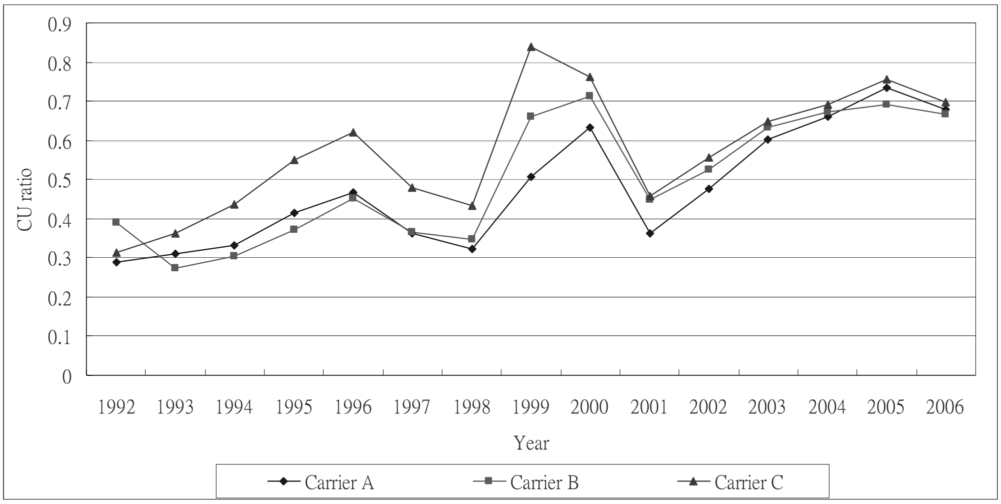


Figure 1: Capacity utilization for the largest container shipping lines in Taiwan

With reference to the curves shown in Figure 1, the CU ratios among the three shipping lines present an approximate twofold growth from around 0.3 to 0.7 during the sample period. Regardless of the persistently under-utilized capacity, the improvement of CU ratios seems to illustrate the contribution of shipping alliance on the utilization of fleet capacity. For examining this viewpoint, a linear multiple regression equation has been developed relating CU ratios to both the ocean freight level and the extent of joining shipping alliance for a carrier. The regression equation derived is:

$$CU_{it} = \underset{(-3.17)}{-5.11} - \underset{(-3.76)}{0.167} D_{1t} - \underset{(-4.2)}{0.207} D_{2t} + \underset{(3.47)}{0.771} P_{it} + \underset{(5.35)}{0.349} RA_{it} \qquad R^2 = 0.51 \qquad (12)$$

where $CU_{it,}$ and P_{it}, represent the capacity utilization ratio and freight level for carrier i at time t, respectively, and D_{1t} and D_{2t} are two dummy variables to separate the realized CU ratios of a carrier from the ones of the other two carries. A variable, RA_{it}, is introduced into equation (12) to express the ratio of service routes joining a shipping alliance to total service routes operated by a carrier i at time t. The data regarding the numbers of service routes operated individually and jointly could be collected from the *Containerisation International Yearbook*. The data of CU ratios are based on the estimates derived in this study. The freight level is computed by giving a set of weights to the six freight indices from the *Review of Maritime Transport*, reported by United Nations secretariat. In reflecting with the significant difference of annual containers shipped in the six trade routes, a triple weight has been put on the routes of Asia-North America and Asia-Europe.

The estimated coefficients have been expressed in equation (12) and the corresponding t ratios have been shown in parenthesis. Surprisingly, the significantly positive effect of ocean freight on the CU ratios demonstrates that the performance of fleet utilization for a shipping line might be mainly dependent on the market demand. That is, a strong shipping demand might have not only pulled the freight level up, but also improved the utilization of containership fleet. In the meantime, the estimated coefficient of RA_{it} variable has also shown a significantly positive effect on the CU ratios. By coordinating sailing schedules, marketing efforts and sharing slot space, in fact, a shipping alliance enables individual members to allocate capacity without purchasing or chartering entire vessels. As a result, the emergence of shipping alliances has significantly improved the utilization of fleet capacity for the member carriers.

Beginning in the late 1990s, international container shipping industry has undergone a revolutionary change in production practice. Lots of mega-carriers with slot capacity more than 6000 TEUs have been deployed in the major trade routes. The deployment of mega-carrier has not only changed the competitive behaviors in the market, but also shifted patterns of container shipping operation. In practice, the activities related to stevedoring operation and containership operation still play the key roles on improving operational efficiency for a shipping line. On measuring the operational efficiency for a container line, the CU ratios are measured to investigate the extent of fleet utilization under a given fleet capacity installed. On the other hand, the efficiency difference is another issue to study the bias of technological change on the input mix. Theoretically, the estimated coefficients shown in Table 1 could be utilized to examine the impact of technological change on the input mix, if the derived cost share function is differentiated with the technology variable. Because the average containership size deployed by a shipping line is set as the proxy of technology variable in this study, the impact of technological change on the variations of input cost shares could be measured by taking derivative of the cost share equation to the average containership size. This is expressed as:

$$\frac{\partial S_i}{\partial T} = \delta_{Ti} \qquad \forall \ i \tag{13}$$

According to the definition of efficiency difference in the Hicksian sense, if the δ_{Ti} in equation (13) is positive, zero or negative, it implies that the efficiency gain follows a pattern of input *i*-using, *i*-neutral or *i*-saving, respectively. Obviously, the parameter, δ_{Ti} could be used to investigate the impact of deploying large containership on the bias of input mix. With reference to the estimated parameters, δ_{TL}, δ_{TF} and δ_{TM} shown in Table 1, it reveals that the deployment of large containerships have significantly changed the distribution of factor shares in production for a shipping line. Both significantly positive value of δ_{TF} and negative value of δ_{TM} indicate a technological change in production toward fuel-using and intermediate-material-saving patterns as a shipping line delivers larger containerships into market. In contrast, the insignificant estimate of δ_{TL} indicates that the deployment of large containership brings a neutral impact on the labor use for a container shipping line. In other words, the deployment of large containership has proportionally increased the labor cost to keep the labor cost share unchanged.

Over all, the trend of introducing larger containership into market has significantly changed the pattern of input mix toward more fuel consumption and away from intermediate material input. This result suggests that the deployment of large containership has not brought about a proportionally reduction on the fuel consumption. As a result, the cost efficiency is mainly created by the gain at adopting more efficient stevedoring-related activities. The changed operational pattern with more fuel consumed has implied that a container shipping line with deploying large containership to improve the competitiveness should have paid more attention on controlling the fuel cost. In addition, the finding of insignificant effect on the labor cost share has also suggested that the raising containership size has no significant effect on saving labors employed.

6. Conclusion

By applying the cost theory in economics, in this paper, a theoretical model on estimating capacity output has been developed and utilized to investigate the fleet utilization for three container shipping lines in Taiwan. A translog variable cost function is used to estimate the output where the short-run average cost is minimized. The findings show that the three shipping lines have all invested too much fleet capacity into market during the past decade. Meanwhile, the strategic alliance prevailing in shipping market has played a positive role on improving the utilization of fleet capacity because the space-sharing activities have significantly released the pressure for a shipping line at endless capacity investment. This paper also discover that the scale enlargement in containership size have changed the factor inputs combination in production toward a more fuel-using and intermediate material-saving pattern.

References

Azeez, A. (2001), Utilization of optimal capacity in Indian manufacturing 1974-1996. Applied Economics Letters 8, 623-628.

Berndt, E. and Morrison, C. (1981), Capacity utilization measures: underlying economic theory and alternative approach, American Economic Review 71, 48-52.

Cassels, J. (1937), Excess capacity and monopolistic competition. Quarterly Journal of Economics, 51, 426-443.

Caves, D., Christensen, L. and Swanson, J. (1981), Productivity growth, scale economics, and capacity utilization in US railroads, 1955-74. American Economic Review 71, 994-1002.

Christensen, L. and Greene, W. (1976), Economies of scale in U.S. electrical power generation. Journal of Political Economy 84, 655-676.

Christensen, L., Jorgenson, D. and Lau, L. (1975), Transcendental logarithmic utility function. American Economic Review 65, 367-383.

Coelli, T., Grifell-Tatje, E. and Perelman, S. (2002), Capacity utilization and profitability: a decomposition of short-run profit efficiency. International Journal of Production Economics 79, 261-278.

Containerisation International Yearbook, 1992-2007, London: Informa, UK Limited.

Dixit, A. (1980), The role of investment in entry-deterrence. The Economic Journal 90, 95-106.

Masson, R. and Shaanan, J. (1986), Excess capacity and limit pricing: an empirical tes, Economica, Vol. 53, 365-378

Nelson, R. (1989), On the measurement of capacity utilization. The Journal of Industrial Economics 37, 273-286.

Review of Maritime Transport, 1996-2007, United Nations secretariat

Spence, A. (1977), Entry, capacity, investment, and oligopolistic pricing. Bell Journal of Economics 8, 534-44.

Viton, P. (1981), A translog cost function for urban bus transit. The Journal of Industrial Economics 26, 287-304.

Demand characteristics of Chinese transport logistics services

Jiaqi Yang

School of Transportation, Wuhan University of Technology, Wuhan, China.

Abstract

China's entrance into the WTO in 2001 led to the establishment of a market-oriented economy for its transport logistics, due to central and local government's attention on the efficiency and effectiveness of transport logistics services, resulting in encouraging progress on the improvement of infrastructure and networking. Nevertheless, until now, the Chinese transport logistics market has remained fairly underdeveloped, in which the demand for transport logistics service matching with the increase of China's economy is plagued with impeding factors. In this context, this research investigates the demand characteristics of Chinese transport logistics services and the factors creating the features. The demand for Chinese transportation logistics is characterized by four aspects. First, compared with developed countries, Chinese enterprises are reluctant to outsource their logistics activities; no signs indicate a spectacular increase in the willingness to operate their logistics services externally. The second feature concerns the logistics functions outsourced, which are now transformed from quantity-oriented to quality-oriented services, diversity and individuality. The third feature creates a state of disequilibrium between China's eastern coastal area and its central and western areas; coastal regions have grown rapidly, benefiting from logistical accessibility and well-developed infrastructure, while interior areas have fallen behind. The last feature concerns market segmentation.

Keywords: Demand characteristics, Chinese transport logistics market, Outsourcing, Regional disparity

1. Introduction

During the central planning years, the concept of logistics did not exist in China; or, to be more precise, the concept of logistics was totally different from that in developed market economies. China's state-owned enterprises (SOE) were actually production units rather than real enterprises. The production was arranged based on the state plan. Firms did not need to bother with the distribution or marketing of products because the government arranged for the distribution of products through its own distribution and transport system/channels. Raw materials and in-process-products needs were also assigned by the government. Since the opening of China's markets in 1978, the distribution and logistics system has been fraught with infrastructure problems and difficult legal issues (Powers, 2001). China's logistics is dynamic and complex, with the characteristics of both planned and market-oriented economies. China's entrance into the WTO in 2001 led to the establishment of a market-oriented economy for its transport logistics, due to central and local government's attention on the efficiency and effectiveness of transport logistics services, resulting in encouraging progress on the improvement of infrastructure and networking.

Nevertheless, until now, the Chinese transport logistics market has remained fairly underdeveloped for two main reasons: (1) the dominance of SOEs and their tendency to own and operate all functions themselves; and (2) the predominance of cellular economies with limited, local distribution areas (Shaw and Wang, 2001). Chinese transport logistics services have witnessed several features that have been increasingly evolving since the beginning of Chinese economic reform; these will be discussed in the following sections.

2. The Willingness of Chinese Enterprises to Outsource their Logistics Services and the Type of Logistics Functions Outsourced

Outsourcing is the strategic decision to contract out one or more activities required by the organization to a TPL specialist (Hong *et al*., 2004). Thus, the overall market capacity for the demand of logistics service, in particular TPL services, depends to a large extent on the firms' willingness to outsource logistics activities, including transportation and warehousing, to outside firms. Obviously, the degree of outsourcing varies and differs from one industry to the next. Table 1, which highlights the annual survey results released by the *China Association of Storage and Warehousing (*CASW*)* since 2001, provides the degree of logistics activities outsourcing by Chinese commercial and manufacturing firms in 2001 and 2004. The table shows that Chinese enterprises are reluctant to outsource their logistics activities, and there are no signs of spectacular increase in the willingness to operate their logistics services externally. Compared to developed countries and areas, the degree of logistics outsourcing in China remains low. A survey carried out by Capgemini (2004) revealed that Western European respondents continue to spend a larger portion of their logistics budget (61 percent) on TPL services than do those in North America (44 percent) and Asia-Pacific (49 percent), but Latin American respondents spend more of their logistics budget (65 percent) on TPL services. Nevertheless, the low level of outsourcing in China presents to some extent a great market potential for TPL, considering China's rapid development of manufacturing, international trade, and domestic wholesale/retail trade.

The lack of awareness of concept and importance of logistics are the main factors affecting Chinese enterprises in outsourcing their logistics services (Hong *et al*., 2004). CASW 2001 survey investigated the importance of logistics for commercial and manufacturing firms. Among the 450 large and medium-sized industrial and commercial enterprises surveyed, 48.7 percent thought that the impacts of logistics were not obvious for their enterprises, and only 7.9 percent thought that logistics were very important for them. In this context, the concept of logistics is not popularized in Chinese enterprises. A great number of enterprises—especially SOEs—have not yet realized the importance and urgency of the use of TPL, pursuing instead the traditional "big-whole and small-whole" model (meaning whatever the size of the firm, everything should be fully equipped in-house on their own) to operate logistics on their own, which inevitable adds up to overall costs.

Other factors impeding logistics outsourcing include the traditional thinking of self-reliance; inefficient management on the shippers' side; and low service quality and high operational costs on TLP's side. Furthermore, unfavourable logistical circumstances, such as segregated administration regimes, the comparatively low degree of commercialization, diversification, and specialization, make Chinese firms run their logistics activities in-house.

Table1: Degree of logistics service outsourcing: 2001 and 2004

(%)

Logistics Service Providers (LSP)	First-Party Logistics (In-house)		Second-Party Logistics (Supplier/Buyer)		Third-Party Logistics (TPL)		Total
	2001	2004	2001	2004	2001	2004	2001/2004
Supply of Materials for Manufacturer	8	25	71	56	21	19	100
Distribution of Finished Products for Manufacturer	43	16	36*	53*	21	31	100
Distribution of Goods for Commercial Firms (Retailer/Wholesaler)	13	78	74	5	13	17	100

*Partially outsourcing and partially in-house

Source: CASW, 2002, 2005

As mentioned in McKinsey & Company's (Shaw and Wang, 2001) report *China's Evolving Logistics Landscape*, most domestic logistics and goods transport needs have largely been met by the SOEs themselves, as they have historically owned and operated their own trucks to deliver goods downstream to distributors and wholesalers. These distributors and wholesalers in turn have bought the finished goods and moved them to consumers through a highly antiquated and fragmented retail trade, leaving manufacturers out of the logistics loop.

Compared to Chinese SOEs, Sino-foreign joint ventures have a much stronger disposition to outsource their logistics and transport activities. Currently, approximately 400,000 joint-venture enterprises operate in China (CASW, 2002). The experiences in outsourcing logistics in the developed market economy of these foreign firms make them realize the importance of logistics for enterprise. Through outsourcing, they can focus on running their core business in order to improve productivity and profit. The trend has been stirred by the Chinese entry into the WTO. In light of China's promises to the WTO, the Chinese logistics market—which includes road haulage, forwarding/NVOCC, warehousing and distribution—is further opened to foreign investment. As such, increasing numbers of foreign enterprises have entered into the Chinese market, which will further enlarge market demand for Chinese logistics services. According to *China Logistics Development Report* (CFLP, 2005-2006), the outsourcing of transport logistics services will continue to expand roughly by 33 percent annually through 2010; the market value will increase from 40 billion RMB in 2004 to 230 billion RMB in 2010 thanks to the improvement of Chinese enterprises' supply chain management and stronger MNC interest and demand for TPL. MNCs relying on China as a global sourcing base are inclined to use—and are experienced in using—TLP, especially those of TPL providers with which the MNCs have established relationships at home. More than 90 percent of MNCs in China currently contract at least a portion of their logistics business to TPL providers.

Meanwhile, China's approximately 2 million private enterprises, accounting for one third of China's GDP, have developed a stronger willingness to outsource logistics services than SOEs (CASW, 2002).

The type of the logistics functions outsourced creates the second demand feature of Chinese transport logistics services. Obvious differences exist between comprehensive transport logistics services and traditional transport services. Traditional transport demand is limited to freight transport. With the globalization and specialization of the world economy and trade, demand for transport logistics has greatly changed. On the one hand, transport logistics demand has transformed from a quantity-oriented service to the quality-oriented one. The value attached to transportation is growing in tandem with the number of specialized cargoes and special service requirements. This requires the restructuring of operational models in the transport industry to better meet shipper demand for quality services. On the other hand, shipper demand demonstrates diversity and individuality, driving transport logistics providers to improve their flexibility and provide "tailor-made" logistics services to meet the changing demand.

Shippers' changing demand is evidenced by the outsourcing functions of Chinese firms. Compared to 2001, current logistics functions outsourced, as shown in Figure 1, remain limited to transport, distribution, and warehousing (CASW, 2005). Nevertheless, the scope of logistics activities outsourcing appears different from earlier years in that current Chinese manufacturers and commercial firms are showing a tendency to concentrate their outsourcing activities primarily on transport, distribution, and warehousing rather than outsource their logistics activities extensively, as found in CASW' first survey in 2001. This phenomenon suggests that Chinese enterprises gradually understand the essence of the logistics concept and no longer simply imitate their foreign counterparts as they did when first learning logistics ideas from developed countries a few years ago.

The evolution of logistical understanding can also be illustrated by Chinese enterprises' perspective of outsourcing functions. Although the functions of transport, distribution, and warehousing still constitute a dominant part of outsourcing, information-based and value-added functions are present to a much larger extent than some years ago. For instance, about one third of commercial firms demand value-added and information-based logistics services, such as packaging/processing, logistics information system designing, labelling/bar-coding, and restructuring of logistics systems. Also, about one third of manufacturing firms require value-added and information-based services such as logistics consulting, customs clearance, logistics information system designing, and restructuring of logistics systems. This evolution is partially consistent with Laarhoven et al. (2000)'s finding, in their survey of shippers in European countries involved in outsourcing their transport, warehousing, and other logistics activities.

3. Regional Disparity in Transport Logistics Demand

With the shift to a world manufacturing centre and rising living standards among citizens, transport logistics is increasing in importance for Chinese manufacturing and commercial firms, as restructuring of the firms often entails relying on external vendors—an attribute of Just-in-Time (JIT) production systems (Shaw and Wang, 2001). However, the trend has been unevenly spread, with the most dynamic increases in China's eastern coastal area, due to the faster economic and international trade growth in this area. In 2004, the top nine provinces/municipalities ranked in terms of volume of external trade—including Guangdong, Jiangsu, Shandong, Shanghai, and Tianjin—accounted for more than 90 percent of China's total. These nine provinces/municipalities are all along the coast. The remaining 22 interior provinces and municipalities, constituting 68.3 percent of the total population, accounted for less than 10 percent of the external trade.

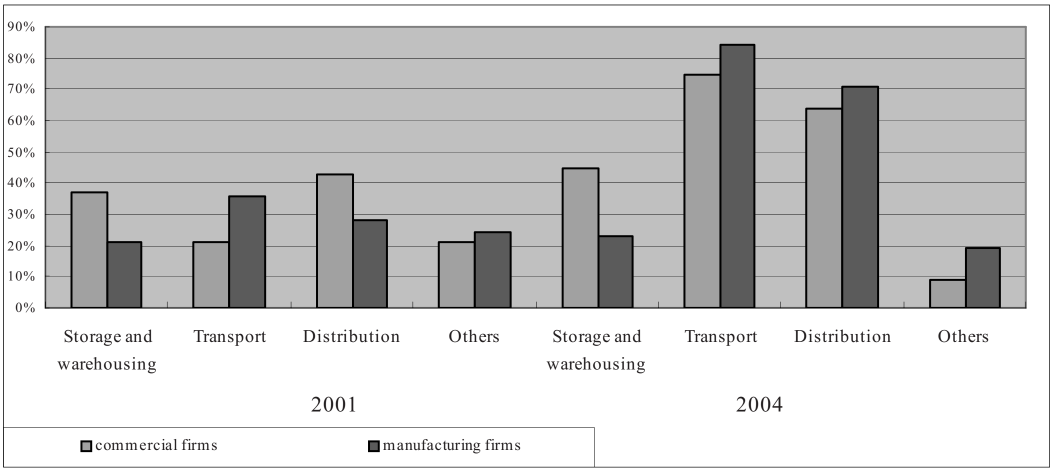


Figure 1: Comparing Logistics activities outsourcing in 2001 and 2004
Source: CASW, 2002, 2005

Meanwhile, regional income disparities between eastern and western regions continue to worsen (Gezen, 2005). As a result, distribution/retail in China's eastern region is far better developed than that in the western region. The proportion of the number of chain/retail stores in the eastern region to China's total reached 75.3 percent in 2003, up 1.4 percent from the previous year, as shown in Figure 2. Of this, the number of chain/retail stores in Shanghai, Beijing, Jiangsu, and Guangdong accounted for almost 50 percent of the total. Accordingly, the turnover of chain stores and retail firms in the eastern coastal area accounted for over 80 percent of China's total in 2003. The same four provinces/municipalities—Shanghai, Beijing, Jiangsu, and Guangdong—created 70 percent of China's total turnover of chain stores and retail; Shanghai alone occupied one third of the total (Report of China Logistics Development, 2005).

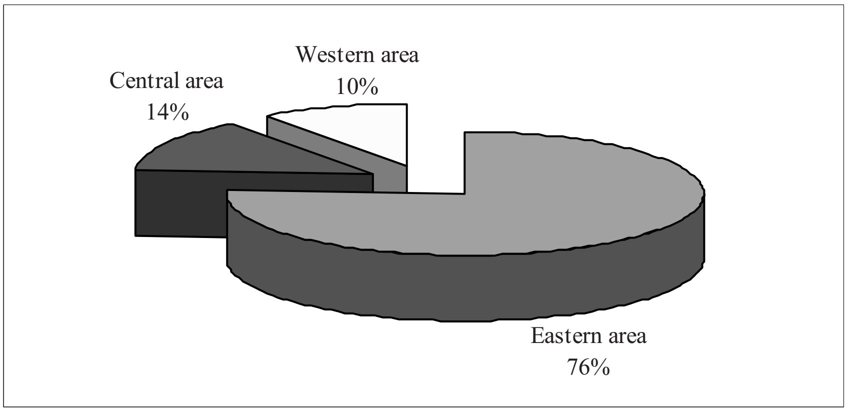


Figure 2: Regional distributions of China's chain stores
Source: Report of China Logistics Development, 2005

Consequently, the third feature of China's transport logistics demand presents a difference between China's eastern coastal area and its central and western areas; coastal regions have grown rapidly, benefitting from logistical accessibility and well-developed infrastructure, while those in the interior have fallen behind. In fact, transport logistics demand has remained concentrated in the central and coastal provinces, around the three major areas of Bohai Bay (Beijing/Tianjin), the Yangtze River Delta (Shanghai/Ningbo), and the Pearl River Delta (Guangzhou/Shenzhen). Cargo movement and industry output are highly concentrated in these areas too; the country's top seaports/airports are located here, and the cities of these three areas are well connected by road networks. Moreover, the populace enjoys some of the highest per capita incomes in the country (Shaw and Wang, 2001). Figure 3 and Table 2 show the dominancy of eastern coastal seaports and airports in China's freight cargo handled.

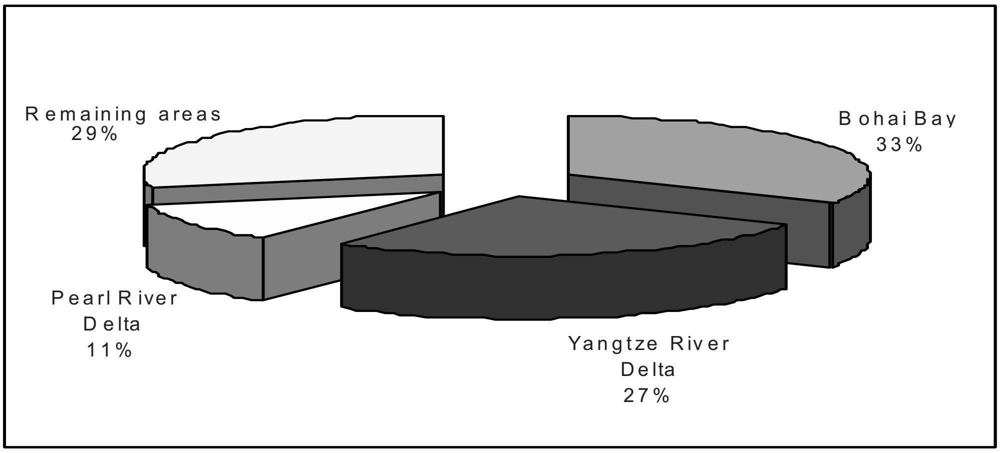


Figure 3: Shares of throughputs of eastern coastal seaports in 2003
Source: MOC1 and Report of China Logistics Development, 2005

Table 2: Freight volumes of eastern coastal airports: 2002-2003

(Million tonnes)

Airports		2002	2003
Total		40183.4	45174.4
Eastern Coastal Areas	Bohai Bay	9080.1	9718.8
	Yangtze River Delta	12680.7	16335.8
	Pearl River Delta	8053.9	8241.2
% Eastern coastal areas of the total		79.97	81.36

Source: Report of China Logistics Development, 2005

Apart from undeveloped manufacturing industry, international trade and domestic distribution/retail trade in the western area, an insufficient transport logistics infrastructure, and an undeveloped network in this region affect the regional imbalance of China's transport logistics demand. Current Chinese government efforts aim to narrow the widening wealth and income disparities between the two regions by attracting more domestic and foreign investment into the interior region (Nogales and Graham, 2004). Following policy changes brought about by China's Great Western Development Strategy (launched in 2000) and Central Rising Strategy (launched in 2006), which boosts public spending on infrastructure and offers private investment incentives to encourage industry to migrate inland, production centres are being created in places in central and western China that have low unit production costs (Kwan and Knutsen, 2006). This tendency will even out to a great extent the imbalance of transport logistics between eastern and western areas.

4. Market Segmentation

The value of the Chinese transport logistics market has consistently experienced double-digit growth each year since the beginning of the decade (Datamonitor, 2006). In 2004, the market generated total revenues of 2344.2 billion RMB, representing an increase of 13.9 percent over the previous year. Focusing its study on the retail, automotive, consumer, hi-tech, and pharmaceuticals segments, Datamonitor (2006) concludes that the retail sector is the most profitable for China's logistics market, generating 480 billion RMB ($62.1 billion) of revenue in 2005, or the equivalent of 76.2 percent of the total market's value. The consumer segment, whose 95.9 billion RMB ($12.4 billion) value generates 15.3 percent of total market revenues, is the second most profitable after retail, as shown in Figure 4. The overall growth of the Chinese economy and the concomitant rise in consumer demand and disposable income should ensure sustained growth in all segments of the transport logistics market.

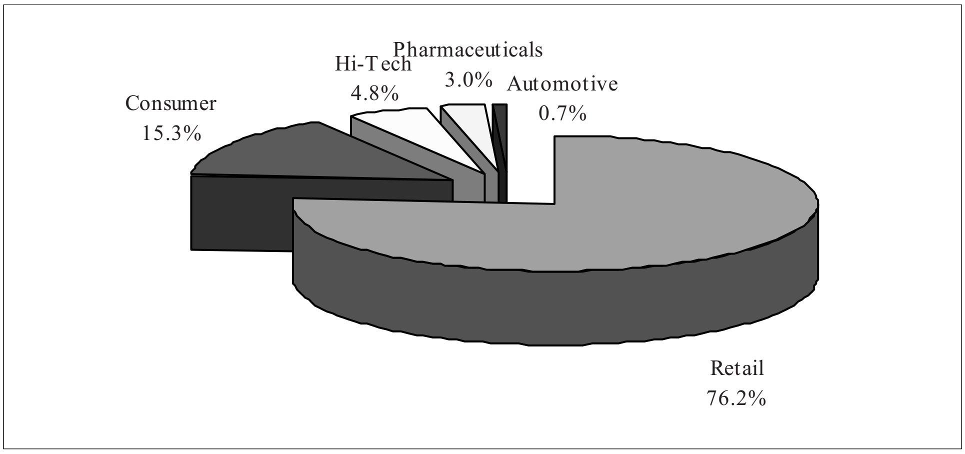


Figure 4: China Logistics Market Value*: Percent share, 2005
*The logistics market value is composed of all expenditures (in-house and outsourced) from the transportation, distribution, and management of retail, consumer electronics, automotive, hi-tech, and pharmaceutical sectors.
Source: Datamonitor, May 2006

Indeed, China's distribution and retail trade has witnessed a great upsurge over the past few years. Large chain stores and hypermarkets are growing rapidly, reflecting the preference of consumers for larger, more appealing stores offering mixed choice assortments, low prices, and trusted brands (Shaw and Wang, 2001). The turnover of top 30 chain stores amounted to 374.4 billion RMB in 2004, up 32.9 percent from the previous year. Table 3 depicts the turnovers of China's top five chain stores in 2004. The emergence of new retail channels is creating demand for transport logistics solutions. Just one consumer electric appliances chain store, Guomei, has increased its number of stores by more than 30 times, growing from 7 to 227 in just 6 years (1998 to 2004). In the process, its presence has expanded from Beijing to Shanghai and several second-tier cities, including a number in northeast and southwest China. Growing sales from such large chain stores, including Guomei, Suning, etc., made up about 7.2 percent of total retail sales in China by 2004 (Report of China Logistics Development, 2005). In this context, while traditional trade in China still cannot be ignored, modern retailers are making real inroads into the first-, second-, and even third-tier cities, creating a greater need for specialized transport logistics solutions to serve these outlets and an ability to bypass trade distributors—that is, wholesalers (Shaw and Wang, 2001).

In addition to the rapid growth of the economy and the concomitant increase in living standards, the deregulation and decentralization of Chinese transport logistics have contributed to the fast development of the distribution and retail industries. In recent years, China's regulatory environment has shifted to accommodate much broader business scopes and more efficient operating structures for wholesalers, retailers, distributors, and other companies by removing layers of bureaucracy. Under this loosened business environment, more and more foreign-invested companies are now using foreign-invested commercial enterprises (FICEs) for retail, wholesale, franchising, and commission-based agency services to trim the fat from current distribution channels and provide better products and services to customers. As a result, distribution and retail industries are facing new opportunities and more choice for their logistics needs.

Table 3: Top five chain stores in China, 2004

Ranking	Chain store	Turnover (Billion RMB)	Growth (%)	Stores (Number)	Growth (%)
1	Bailian Group	67.6	22.5	5493	25.1
2	Guomei Electric Appliance	23.9	34.3	227	63.3
3	Dashang Group	23.1	27	120	25.0
4	Suning Electric Appliance	221.1	79.6	193	30.4
5	Carrefour (China)	16.2	20.9	62	51.2

Source: Report of China Logistics Development, 2005

At the same time, Chinese manufacturers and MNCs investing in China in particular are increasingly adopting integrated national approaches and requiring time-definite delivery of smaller shipment sizes, especially in sectors such as auto components and high tech. Over the past few years, China is witnessing a booming demand for cars with the rising living standards. As a result, Chinese automotive production has increased rapidly; China manufactured 2.24 million automobiles in 2004, up 13.3 percent from previous year. For the automotive and hi-tech manufacturers, consumers are seeking more sophisticated road transport and distribution solutions, leading automotive and hi-tech logistics to become a prominent segment in recent times.

Express and parcel post is another prominent emerging market whose growth reached 23 percent in 2003. According to the US Coalition of Service Industries (USCSI), the compound annual growth rate (CAGR) of the market is estimated to be around 33 percent for the next three-year period, spanning 2004 to 2006 (Report of China Logistics Development, 2005). Datamonitor (2005) estimates China's express delivery market to be valued at $3.5 billion, noting that the driving factor behind China's growth is its increasing export activity with Europe and the U.S. China, the centre of current Asia-Pacific activity, will become the sixth largest express market in the world by 2010, provided it continues growing at an average 20 percent per year. Due to the deregulation of the Chinese distribution regime, express and parcel post alternatives are available to meet manufacturers' needs for delivery of time-sensitive items not only through international providers such as Fedex or TNT, but also through numbers of emerging domestic private express companies as well as the traditional supplier, China Post. As the express firms respond to a fast-growing demand within China, they are looking to less-than-truck-load (LTL) business as an economic and practical alternative to air transport.

In the consumer products segment, fast moving goods (FMCG) is a quickly emerging market in China. After two decades of rapid development, the spending power of Chinese consumers has risen significantly. Because of the increase in disposable income, consumption behaviour has shifted towards the luxurious level. People are increasingly receptive to quality products. China's FMCG market has grown exponentially over the past decade. Changes in distribution infrastructure and relationships have directly affected the evolution of China's FMCG market. However, in general, the ratio of consumption spending to GDP in China remains relatively low. Thus, huge demand exists for quality, particularly for international brand goods. To this end, the total amount of FMCG consumption will further increase. Actually, China's FMCG has become a highly competitive market, featuring a variety of products, fast circulation, and low profits. In order to cut down on logistics costs, FMCG firms either establish their own logistics network to touch the end market directly, or outsource logistics to large and respectable logistics enterprises. This presents a prospective boom in the FMCG market.

5. Conclusions

The demand features of Chinese transportation logistics is categorized into four aspects. The first feature is that, compared with developed countries, Chinese enterprises are reluctant to outsource their logistics activities; no signs indicate a spectacular increase in the willingness to operate their logistics services externally. The lack of awareness of the concept and importance of logistics is the main factor hindering the willingness of outsourcing. Other factors include: traditional thinking of "self-reliance" and inefficient enterprise management on shippers' side; low-level service quality and high operational costs on TLPs' side; unfavourable logistics circumstances, such as segregated administration regimes; and comparatively low degree of commercialization, diversification, and specialization. The second feature of China's transport logistics demand concerns the logistics functions outsourced, which are now transformed from quantity-oriented to quality-oriented services, diversity and individuality. The third feature creates a state of disequilibrium between China's eastern coastal area and its central and western areas; coastal regions have grown rapidly, benefitting from logistical accessibility and well-developed infrastructure, while interior areas have fallen behind. The last feature concerns market segmentation.

References

Capgemini (2004), Logistics outsourcing is an important driver of topline growth corporate strategy according to new global study. Accessible from http://www.capgemini.com/resources/news.
China Association of Storage and Warehousing (CASW) (2002 and 2005), Survey in the demand and supply of Chinese logistics market in 2001/2005 (in Chinese). Application of Logistics Technology Vol.3: 9-21.

China Federation of Logistics & Purchasing (CFLP) (2005- 2006), China Logistics Development Report. China Logistics Publishing House: Beijing.

Datamonitor (2005), Express industry sees China as a strategic growth market. Accessible from http://www.logisticstoday.com/sNO/7938/LT/displayStory.asp

Datamonitor (2006), Logistics in China - Industry Profile. Accessible from http://www.datamonitor.com/industries/taxonomy/

Gezen, A. (2005), Going intermodal. China Business Review 32(4):8-16

Hong, J. (2004), Logistics outsourcing by manufacturers in China: a survey of the industry. Transportation Journal Winter: 17-25.

Kwan, C. and Knutsen, K. (2006), Intermodal Revolution. China Business Review 33(4): 20-25.

Laarhoven, P., Berglund, M. and Peters, M. (2000), Third-party logistics in Europe-five years later. International Journal of Physical Distribution & Logistics Management 30(5): 425-442.

Nogales, A. and Smith G. (2004), China's evolving transportation sector. China Business Review 31(2): 24-29.

Powers, P. (2001), Distribution in China: the end of the beginning. China Business Review 28(4): 8-12.

Report of China Logistics Development (in Chinese) (2005), China Machine Press: Beijing.

Shaw, S. and Wang F. (2001), China's Evolving Logistics Landscape, McKinsey & Company, Greater China Office: Beijing.

Use of fuzzy AHP to determine port hinterland development assessment criteria for free trade zone

Yi-Chih Yang

Department of Shipping and Transportation Management, National Kaoshiung Marine University,
Kaohsiung, 81143 Taiwan, R.O.C.
Tel: +886-7-3617141 ext 3158
E-mail: hgyang@mail.nkmu.edu.tw

Abstract

The goal of this paper is to highlight the functions of port hinterland and use the fuzzy AHP method to identify key success factors for port hinterland development in the Free Trade Zone, and provide specific suggestions for the government.

The key success factors of the FTZ port hinterland in Taiwan are integration of customs and port logistics information, efficiency of port operations, exemption and deduction of custom duties and value-added tax for cargo, stability of political climate, economic scale of market, soundness of investment system and incentive measures, exemption or deduction of corporate and local taxes, direct shipping across the Taiwan Strait, cost of labor, convenience of customs clearance procedures and one-stop administrative service window, efficiency of intermodal transport network, transport and distribution costs, cost of land, adequacy of port logistics facilities, frequency of ship sailing and diversification of shipping route and adequacy of the port hinterland for logistics functions.

Keywords: Free Trade Zone, Port, Hinterland, Fuzzy AHP

1. Introduction

The "Challenge 2008: National Development Focus Plan" called for the establishment of free trade zones as one of its important measures, and the government of Taiwan promulgated the *Act for the Establishment and Management of Free Trade Zones* (hereinafter referred to the FTZ Act) in July 2003. The FTZ Act aims to promote the development of a new operational model for international logistics and management, accelerate trade liberalization mechanisms, enhance national competitiveness, and facilitate national economic development.

A "designated FTZ" refers to an area situated within the controlled district of an international airport or an international seaport designated with the approval of the Executive Yuan(means premier office), or an adjacent area demarcated as a controlled area, or an industrial park, export processing zone, science-based industrial park, or other area approved by the Executive Yuan for the establishment of a controlled district for the purpose of conducting domestic and foreign business activities in which a comprehensive goods tracking system can be implemented by means of technical facilities.

Article 1 of the FTZ Act being enacted in July 2003, states that the goals of Taiwan's port free trade zones are to develop an operating model for global logistics and management systems, aggressively promote trade liberalization and globalization, facilitate the smooth flow of personnel, goods, finance, and technology, enhance national competitiveness, and further national economic development (Executive Yuan, 2004). In accordance with the FTZ Act, the Central government approved management plan formulated by port authorities, such as they ratified *Regulations Governing Entry, Exit, and Residence in the Keelung Free Trade Zone* in 2004. In addition, The other four FTZs, which include the Port of Kaohsiung, Port of Taipei, Port of Taichung, and Taoyuan Air Cargo Park, where were approved by the Executive Yuan in 2005.

Taiwan's FTZ's mechanism can benefit companies in many respects. For instance, they can promote various production and trade activities, including transhipment, distribution, reassembly, consolidation of containers from different countries, simple processing, and intensive processing. They can conserve as much as 95% of the time required for logistics, and can boost added value by 4-30 times for simple or intensive processing in the FTZ. In addition, goods and equipments flow inside the FTZ are exempt from taxation, including import duties, commodity tax, business tax, tobacco and alcohol tax, health donation for tobacco products, trade promotion service fees, and harbour service fees, which can enable substantial cost savings.

As for operating efficiency, FTZs are deemed to be within the national territory but outside the customs territory of Taiwan. Goods within an FTZ may be transported freely without examination, inspection, customs clearance, or escorted shipment. Therefore, such goods can enjoy high degree of freedom and flexibility of movement, and can exercise value-added processing (Centre for Economic Deregulation and Innovation, 2006).

Table 1: Total trade number of Free Trade Zone

Unit: NT$1,000

	2006			2007(1-7)		
	Import	Export	Sub- trade	Import	Export	Sub-total
Keelung Port FTZ	50,195	18,993	69,188	675,592	681,827	1,357,419
Taipei Port FTZ	3,052,689	1,378,341	4,431,030	1,172,440	1,742,439	2,914,879
Kaohsiung Port FTZ	4,464,397	3,008,376	7,472,773	1,377,268	1,809,167	3,186,435
Taichung Port FTZ	715,307	407,887	1,123,194	803,610	554,419	1,358,029
Taoyuan Air Cargo Park FTZ	68,646	991,581	1,060,227	8,561,650	73,787,835	82,349,485
Total	8,351,234	5,805,178	14,156,412	12,590,560	78,575,687	91,166,247

Source: Ministry of Transportations and Communications, Statistics of Free Trade Zone, http://www.motc.gov.tw.

Table 2: Tenant companies numbers of Free Trade Zone

Port Name	Operation date	2005	2006	2007.6
Keelung Port FTZ	2004.9.	4	8	13
Taipei Port FTZ	2005.6	1	1	1
Kaohsiung Port FTZ	2005.1	7	11	15
Taichung Port FTZ	2005.12	3	7	12
Taoyuan Air Cargo Park FTZ	2005.11	5	75	79
Total Amount		20	102	120

Source: Ministry of Transportations and Communications, Statistics of Free Trade Zone, http://www.motc.gov.tw.

Table 1 published by Ministry of Transportations and Communications represent total trade number of FTZs has increased from 14,156,412 thousand NTD of 2006 to 91,166,247 thousand NTD of 2007(Jan. to July), Keelung port FTZ has risen from 69,188 thousand NTD of 2006 to 1,357,419 thousand of 2007(Jan. to July). However Kaohsiung port FTZ were decreased from 7,472,773 NTD of 2006 to 3,186,435NTD of 2007(Jan. to July). Table 2 depict tenant companies number statistics of FTZs has heighten from 20 firms of 2005 to 120 firms of June 2007 after operating date commenced, Kaohsiung FTZ has increased from 7 firms of January 2005 to 15 firms of June 2007, the number of

Keelung port has added from 4 firms of September 2004 to 13 firms of June 2007. The main purposes of Taiwan's FTZ system are to extend the current functions of the "Global Logistic Development Project," confronting with severe competitive pressure from neighboring countries (e.g. China, Hong Kong, Singapore, Japan, and Korea) which are aggressively establishing their own FTZs nearby the seaport and airport hinterland areas by reinforcing transshipment and logistic functions.

Aforementioned by the Center for Economic Planning and Development, Executive Yuan, the essential role of an FTZ consists of simplifying business transaction procedures, facilitating free flow of commodities in the FTZ, exempting goods from customs administration and customs clearance procedures, allowing foreign managers who engage in business activities in an FTZ to receive 72-hour visas, and providing other preferential measures.

The Fuzzy Analytic Hierarchy Process (FAHP) method is used to determine the weighting for evaluation criteria among decision makers, the subjectivity and vagueness in the alternatives selection process is dealt with by using fuzzy numbers for linguistic terms (Hsieh et al., 2004). We use FAHP to identify the evaluation criteria weights of FTZ's port hinterland development by performing two questionnaires with expert groups including scholars, port authority officers, managers of shipping companies or global logistics.

Therefore, the purpose of this paper is to:
1. Review the performance of Taiwan FTZ's port hinterland.
2. Highlight the multi functions of FTZ's port hinterland development
3. Identify assessment criteria of FTZ's port hinterland development.

2. Assessment Criteria of FTZ's Port Hinterland Development

The experiences of FTZ mainly in East and Southeast Asia, Latin America and African regions have been evaluated over the wide range of issues such as the level of employment creation, technology up-grading and foreign exchange earnings that the macro level, and firm and industrial performance, wages and working conditions and industrial level at the micro level(Young, 1987).

Miyagiwa (1986) claims that achieving a higher national income and a lower level of unemployment is dependent upon the FTZs location within the economy. The chief goal of a free trade zone is to facilitate export-oriented industrial development. Most well-developed free trade zones around the world offer a variety of benefits to the companies that operate in them. These benefits include tax exemptions, duty-free imports, and personnel recruitment services. The main advantage of a free trade zone is that it permits the free unhindered import of various materials and components for assembly and manufacture for subsequent export without duties and taxes. Frankel (1987) confirmed that benefits to the host country are usually obtained from large infrastructure and other investment, technology transfer, employment and jobs, added port and transport revenues or transport cost savings, foreign exchange earnings as investment, salaries, and port or transport are paid in foreign exchange, and attraction of export and transshipment revenues.

Furthermore, Ghanem (1997) concluded that FTZs offer economic incentives including no restrictions on investment activities, reduced tariffs for importation to the home market, freedom of importation from domestic or foreign markets, exemption from taxes and duties, and exemption of capital assets from customs duties. Free trade zones provide the same export policies to those foreign and domestic firms within a designated geographical area. In addition, income tax incentives and extremely liberal foreign exchange regulations are offered in most FTZs as additional incentives to firms engaging in export activities (World Bank, 1990).

Free trade zones are established with the objective of having a positive effect on the economy. From a national perspective, governments look for the following outcomes from free trade zones: generation of foreign exchange earnings, provision of jobs and creation of income, attraction of foreign direct investment, and generation of technological transfer (World Bank, 1990). FTZs encourage activities such as assembling, packaging, destroying, storing, cleaning, exhibiting, re-packing, distributing, sorting, grading, testing, labelling, repairing, combining with foreign or domestic content, and processing (Engman et al., 2007). In line with past experiences in other countries, the free trade zones established in Taiwan have had the following economic effects (Tseng, 2004): absorption of foreign capital investment, enhancement of foreign currency earnings, expansion of employment, promotion of technology transfer, improvement of local industry economics, expansion of port operating revenue, and attraction of cargo consolidation and transhipment business.

Most of the world's FTZs have the similar objectives of attracting foreign investment, promoting industrialization, creating job opportunities for local labour, gaining access to foreign knowledge and technology, and generating economic benefits for the host countries. Key elements of FTZs are incentive packages offered to foreign investors. These include unlimited duty deductions or exemptions from import duties on raw materials, intermediate inputs, and capital goods used in the production of exported products, exemption from the payment of sales tax on exported products and on all goods and services purchased domestically and used in the production of exported products, tax holidays, rebates, or reduced tax rates on corporate income or profits that are linked with corporate export performance or with exports as a percentage of total production.

Jin and Li (2007) argued that designated port hinterlands integrating logistics, trade, industry, distribution and living functions, etc, facilitate logistics services and value-added services through linkage between the hinterland and various destinations can play major roles in national economic development.

Li (2006) suggested that port hinterland areas may be classified as infrastructure and superstructure, where the former includes drainage, power, energy, and roads, and latter includes warehouses, buildings, convention centers, logistics facilities, greenbelts, parks, and other facilities. Yang (2003) proposed that, in order to improve efficiency and support essential logistics facilities (including hinterland distribution, processing, fabrication, packing facilities, etc.) port hinterland development strategies should reflect the relationships between facility clusters and land use functions, such as through tight concentrations of functions and facilities, or through linkage between port and hinterland.

The Korean Port Law defines port hinterland as space for the establishment and development of a port's supporting facilities and waterfront facilities. The hinterland areas of international trading ports are expected to create added value, strengthen port-related industries, and promote the convenience of users. The 2003 "Study of the Developmental Directions and Prospects of Pyeongtaek Port in the 21st Century," published by the Korea Maritime Institute (KMI, 2003), suggested that the purpose of port hinterland is to facilitate and promote port functions such as high value-added services, reduce logistics costs, and directly foster port-related industries. The functions of port hinterland include container logistics, processing, fabrication, commerce, business, research, venture incubation, and waterfront recreation, etc. It is therefore necessary to use hinterland areas for various purposes reflecting different functions. The 2002 "Integrated Port Hinterland Development Project," published by Korea's Ministry of Maritime Affairs and Fisheries, describes the basic concepts and functions of port hinterlands, and encompasses their logistics functions, fabrication and processing functions, commercial and business functions, waterfront recreation functions, and port-supporting functions (see Table 3).

Table 3: Classification of Port Hinterland Functions

Classification	Function & facilities	Content
Logistics function	Chief function	Large scale transshipment and storage functions, some fabrication, processing, labeling, packing, and other value-added service functions, and certain supporting functions from container terminal
	Investment facilities	Logistics warehouse, distribution center, storage warehouse, and empty container storage yard
Fabrication and processing function	Chief function	Large-unit value-added logistics services focusing on cargo fabrication and processing, and not including transshipment, storage, and other basic logistics functions
	Investment facilities	Fabrication facility, processing facility, and other related auxiliary facilities
Waterfront recreation function	Chief function	Employs the outer edge of waterfront facilities to provide certain leisure and entertainment functions to local residents and port users
	Investment facilities	Leisure, recreation, maritime park, etc.
Commercial and business function	Chief function	Apart from major transshipment and storage functions, commodity promotion, exhibition, and conference-related functions established in fabrication or processing areas
	Investment facilities	Exhibition centers, business facilities, convention centers, restaurants, lodging, and entertainment facilities
R&D and venture function	Chief function	Educational, research, and venture business functions in port hinterland areas contributing to the development of sustainable logistics and port industries
	Investment facilities	Educational, research, and venture facilities, etc.

Source: KMI (2003), Study of the Developmental Directions and Prospects of Pyeongtaek Port in the 21st Century

Korea Maritime Institute and Cun-Ang University(2000) concluded that the key factors determining the success of a FTZ hinterland can be summarized as management and strategic advantages for global enterprises, appropriate functions and type of free trade zone, scale of hinterland economic area, soundness of investment conditions, efficiency of infrastructure facilities, and character of the socioeconomic culture. KMI(2000) proposed that the chief factors determining the success of a FTZ are adequate infrastructure facilities, simplified customs clearance procedures, preferential tax measures, an excellent labor force, and an advanced integrated information system. Yang (2003) suggested that factors determining tenant enterprise willingness to invest in an FTZ can be classified along various dimensions, such as political-economic environment dimension (politics, policies, and economics), production dimensions (cluster effect, acquisition of materials), cost dimension (cost of labor, energy cost, port duties, tax incentive terms), and infrastructure dimension (services, location, facilities).

These major preliminary assessment factors collected from expert interview and literature review can be summarized as five assessment dimensions (including Political-Economic Environment, Operational Environment, Cost Environment, Infrastructure Facility Environment, and Preferential Incentive Environment) containing 20 assessment criteria (stability of political climate, direct shipping across the Taiwan Strait, economic scale of market, globalization of international trade and foreign currency exchange systems, administrative efficiency of local government agencies, convenience of customs clearance procedures and one-stop administrative service window, efficiency of port operations, integration of customs and port logistics information, cost of labor, cost of land, raw material purchasing costs, transport and distribution costs, adequacy of port logistics facilities, adequacy of the port hinterland for logistics functions, efficiency of intermodal transport network, frequency of ship sailing and diversification of shipping route, soundness of investment system and incentive measures, exemption or deduction of corporate and local taxes, exemption and deduction of custom duties and value-added tax for cargo, and financial aid for invested companies)(see Table 4).

Table 4: Main assessment criteria for port hinterland of Free Trade Zone

Dimension	Assessment Criteria	Literature Source
Political-Economic Environment	stability of political climate	(Qing Qiao Yang,2003)
	direct shipping across the Taiwan Strait	
	economic scale of market	(KMI & CAU, 2000; KMI, 2005)
	globalization of international trade and foreign currency exchange systems	(Young,1987; KMI, 2003)
Operational Environment	administrative efficiency of local government institution	(Qing Qiao Yang, 2003)
	convenience of customs clearance procedures and one-stop administrative service window	(KMI & CAU, 2000; KMI, 2000)
	efficiency of port operations	(KMI & CAU, 2000)、(Qing Qiao Yang,2003)
	integration of customs and port logistics information	(KMI & CAU, 2000)
Cost Environment	cost of labor	(Qing Qiao Yang,2003)
	cost of land	(Qing Qiao Yang,2003)
	Raw Material Purchasing Cost	(KMI & CAU, 2000)
	transport and distribution costs	(KMI & CAU, 2000; KMI, 2000)
Infrastructure Facility Environment	adequacy of port logistics facilities	(KMI & CAU, 2000; KMI, 2000)
	adequacy of the port hinterland for logistics functions	(KMI & CAU, 2000)
	efficiency of intermodal transport network	(KMI, 2003; KMI, 2000)
	frequency of ship sailing and diversification of shipping route	(KMI & CAU, 2000; KMI, 2003)
Preferential Incentives Environment	soundness of investment system and incentive measures	(KMI & CAU, 2000; KMI, 2000)
	exemption or deduction of corporate and local taxes	(KMI & CAU, 2000)、(Ghanem, 1997)、(KMI, 2005)
	exemption and deduction of custom duties and value-added tax for cargo	(Ghanem, 1997; Young,1987)
	financial aid for invested company	(KMI,2005; KMI & CAU, 2000; KMI, 2000)

3. Methodology

AHP approach is the commonly used one for multi-criteria problem. However, the relative weights base upon this measurement in which information is incomplete or imprecise or views that are subjective or endowed the linguistic characterizes creating a fuzzy environment. The use of fuzzy numbers would be more suitable in that situation. Therefore, the fuzzy AHP approach is used to measure relative weights for evaluation these key factors.

The systematic steps for evaluating relative weights using fuzzy AHP to be taken are described below(Xu, 1998; Ding, 2006).

Step 1. Develop a hierarchical structure

A hierarchy structure is the framework of system structure. We can skeletonize a hierarchy to evaluate research problems and benefit the context. The hierarchy of preliminary factors can be constructed as same as Figure 1. Figure 1 is a hierarchical structure with k activity systems on the $L+1$ layer, $p+...+q+...+r$ functional activities on the $L+2$ layer and $e_1+...+e_p+...+f_1+...+f_q+...+g_1+...+g_r$ functional-related activities on the $L+3$ layer, respectively (see Figure 1).

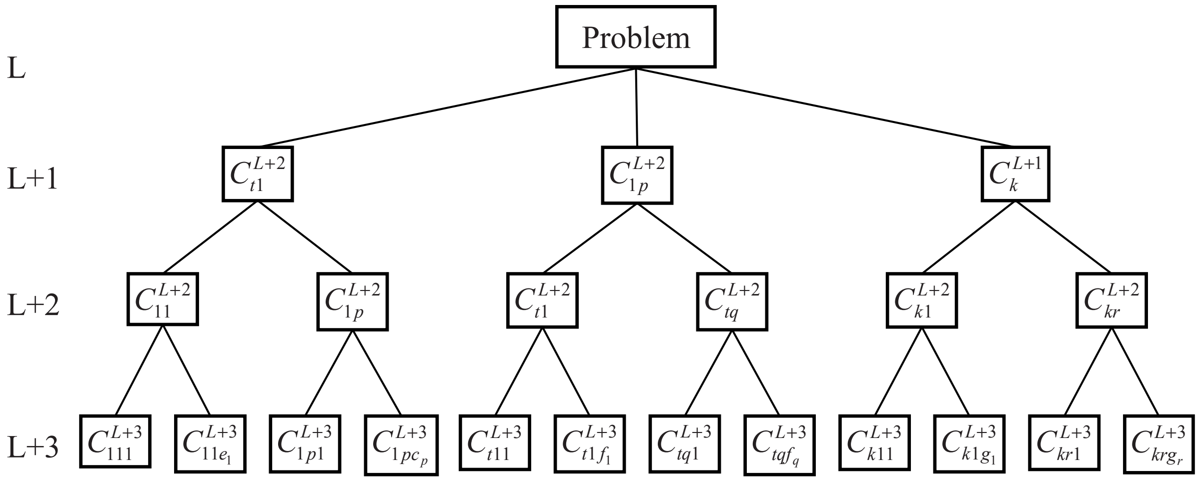


Figure 1 hierarchical structure of FAHP

Step 2. Collect pair-wise comparison matrices of decision attributes

We choose experts to collect pair-wise comparison matrices of decision attributes, which is represented the relative importance of each pair-wise attribute.

(1) Let x_{ij}^{h}, h=1, 2, …, n, be the relative importance given to activity system i to activity system j by expert h on the L+1 layer. Then, the pair-wise comparison matrix is defined as $[x_{ij}^{h}]_{k\times k}$.

(2) Let x_{uv}^{h}, h=1, 2, …, n, be the relative importance given to functional activity u to functional activity v by expert h in the L+2 layer. Then, the pair-wise comparison matrix with respect to each activity system, i.e., $C_{1}^{L+1}, C_{t}^{L+1}, C_{k}^{L+1}$ is defined as $[x_{uv}^{h}]_{p\times p}, [x_{uv}^{h}]_{q\times q}, [x_{uv}^{h}]_{r\times r}$.

(3) Let x_{yz}^{h}, h=1, 2, …, n, be the relative importance given to functional-related activity y to functional-related activity z by expert h on the L+3 layer. Then, the pair-wise comparison matrix with respect to each functional activity, i.e., $C_{11}^{L+2}, C_{1p}^{L+2}, C_{t1}^{L+2}, C_{tq}^{L+2}, C_{k1}^{L+2}, C_{kr}^{L+2}$ is defined as $[x_{yz}^{h}]_{e_1\times e_1}, [x_{yz}^{h}]_{e_p\times e_p}, [x_{yz}^{h}]_{f_1\times f_1}, [x_{yz}^{h}]_{f_q\times f_q}, [x_{yz}^{h}]_{g_1\times g_1}, [x_{yz}^{h}]_{g_r\times g_r}$.

Step 3. Transform relative importance into triangular fuzzy number

The generalized means is a typical representation of many well-known averaging operations, e.g., min, max, geometric mean, arithmetic mean, harmonic mean, etc. The min and max are the lower bound and upper bound of generalized means, respectively. Besides, the geometric mean is more effective in representing the multiple decision-makers' consensus opinions (Saaty, 1980). To aggregate all information generated by different averaging operations, we use the grade of membership to demonstrate their strength after considering all approaches. For the above-mentioned reasons, the triangular fuzzy numbers characterized by using the min, max and geometric mean operations are used to convey the opinions of all experts.

Let $x_{ij}^{h} \in [1/9,1] \cup [1,9]$, h=1, 2, …, n, $\forall i, j = 1,2,...,k$, be the relative importance given to activity system I to activity system j by expert h on the L+1 layer. After integrating the opinions of all n experts, the triangular fuzzy numbers can be denoted by

$$\tilde{A}_{ij}^{L+1} = (c_{ij}, a_{ij}, b_{ij}),$$

Where

$$C_{ij} = \min\{x_{ij}^1, x_{ij}^2, ..., x_{ij}^n\}, \quad a_{ij} = \{\prod_{h=1}^{n} x_{ij}^h\}^{1/n},$$

$$b_{ij} = \max\{x_{ij}^1, x_{ij}^2, ..., x_{ij}^n\}.$$

By the same concept, we can integrate the opinions of all n experts on the L+2 layer, i.e., the triangular fuzzy numbers can be denoted by

$$\tilde{A}_{uv}^{L+2} = (c_{uv}, a_{uv}, b_{uv}),$$

$$\forall u, v = 1, ..., p; ...; \quad \forall u, v = 1, ..., q; ...; \quad \forall u, v = 1, ..., r,$$

Where

$$c_{uv} = \min\{x_{uv}^1, x_{uv}^2, ..., x_{uv}^n\}, \quad a_{uv} = \{\prod_{h=1}^{n} x_{uv}^h\}^{1/n},$$
$$b_{uv} = \max\{x_{uv}^1, x_{uv}^2, ..., x_{uv}^n\}.$$

For saving space, the equation of triangular fuzzy number is omitted to reason by analogy on the L+3 layer.

Step 4. Build fuzzy positive reciprocal matrices

We use the integrated triangular fuzzy numbers to build fuzzy positive reciprocal matrices. For the L+1 layer, the fuzzy positive reciprocal matrix can be denoted by

$$A = [\tilde{A}_{ij}^{L+1}] = \begin{bmatrix} 1 & \tilde{A}_{12}^{L+1} & \Lambda & \tilde{A}_{1k}^{L+1} \\ 1/\tilde{A}_{12}^{L+2} & 1 & \Lambda & \tilde{A}_{2k}^{L+1} \\ M & M & O & M \\ 1/\tilde{A}_{1k}^{L+1} & 1/\tilde{A}_{2k}^{L+1} & \Lambda & 1 \end{bmatrix}$$

Where

$$\tilde{A}_{ij}^{L+1} \otimes \tilde{A}_{ji}^{L+1} \cong 1, \quad \forall i, j = 1, 2, ..., k.$$

For saving space, the equations of fuzzy positive reciprocal matrices are omitted to reason by analogy on the L+2 and L+3 layers.

Step 5. Calculate the fuzzy weights of the fuzzy positive reciprocal matrices

Let $\tilde{Z}_i^{L+1} \cong (\tilde{A}_{i1}^{L+1} \otimes \tilde{A}i2^{L+1} \otimes ... \otimes \tilde{A}_{ik}^{L+1})^{1/k}, \forall i = 1, 2, ..., k,$ be the geometric mean of triangular fuzzy number of i th activity system on the L+1 layer. Then, the fuzzy weight of i the activity system can be denoted by

$$\tilde{W}_i^{L+1} \cong \tilde{Z}_i^{L+1} \otimes (\tilde{Z}_1^{L+1} \oplus \tilde{Z}_2^{L+1} \oplus ... \oplus \tilde{Z}_k^{L+1})^{-1}.$$

For being convenient, the fuzzy weight is denoted by $\tilde{W}_i^{L+1} = (w_{ic}, w_{ia}, w_{ib})$. For saving space, the equations of fuzzy weights are omitted to reason by analogy on the L+2 and L+3 layers.

Step 6. Defuzzify the fuzzy weights to crisp weights

For solving the problem of defuzzification powerfully, the graded mean integration representation method, proposed by Chen and Hsieh (2000), is used to defuzzify the fuzzy weights. Let $\tilde{W}_i^{L+1} = (w_{ic}, w_{ia}, w_{ib})$, $\forall i = 1,2,...,k,$ be k triangular fuzzy numbers. By the powerful method, the graded mean integration representation of crisp weights k can be denoted by $\tilde{W}_i^{L+1} = \dfrac{w_{ic} + 4w_{ia} + w_{ib}}{6}, \forall i = 1,2,...,k.$

For saving space, the defuzzifications of fuzzy weights are omitted to reason by analogy on the L+2 and L+3 layers.

Step 7. Normalize the crisp weights

For being convenient to compare the relative importance between each layer, these crisp weights are normalized and denoted by

$$NW_i^{L+1} = \frac{W_i^{L+1}}{\sum_{i=1}^{k} W_i^{L+1}}.$$

Step 8. Calculate the integrated weights for each layer

Let NW_i^{L+1}, NW_u^{L+2} and NW_y^{L+3} be the normalized crisp weights in the L+1, L+2 and L+3 layers, respectively. Then,

(1) The integrated weights of each activity system on the L+1 layer is

$$IW_i^{L+1} = NW_i^{L+1}, \quad \forall i = 1,2,...,k.$$

(2) The integrated weights of each functional activity in the L+2 layer is

$$IW_u^{L+2} = NW_i^{L+1} \times NW_u^{L+2}, \quad \forall i = 1,2,...,k;$$
$$\forall u = 1,..., p;...; \quad \forall u = 1,..., q;...; \quad \forall u = 1,..., r.$$

(3) The integrated weights of each functional-related activity on the L+3 layer is

$$IW_y^{L+3} = NW_i^{L+1} \times NW_u^{L+2} \times NW_y^{L+3}, \quad \forall i = 1,2,...,k;$$
$$\forall u = 1,..., p;...; \quad \forall u = 1,..., q;...; \quad \forall u = 1,..., r.$$
$$\forall y = 1,..., e_1;...; \quad \forall y = 1,..., e_p;...;...; \quad \forall y = 1,..., g_1;...; \quad \forall y = 1,..., g_r.$$

4. Empirical Study

In this section, for confirming key successful factors of FTZ's port hinterland development empirical examination and analysis are implemented as follows.

4.1. Questionnaire Design, Data Collect and Geometric Average Analysis

The AHP problem is involved the group decision-making, where Robbins suggested five or seven decision-makers are suitable for dealing with group decision-making problem. The respondents are divided into three groups being composed of government official, academics researcher, senior manager of ocean carrier, freight forwarder, and warehouse operator respectively. The author selected some experts list in each group and carried out the survey questionnaire by mail on the August in 2007. A total of 22 valid questionnaires (11 government official, 3 expert academics, and 8 senior managers of ocean carriers containing ocean carrier, freight forwarder and warehouse operator) were collected from the twenty-two respondents, or which return mail ratio represents about 34%of the total Sixty four mails.

Table 5 shows that according to geometric average statistics of assessment dimension and criteria the author discover that the important ranking order of assessment criteria is direct shipping across the Taiwan Strait, frequency of ship sailing and diversification of shipping route, convenience of customs clearance procedures and one-stop administrative service window, efficiency of port operations, integration of customs and port logistics information, adequacy of port logistics facilities, soundness of investment system and incentive measures, exemption or deduction of corporate and local taxes, adequacy of the port hinterland for logistics functions, economic scale of market, efficiency of intermodal transport network, cost of labor, soundness of investment system and incentive measures, cost of land, stability of political climate.

Table 5: Geometric Average Statistics of Assessment Dimension and Criteria

Assessment Dimension	Assessment Criteria	Code	Geometric Average	Rank
Political-Economic Environment A1	stability of political climate	a1	7.13	16
	direct shipping across the Taiwan Strait	a2	8.28	1
	economic scale of market	a3	7.44	10
Operational Environment A2	convenience of customs clearance procedures and one-stop administrative service window	b1	7.85	3
	efficiency of port operations	b2	7.68	4
	integration of customs and port logistics information	b3	7.67	5
Cost Environment A3	cost of labor	c1	7.33	12
	cost of land	c2	7.18	15
	transport and distribution costs	c3	7.26	14
Infrastructure Facility Environment A4	adequacy of port logistics facilities	d1	7.59	6
	adequacy of the port hinterland for logistics functions	d2	7.49	9
	efficiency of intermodal transport network	d3	7.36	11
	frequency of ship sailing and diversification of shipping route	d4	8.04	2
Preferential Incentives Environment A5	soundness of investment system and incentive measures	e1	7.52	7
	exemption or deduction of corporate and local taxes	e2	7.49	8
	exemption and deduction of custom duties and value-added tax for cargo	e3	7.30	13

4.2. FAHP Analysis

For constructing a fuzzy positive reciprocal matrix we used data concerning the relative importance of the 22 valid questionnaires to compile fuzzy pair-wise comparison matrices, and then transformed this data into fuzzy positive reciprocal matrices using the geometric mean method (see Table 6). To calculate fuzzy weights, defuzzifying the fuzzy weights, and normalizing the crisp weights we use geometric means of the triangular fuzzy numbers and the fuzzy weights of the four functional related activities shown in Table 6. The fuzzy weights are defuzzified employing the graded mean integration representation method, yielding the crisp weights, which are then normalized. The results are as shown in Tables 7. For instance, a case of A1 (Political –Economic Environment) geometric Average number is 0.349, 1.128, 3.314; fuzzy weight is 0.023, 0.223,1.919; after defuzzified weight ((0.023+4*0.223+1.919)/6) is 0.472 and Normalized weight (0.472/(0.472+0.473+0.346+0.407+0.472)) is 0.2177. Because of the space limitation, the step by step computations and fuzzy comparison matrices of all the criteria are not show.

Moreover, to calculate integrated weights for each layer we used the same FAHP computation process for each criterion (assessment dimensions and criteria) in two layers to obtain the normalized weights, obtaining the integrated weights for each layer shown in Table 7. For example, weight value of political-economic environment is 0.2177(a), weight value of stability of political climate is 0.363(b), then integrated weight value(c) is 0.0790 based on a simple mathematic formula (a*b=c). according to rank order of integrated weight value are integration of customs and port logistics information(0.0831), efficiency of port operations(0.0811), exemption and deduction of custom duties and value-added tax for cargo(0.0786), stability of political climate(0.0790), economic scale of market(0.0734), soundness of investment system and incentive measures(0.0710), exemption or deduction of corporate and local taxes(0.0676), direct shipping across the Taiwan Strait(0.0651), cost of labour(0.0547),convenience of customs clearance procedures and one-stop administrative service window(0.0539), efficiency of intermodal transport network(0.0530), transport and distribution costs(0.0538), cost of land(0.0512), adequacy of port logistics facilities(0.0482), frequency of ship sailing and diversification of shipping route(0.0465), adequacy of the port hinterland for logistics functions(0.0403).

Table 6: Relevant weight calculations for each assessment dimension

	A1	A2	A3	A4	A5
A1	(1,1,1)	(0.25,1.055,4)	(0.2,0.717,4)	(0.2,0.746,3)	(0.25,0.969,4)
A2	(0.25,0.948,4)	(1,1,1)	(0.2,0.74,4)	(0.2,0.992,3)	(0.25,0.77,4)
A3	(0.25,1.395,5)	(0.25,1.351,5)	(1,1,1)	(0.5,1.651,5)	(0.333,1.681,5)
A4	(0.333,01.34,5)	(0.333,1.008,5)	(0.2,0.606,2)	(1,1,1)	(0.33,0.999,5)
A5	(0.25,1.032,4)	(0.25,1.299,4)	(0.2,0.595,3.03)	(0.2,1.001,3.03)	(1,1,1)
Geometric Average	(0.349,1.128,3.314)	(0.349,1.133,3.314)	(0.276,0.718,2.496)	(0.331,1.041,2.673)	(0.369,1.046,3.314)
Fuzzy Weight	(0.023,0.223,1.919)	(0.023,0.224,1.919)	(0.018,0.142,1.491)	(0.023,0.205,1.596)	(0.024,0.206,1.979)
Defuzzified Weight	0.472	0.473	0.346	0.407	0.472
Normalized Weight	0.2177	0.2181	0.1596	0.1874	0.2172

Table 7: Integrated weigh value of Assessment Criteria for FTZ's Port hinterland

Assessment Dimension	Weight Value(a)	Assessment Criteria	Code	Weight Value(b)	Integrated Weight Value(c=a*b)	Rank
Political-Economic Environment A1 (0.2177)	0.2177	stability of political climate	a1	0.363	0.0790	4
	0.2177	direct shipping across the Taiwan Strait	a2	0.299	0.0651	8
	0.2177	economic scale of market	a3	0.337	0.0734	5
Operational Environment A2 (0.2181)	0.2181	convenience of customs clearance procedures and one-stop administrative service window	b1	0.247	0.0539	10
	0.2181	efficiency of port operations	b2	0.372	0.0811	2
	0.2181	integration of customs and port logistics information	b3	0.381	0.0831	1
Cost Environment A3 (0.1596)	0.1596	cost of labor	c1	0.343	0.0547	9
	0.1596	cost of land	c2	0.321	0.0512	13
	0.1596	transport and distribution costs	c3	0.337	0.0538	12
Infrastructure Facility Environment A4 (0.1874)	0.1874	adequacy of port logistics facilities	d1	0.257	0.0482	14
	0.1874	adequacy of the port hinterland for logistics functions	d2	0.215	0.0403	16
	0.1874	efficiency of intermodal transport network	d3	0.283	0.0530	11

	0.1874	frequency of ship sailing and diversification of shipping route	d4	0.248	0.0465	15
Preferential Incentives Environment A5 (0.2172)	0.2172	soundness of investment system and incentive measures	e1	0.327	0.0710	6
	0.2172	exemption or deduction of corporate and local taxes	e2	0.311	0.0676	7
	0.2172	exemption and deduction of custom duties and value-added tax for cargo	E3	0.362	0.0786	3

5. Conclusions

Many problems have remained unsettlement and need urgent solution to deal with since central government designated four seaports and one airport FTZs, the concept of FTZ's port hinterland is rather ambiguous for the government agencies concerned. For example the exiting functions of FTZ's port hinterland are restricted to the space provided for fabrication, processing, logistics, and trade, etc. This paper aims to clarify multi-functions of port hinterland development, and explore assessment criteria for FTZ's port hinterland development by FAHP approach

Some findings of the paper are summarized as follows:

(1) Port hinterland areas play a complementary role between ship cargo handling areas and local urban districts. In view of the port cluster perspective, FTZ can extend current pure logistic function to multi-functions combined with port hinterland development, such as establishing a variety of parks in the FTZ evolving with international trade and exhibition park, technologic R& D developing park, venture business park, assembling and processing park, waterfront leisure park and academic research Park.

(2) With respect to ranking order of weight, assessment dimensions of FTZ's port hinterland are preferential incentive environment, political-economic environment, operational environment, and infrastructure facility environment. Especially for preferential incentive measures, main preferential incentives of FTZ's port hinterland containing exemption of imported cargo from customs duties, commodity tax, business tax, tobacco and alcohol tax, public health and welfare dues on tobacco products, trade promotion service fees, and harbour service dues. In contrast, FTZ enterprises in Korea can accept not only the aforementioned exemption or reduction of duties and taxes on goods, but also exemption of customs duties on construction materials, exemption of enterprise taxes including corporate tax, income tax, acquisition tax, registration tax, property tax, and composite land tax, and exemption of transportation-related charges. Besides, the central or local governments also provide tenant the funding for technology development and manpower training.

(3) The key successful factors of FTZ's port hinterland development based on ranking order of integrated weight are integration of customs and port logistics information, efficiency of port operations, exemption and deduction of custom duties and value-added tax for cargo, stability of political climate, economic scale of market, soundness of investment system and incentive measures, exemption or deduction of corporate and local taxes, direct shipping across the Taiwan Strait, cost of labour, convenience of customs clearance procedures and one-stop administrative service window, efficiency of intermodal transport network, transport and distribution costs, cost of land, adequacy of port logistics facilities, frequency of ship sailing and diversification of shipping route, adequacy of the port hinterland for logistics functions.

(4) Due to recent political climate between the governments of Taiwan and China getting better, Direct shipping route has started to available for Taiwan, Hong Kong and China-flag ships to navigate after Two government across Taiwan Strait signed MOU of direct sailing by sea transport and air transport in December 2008. Besides, they started to hold a formal dialogue on setting up branch office and business travelling visa based on mutual benefit, thus the prospect of FTZ's port

development seem like brighter in the near future, and Chinese company would become a crucial source of potential tenants for FTZ management authority. .

(5) Customs offices could organize the combined audit taskforces to inspect the autonomous management performed by free trade zone enterprises with regard to control and management, on-line computerized customs clearance and other accounting-related operation, and also conduct physical inventory of the goods involved (Article 19 of the FTZ Act). In other words, customs offices could inspect FTZ enterprises at any time in order to avoiding smuggling or other illegal acts freely. Whereas FTZ is located outside the nation's custom territory, any interference in private business activities from customs officer may decrease the investment willingness of domestic and foreign enterprises. Hence our government should easy exiting inspection regulations and mechanise to keep the liberty of operating activity and management system in the FTZ.

(6) According to FTZ act, any government agency located in or adjacent to an international port has a qualified condition of becoming a FTZ management agency, it arises to the similar problems happening to Port of Taichung and Port of Kaohsiung. As harbour bureaus and economic export zone administrations were located within the same port areas, they have the same right of apply for FTZ establishment. However their marketing targets were similar because of focusing on international logistics service providers and warehouse operators. In the end, two government agencies became competitors and attract domestic and foreign enterprises via same measures (like lower land rent and management fee). We suggested that our government should hold the periodical inter-organizational meeting and request the two administrative agencies to establish cooperative partnerships over promotion for FTZ business.

References

Brenes, E.R., Ruddy, V. and Castro, R. (1997), Free Zones in El Salvador. Journal of Business Research 1997; 38: 57-65.

Centre for Economic Deregulation and Innovation. FTZs in Taiwan 2006. http://www.cepd.gov.tw/.

Chen, S.H. and Hsieh, C.H. (2000), Representation, ranking, distance, and similarity of L-R type fuzzy number and application, Australian Journal of Intelligent Processing Systems 2000; 6(4): 217-229.

Ding, J.F. (2006), Using fuzzy AHP to investigate key factors influencing the quality of service delivery system for port of Kaohsiung, Advanced in Fuzzy Sets and Systems 2006; 1(1): 63-81.

Engman, M., Onodera, O. and Pinali, E. (2007), Export processing zones: past and future role in trade and development. OECD trade policy working paper No.53 2007. http.//www.oecd.org.

Executive Yuan (2004), Act for the establishment and management of free trade zones 2004. http://www.ey.gov.tw/mp?mp=1

Frankel, E.G. (1987), Port planning and development. New York: A Wiley-Interscience Publication 1987.

Ghanem, S. (1997), Local economical development process and government policy-making: an international perspective, Economic Development Review 1997, pp.30-32.

Hsieh, T.Y., Lu, S.T. and Tzeng, G.H. (2004), Fuzzy MCDM approach for planning and design tenders selection in public office buildings, international journal of project management 2004; 22:57u3-584.

Jin, X.S. and Li, X.Y. (2007), a diagnosis on promoting strategies of port hinterland access as a global logistic center. Korean Maritime Institute 2007.

Korea Maritime Institute -- KMI. (2000), A study of free trade zone establishment in the Korea seaport 2000.

Korea Maritime Institute – KMI and Cun-Ang University -- CAU. (2000), An studying on destination and operation of Gwangyang port Custom Free Zone 2000.

Korea Maritime Institute – KMI. (2003), An studying on development direction and prospect of Pyeongtaek port in era of 21 century 2003.

Korea Maritime Institute (2005), Free trade zone and port hinterland development 2005. UNESCAP, http://www.unescap.org.

Li, X.Y. (2006), case studying on port hinterland development in Japan. Korea Marine Institute 2006.

Ministry of Transportations and Communications, Statistics of Free Trade Zone, http://www.motc.gov.tw.

Miyagiwa, K.F. (1986), Reconsideration of the welfare economic of a free trade zones. Journal of International Economics 1986; 21:337-350.
Saaty, T. L. (1980), The Analytic Hierarchy Process. New York: McGraw-Hill Companies; 1980.

National Chen Kung University Institute of transportation and communication management science master essay 2003.

Tseng, L.A. (2004), A studying on Free trade zone development of Kaoshiung port, Kaohsiung Port Authority 2004.

World Bank (1990), Free Trade Zones in export Strategies 1990.

Xu, C.H. (1998), Fuzzy Delphi and AHP approach. Journal of Fuzzy System 1998; 4(1):59-72.

Yang, Q.Q. A studying on influencing factors of FTZ investment inclination for tenant enterprise.

Young, L. (1987), Intermediate good and the formational of duty-free zones", Journal of Development Economics 1987, 25, pp.369-384.

Impacts of port relocation on the location of port-relation industries

Zhongzhen Yang and Ying Li

College of Transportation Management, Dalian Maritime University, Dalian, Liaoning, 116026, China.

Abstract

At present many Chinese seaport cities completely or partly move the port out of the cities. This relocation will certainly change the urban land use pattern, especially land use pattern near the old and new port areas. This study aims to analyze the impacts of port relocation on movement of port-related industries. First, we make a questionnaire survey to know the location behaviors of port-related companies. Second, we develops an industrial location potential model, which considers the location behaviors of the industries in terms of transportation accessibility, land price, aggregated degree of similar and relevant companies, and abundance of service facilities. Then, by taking the relocation of Dalian port as an example, we calculate the relocation behavior of port-related companies with the developed model.

Key words：Port relocation, Industrial location model, Land use pattern

1. Background

Due to the continuous economic growth in China, the number and scale of Chinese ports were increased rapidly in the last 10 years. In 2006, there are 10 ports whose throughputs are over 100 million ton in China. Four of the top ten world container ports are in China. To encourage the economic growth and build international logistics center, many seaport cities actively enhance the investment and construction of port and relevant facilities. In the process, some ports expand on the basis of old ports; some ports abandon the location of old ports and construct modern and advanced ones in new locations. Tianjin port is the representative of the formers. The reason is that Tianjin port is located in the east to its Tanggu District, which is 60km from city center. There are enough coastline and land for development. However, many other ports do not have this advantage and have to abandon old ports and build new ones far from city center, such as Dayaowan port in Dalian, Yangshan port in Shanghai and Yantian port in Shenzhen. Taking Dalian port as an example, it can be seen that Dalian relocated its port to Dayaowan, which is about 30km away from city center, and built dedicated access motorway to Shenyang-Dalian expressway and Dalian-Dandong expressway.

It was well-known that port is a point with intensive production and transportation activities. Many related enterprises tend to be located around the port, and every day large amount of vehicles go into and out the port. In fact, port-related activities are the elements in urban land use system, which is resulted from systematic equilibrium. Thus, the port relocation may break the equilibrium in urban land use system and causes other relocation behaviors, while the relocation may lead to spatial movement of many activities in the short term and bring new equilibrium finally. In the context of the new equilibrium, urban land use system presents a new pattern, and the new pattern will corresponds to a new traffic demand. Therefore, relocation of port will certainly lead to the change of spatial relocation of activities in the whole city and the change of concomitant traffic demand. We think land-use model may be used to analyze the impact of port relocation on the relocation of port-related industries and further on land use system.

Land-use modeling appeared firstly in USA for traffic planning. The most famous land-use model is Lowry model developed in North American in the 1960s, which was formed by equations dealing with the generation and allocation of land use (Helly, 1975). In the model, employees are divided into two categories, namely employees in basic industry and in service or non-basic industry. It is supposed that the activities of service industry are derived from pre-determined employees in basic industry. Thus, after given the total number of employees and population in a region, the spatial distribution of every kind of activity can be determined through solving an array of equations simultaneously, which are established based on spatial interaction relationship and gravity model.

Based on Lowry model, several other models had been developed, e.g. TOMM (Time Oriented Metropolitan Model) took Pittsburgh as the study subject, BASS (Bay Area Simulating Study) took the bay area in San Francisco as study subject, and PLUM (Projective Land Use Model) etc. Among them, Garin-Lowry model has been used successfully to analyze impacts of construction of The Third Airport of London on regional land use. It is found that the construction of a large airport in a region would create a great amount of employment in the basic industry. Due to the inflow of the laborers and the households, the population in the region would increase sequentially. The model revealed that construction of four-runway airport will bring 65,000 employees in the basic industry, and after opening of the airport another 18,000 working posts in the basic industry would be created. Moreover, the total increment in basic industries would bring 42,500 employees in service industry and population increment of 314,000 persons (Batty, 1976).

There are also many other studies on land-use models. For example, Kim (2007) analyzed the impact of new express-way on land-use and population changes in a rural area close to the construction of a new expressway through location modeling with spatial interaction analysis. He considered the relationships among various activities in terms of travel difficulty and contacting demand, and analyzed how a new expressway affects rural areas with location modeling. Sim et al. (2001) assessed whether the regional centre has been successful as an alternative employment center to the CBD (central business district) and has helped to reduce work-travel. The results show that there is great potential in reducing work-travel and reliance on the car so as to alleviate traffic congestion by decentralizing commercial activities from the CBD to the regional centers. The technical note of Yam et al. (2000) develops a multi-linear regression model to forecast traffic generation for large-scale, high-density, multistory public residential housing estates in Hong Kong. Many variables, such as the number of apartments, population, gross floor area, parking spaces, and accessibility, were included in the analysis.

McDonald and Osuji (1995) did an empirical study on residential land values in the vicinity of the new elevated transit line that runs the 11 miles from downtown Chicago to Midway Airport. It is show that in 1990 an increase of 17% in residential land values within one-half mile of the station sites can be attributed to improved access. Ohno (1999) presented a population distribution model, which can estimate change of industrial and population distribution due to the interregional transport development, and can forecast its effect to the nationwide structure. By applying this model to the Shinkansen project in Japan, he looked over the influence of Shinkansen railway on the industrial structure and the generational structure of Japan.

Some other literatures have analyzed the change of land use due to the suburbanization process to the surrounding areas (Ishikawa and Fielding 1998; Rustiadi and Kitamura 1998; Wu 1998; Filon et al. 1999; Gaubatz 1999; Kim et al. 2003). However, it is difficult to find study cases on how the relocation of port specifically changes the urban land use patter and the traffic demand. Due to the function of freight handling of port, port relocation may have more far-reaching impact on city and related sections. This is the reason why we tend to analyze the location behavior of the port-related industries, the changes of land use, and the changes of passenger and freight traffic after the relocation of port in this study.

Therefore, this paper first will classify the industries that closely relate to port business and are influenced by port relocation directly. Second, a Stated Preference (SP) survey is conducted on companies of these industries to study their location behaviors due to the port relocation and a location model for them is established. Third, this paper will analyze the location conditions in the areas surrounding the new and old ports by taking Dalian port as a study subject, and calculates the probabilities of the port-related industries to be located at the new and old port areas. At the same time, sensitivities of location factors on the industries will be analyzed.

2. Port-Related Industries

2.1. Classification of Port-Related Industries

Port-related industry includes companies dealing with businesses related to port production and provide services for port and marine enterprises. The businesses related to port production include marine transportation with vessels and operational production activities for marine transportation with quays and berths. Here, we divide port-related industries into three categories, i.e port enterprise, shipping enterprise and service enterprise etc.

Port enterprise is a company that mainly runs port business. Its business is to transport freight and passengers within the port and provide service for carriers, shippers and passengers.

Shipping enterprises are companies mainly running marine transportation of freight and passengers. Here shipping enterprise means the company dealing with freight transportation, such as Maersk, MSC, COSCO etc.

Serving enterprise is the company providing various services to port and shipping enterprises, such as shipping agency, forwarder, yard company and customs broker etc.

2.2. Relationship between Port-related Industries and the Port

Port enterprise and shipping company provide service for freight owners and passengers through port. They are born from the demand of sea-born freight and passenger transportation, and growth with the increment of the demand. At the same time, port is staring and ending point of shipping lines, while shipping lines are corridors between ports for moving freights and passengers. The ports and shipping lines consist of transportation network covering the whole world.

Because serving companies provide services to port and shipping companies, their businesses certainly relate to the port closely. For example, shipping agency accept the commission from shipping company, ship operator or leaseholder and represent clients to deal with a series of businesses related to vessels in the port, such as the coordination of being used berth or customs declaration for the whole vessel. Freight forwarders accept the commission and represent freight owners to deal with businesses related to marine freight, e.g. storage, checkout, packing, encasements, and transshipment. Yard companies provide services to shipping companies and freight owners, e.g. storage, trailer transportation. From their main businesses, we can know that the port-related enterprises have close relationship with port spatially and naturally. Figure 1 shows the relationships between port and port-related industries.

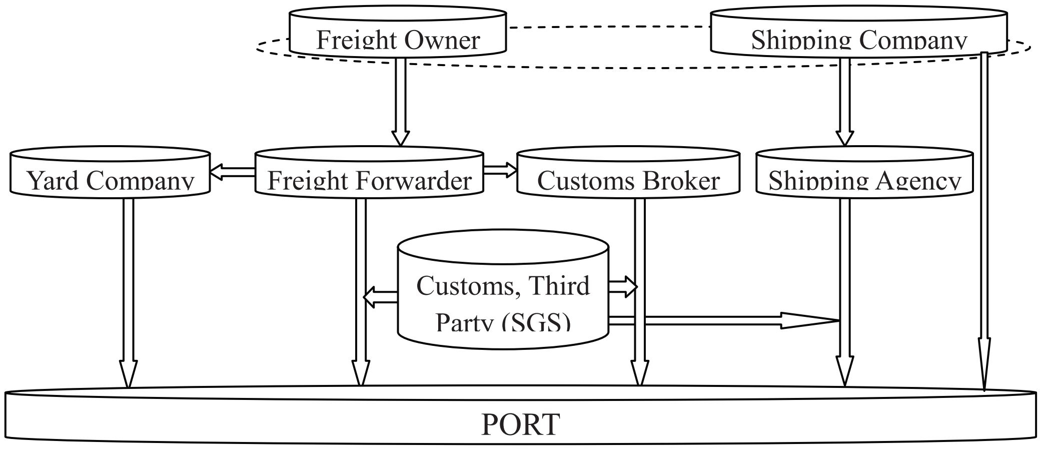


Figure 1: Relationships between Port-Related Industries

2.3. Location of Port-Related Industries

The above analyses show that the businesses of port-related industries are closely related to port spatially and naturally. To save contacting time and cost needed for the businesses, these companies generally tend to locate their offices at the sites around the port. As the result, lots of port-related companies have gotten together around the port. Moreover, due to the aggregated effect, more and more port-related companies will further concentrate around the port to form a densely located region.

For example, in Dalian there are about 40 port companies, 40 shipping companies, 24 container yard companies, 940 freight forwarders, 320 shipping agencies, and 100 customs brokers. Most of them are located in Zhongshan District where is near the old port. Figure 2 shows the distribution of port-related industries in Dalian. It can be seen that almost all shipping companies are located at Zhongshan District, and other types of companies are distributed in every district, but most of them are in Zhongshan District.

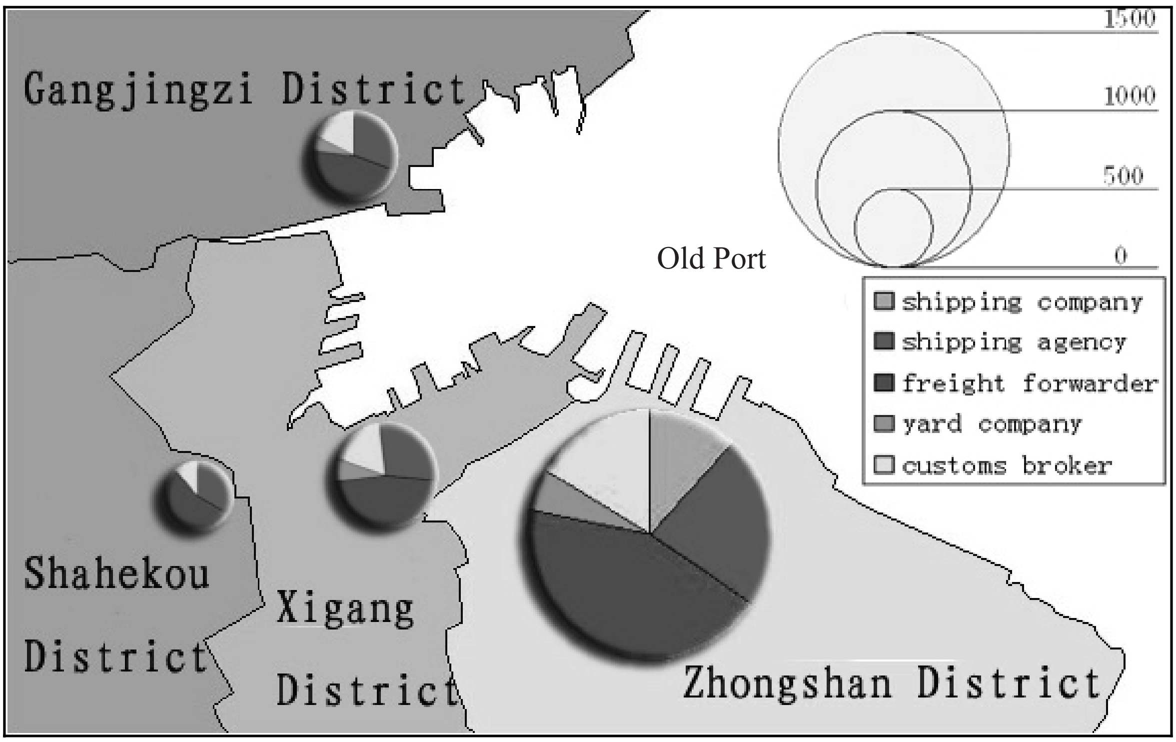


Figure 2: Locations of Port-Related Companies in Dalian

3. Location Model of Port-Related Companies

As mentioned above, in 2001, Dayaowan port (new Dalian port) was opened, which means that port business in Dalian would be moved to the new port gradually. As the result, in 2006, the business of all container transportation and most other transportation in the old port were stopped and moved to the new port. In 2007, the old port area (where almost non production was left) was planned for a re-development. The aim of the re-development is to build the evacuated land into a commercial, leisure, sight seeing and housing zone.

It is found that in Dalian as the relocation of the port most container yard companies have moved to the new port area, while part of shipping agencies, freight forwarders began to set up offices in the new port area or move to new port area completely. Based on Dalian case, we can rationally say that as the development of the new port area, the impacts of the port relocation on the immigration of the port-related industries will be more obviously. Therefore, it is necessary to analyze the impacts of port relocation quantitatively with a land-use model.

3.1. Model Structure

It is commonly believed that an enterprise would choose its location based on the utility offered by the site. According to the stochastic utility theory, the larger the offered utility, the higher the probability a site will be located. Since the above mentioned port-related industries are not manufacturing plant, here we take accessibility to port, rent of office, aggregated degree of similar and relevant companies, and abundance of service facilities as the main factors to determined the location utility. Thus the utility function of a port-related company in a site can be described with Eq.(1).

$$F_{ij} = \alpha A_j + \beta D_j + \gamma S_j - \lambda R_j \qquad (1)$$

Here, F_{ij}= the utility when the company i choose site j as office location; A_j= the accessibility to port at site j; R_j= the rent level of office buildings at site j; D_j= the aggregated degree of similar and relevant companies at site j; S_j= the abundance of service facilities at site j; i $(i=1,2,3,4,5)$ = the type of port-related companies, including shipping company, shipping agency, freight forwarder, yard company and customs broker; $\alpha, \beta, \gamma, \lambda$ = parameters needed to be calibrated. Then among many sites, the probability of company i selecting site j as location can be calculated with Eq.(2), where $k=1,..,K$ is the location alternatives.

$$P_{ij} = F_{ij} / \sum_k F_{ik} \qquad (2)$$

Relocation of port has directly changed the accessibilities of the whole city. As the result, utility of port-related company in a site may get worse or better. Thus, in order to reduce the travel cost and improve working efficiency, the company may consider moving its businesses partially or completely to the sites with better accessibility. If let j (or k) in Eq.(1) and Eq.(2) to represent the sites around the new and old port areas respectively, then the probability of a port-related company moving from old location to new one can be calculated with the equations.

3.2. Calibration of Parameters

A generalized method to calibrate parameters in Eq.(1) is: a) collecting SP (Stated Preference) or RP (Reveal Preference) data of the results of decision-makers; b) calculating the ratios of ever kind of decision makers to select various alternatives; c) substituting Eq.(1) into Eq.(2) and calibrating parameter with the combined equation through regression method (Ben and Lerman, 1985). However, because the cases of port relocation and the relocation of the port-related companies are very few currently, we can not calibrate $\alpha, \beta, \gamma, \lambda$ through regression analysis. Thus, here we develop a calibration method based on Comprehensive Granting Weight method. Details of the method are as follows (Guo, 2002):

Step 0: Let $x_1^i, x_2^i \Lambda\ x_j^i \Lambda\ x_m^i$ represent the samples containing m indices in group i; Here, m (1, 2, 3 ,4) represents accessibility, aggregated degree of similar and relevant companies, the abundance of service facilities, and the rent level of office buildings respectively, i (1, 2, 3, 4, 5) represents shipping company, shipping agency, freight forwarder, yard company and customs broker respectively. For example, x_1^1 is the accessibility to port of a shipping company.

Step 1: Weight with "Difference Animating" method to obtain:

$$w_j^{i,c} = s_j^i / \sum_{k=1}^{m} s_k^i \tag{3}$$

Here, $s_j^i = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_j^i - \overline{x}_j)^2}$, $\overline{x}_j^i = \frac{1}{n}\sum_{j=1}^{m} x_j^i$, and $w_j^{i,c}$ = the weight of index j on the utility of port-related industry i determined with "Difference Animating" method, s_j^i and $\overline{x}_j^i$ are the sample deviation and mean of x_j^i respectively.

Step 2: Weight with "Function Animating" method to obtain:

$$w_j^{i,g} = \overline{x}_j^i / \sum_{k=1}^{m} \overline{x}_k^i \tag{4}$$

Here, $w_j^{i,g}$ is the weight of index j on the utility of port-related industry i determined with "Function Animating" method.

Step 3: Comprehensive weighting. We can get weight by integrating the weights from "Difference Animating" and "Function Animating" methods to obtain: $w_j^i = k_1 w_j^{i,c} + k_2 w_j^{i,g}$. Here k_1, k_2 can be obtained by solving the following optimization problem.

$$\begin{aligned} &Max \quad \sum_{i=1}^{n}\sum_{j=1}^{m}(k_1 w_j^{i,c} + k_2 w_j^{i,g})x_j^i \\ &S.T \quad k_1 > 0, k_2 > 0 \\ &\qquad k_1 + k_2 = 1 \end{aligned} \tag{5}$$

The solutions are as follows：

$$k_1 = \frac{\left(\sum_{i=1}^{n}\sum_{j=1}^{m} w_j^{i,c} x_j^i\right)^2}{\left\{\left(\sum_{i=1}^{n}\sum_{j=1}^{m} w_j^{i,c} x_j^i\right)^2 + \left(\sum_{i=1}^{n}\sum_{j=1}^{m} w_j^{i,g} x_j^i\right)^2\right\}}$$

$$k_2 = \frac{\left(\sum_{i=1}^{n}\sum_{j=1}^{m} w_j^{i,g} x_j^i\right)^2}{\left\{\left(\sum_{i=1}^{n}\sum_{j=1}^{m} w_j^{i,c} x_j^i\right)^2 + \left(\sum_{i=1}^{n}\sum_{j=1}^{m} w_j^{i,g} x_j^i\right)^2\right\}} \quad (6)$$

3.3. Transportation Accessibility

Here the transportation accessibility refers to convenience degree from a given location to an activity site by a travel mode. It reflects the difficulty and easiness of spatial entities overcoming distance obstacles to trade with each others and expresses the relationship between the spatial entities (Kim et al., 2002). Its mathematical meaning can be explained as: the accessibility from i to j is proportional to the scale of land-use activities (employment, population, business service, etc.) in site *j*, and inversely proportional to travel impedance between the two sites, that is:

$$A_i = \sum_j G_j^k / Z(t_{ij}) \quad (7)$$

Here, G_j^k =the scale of *k* type land use in site *j*, *j=1，2…，n*；$Z(t_{ij})$ = the travel impedance function.

3.4. Aggregated Degree of Port-related Industries

The aggregated degree reflects the gathering degree of enterprises in a certain area. It can be defined as:

$$D = \sum_{i \neq j} (T_i \times T_j) / d_{ij} \quad (8)$$

Here, D = the aggregated degree of port-related companies in an area; T_i, T_j = the scale of the companies in site *i*, *j* respectively; d_{ij} = the travel distance between sites *i*, *j*.

3.5. Abundance of Service Facilities

The abundance of service facilities reflects the variety and quantity of service facilities in a region. For port-related companies, we use the ratio of operational area of hotels (office building), restaurant, living and entertainment facilities to the total area to represent this index, and the function is as follows:

$$S_i = m_i / M_i \quad (9)$$

Here, S_i = the abundance degree of service facilities within a region; m_i = the operational area of hotels, restaurant, living and entertainment facilities within the region; and M_i = total area of the region.

3.6. Rental Level of Office Building

The rent level of office building reflects rental cost of office building in a region. Here we use the average rental price of main office buildings to represent this index.

4. Sensitive Analyses

By taking Dalian as a case study, we did a SP survey and a sensitivity analysis based on the collected data. The detailed works are as follows.

4.1. Data Collection

We made a survey on port-related companies in Dalian, which consist of face-to-face interview, mail survey, e-mail survey and telephone interview. The face-to-face interview concerns 17 companies. Mail and e-mail survey sheets were sent to 160 and 200 companies respectively, and 20% and 27.5% companies answered the questions respectively. Table 1 shows the number of the actual samples and the questionnaire sheet is shown in Table 2.

Table 1: Number of each kind of Samples

Shipping Company	Shipping Agency	Freight Forwarder	Yard Company	Customs Broker	Total
13	40	43	10	16	122

Table 2: Part of the Questionnaire Sheet

Factors affecting Location Choice	Importance					Satisfaction				
Mark the following factors with rank 1-5 in terms of their importance and your satisfaction degree.	Not Very	Not	Average	Normal	Very	Not Very	Not	Average	Normal	Very
Travel distance to the port	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Travel time to the port	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Travel distance to some authorities (such as customer)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Housing Price near the office	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Rent level of the Office buildings	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Abundance of service facilities	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Abundance of hotels	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Aggregated degree of same companies	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Aggregated degree of relevant companies	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

4.2. Determination of Parameters and Other Variables

4.2.1. Calibration of Parameters with Weighting Method

We can estimate the weights of transportation accessibility, the level of rental price of office building, aggregated degree of similar and relevant companies and abundance of service facilities in Eq.(1) with Eq. (3), (4), (6) based on the surveyed data. Then the calibrated $\alpha, \beta, \gamma, \lambda$ are obtained in Table 3.

Table 3: $\alpha, \beta, \gamma, \lambda$ of each kind of Company

	α	β	γ	λ
Shipping Company	0.138	0.324	0.427	0.111
Shipping Agency	0.343	0.279	0.122	0.256
Freight Forwarder	0.311	0.284	0.132	0.275
Yard Company	0.839	0.058	0.051	0.053
Customs Broker	0.309	0.284	0.156	0.252

It can be seen that shipping companies have low demand for transportation accessibility (**0.138**). This may be because that they have little businesses to be done in the port. The shipping companies also do not care about the rent price of office building (**0.111**). This indicates that they tend to choose top-grade office buildings as working place to display their status. However, they are sensitive to the abundance of service facilities (**0.427**). Shipping agencies, freight forwarders and customs brokers all pay much attention to the accessibility (**about 0.3**), this may because that they have to go to the port to do many businesses. They also attach much importance to aggregated degree of similar and relevant companies (about **0.3**). Because they are small companies, the shipping agencies are sensitive to rent price of office building (**0.25**). However, they have low demand for abundance of service facilities (**0.12~0.15**). Due to providing special service, yard companies generally tend to be located near the ports. This can be verified by the fact that they concern accessibility seriously (about **0.839**) and do not care about other factors.

4.2.2. Transportation Accessibility to Dalian Port

The planning area of Dayaowan Port is 6.88 square kilometers. It takes about 50 minutes for a car from city center to Dayaowan port. With Eq. (7), we can calculate the accessibilities A_1 and A_2 to the new and old port areas respectively. The results are A_1=8.01 and A_2=0. The reason for A_2=0 is that because there is no freight operation in the old port, the port-related companies can not do any business in the old port.

4.2.3. Aggregated Degree of Port-Related Companies

We measure the scale of a company in terms of the registered capital and the number of employees. The distances between companies are from a GIS database. Putting above data into Eq.(8) , we obtain D_1=2,661,682 and D_2=14,678,265 for the new and old port areas respectively.

4.2.4. Abundance of Service Facilities

We obtained the operating area of service facilities by making a survey on hotels, entertainment facilities and restaurants in Dalian. Then with Eq.(9) we got S_1=3.24, S_2=21.33 for the new and old port areas respectively.

4.2.5. The Rental Price of Office Building

There are about 100 office buildings in central Dalian, which are mainly located in Zhongshan, Xigang and Shahekou districts. The average rental price is 2.06 yuan/m^2/day. Office buildings in Zhongshan district account for 60%~80%. The average rental price is 2.56 yuan/m^2/day. At present, the average rental price in Development Zone, which is near the new port, is 2.14 yuan/m^2/day. So we can think R_1 and R_2 are 2.14 and 2.56 respectively.

4.3. Relocation Probability

Since we determined the $\alpha, \beta, \gamma, \lambda$ with Weighting method rather than statistic regression, the above variable values can not be input into Eq.(1) for calculation directly. We must normalize them to the values between 0-1. For example, the normalization of transportation accessibility is as: $A_1^{'} = A_1 / (A_1 + A_2)$, $A_2^{'} = A_2 / (A_1 + A_2)$. Inputting the normalized variables into Eq.(1)-(2), we can get the probabilities of the port-related companies to choose the areas around the new and old ports under current condition as shown in Table 4.

Table 4: Probability of Choosing New Port and Old Port Areas

	Shipping Company	Shipping Agency	Freight Forwarder	Yard Company	Customs Broker
P_{i1}	0.24	0.58	0.54	0.93	0.52
P_{i2}	0.76	0.42	0.46	0.07	0.48

It can be seen that the probability of yard companies choosing the new port area is the highest (0.93). This is because that their businesses relate to port greatly. While, the probability of shipping companies choosing the new port area is the lowest (0.24), this is because the working condition and the abundance of service facilities around the new port are much worse than that in the city center (the area around old port) and the companies have a little businesses that needs them to go to the port. The probabilities of the other industries choosing new port area are also high (0.58, 0.54, and 0.52).

Except shipping company, the above result is basically consistent with real trend. Through face-to-face interview, we found that in addition to the above location factors some shipping companies pay great attention to the level of top-grade facilities near their working site to show their strength and status. In order to reflect this phenomenon, we add the factor of the level of top-grade facilities (L) for the location choice of shipping company. With surveyed data, we got to know L_1=0.15 and L_2=0.85 respectively, and the weight of L is 0.4. Then with the new method, the changed probabilities of shipping company choosing new port and old port areas are P_{i1}= 0.12 and P_{i2}=0.88 respectively.

The real situation is that most yard companies in Dalian have moved to the new port area, while few of shipping companies had the mind of relocation. However, part of shipping agencies and freight forwarders plan to move to new port area or set up office in new port area in short term. This fact shows that our analysis method is with good accuracy.

5. Conclusions

Port relocation has a great impact on urban land use and traffic demand. We think that the being first impacted industries are the port-related enterprises because they have to do a lot of port-related businesses. Thus relocation of port may change the location of the port-related enterprises first, and the location changes of port and its related enterprises will affect the land use system in the whole city subsequently. During the breaking of the old equilibrium and formation of the new equilibrium in land use system, locations of many bodies will change and then the traffic demand corresponding to the system will also change. This paper develops a location model of port-related enterprises by taking the relocation of Dalian port as an example. In order to formulate the model, port-related enterprises are classified and their location behaviors are analyzed based on the surveyed data.

Because Dalian port was moved outside only for 5-year, at present the relocation case of port-related enterprises is relatively few. Thus we have to use the SP data to do the analyses and the calculated results are the predicted ones. In fact, the impacts of the port relocation on urban land use may continue a long period and proceed in escalated style. It means that the port relocation will first induce the relocation of some port-related enterprises, and then the relocation will cause the demand on other industries (like service industry). As the result, other industries such as office buildings, restaurants and banks will appear around the new port area. The increment of other industries may further change the location factors around the new port area and increase the location merit of the new port area. Subsequently, the port-related enterprises will further enter into the new port area. On the other side, as the new port area gets crowded, land price will rise to control the amount and the speed of the industrial immigration.

The above process continues and repeats to reach equilibrium, thus a new land use pattern in the whole city will appear in a relative long period after the open of the new port. The model developed here is only one part of the whole method to analyze the land use change induced by port relocation. It offers a tool to analyze the relocation behaviors of port-related industries and set a good base for the whole analysis system. Our further study will add systematic analytical function to simulate the location behaviors of all bodies in the whole city based on this model. Then we can forecast the change pattern of land use system due to the port relocation and offer basic information for urban planning and urban transportation planning.

References

Batty, M. (1976), Urban modelling: algorithms, calibrations, predictions, Cambridge University Press, Cambridge.

Ben, A.M. and Lerman, S.R. (1985), Discrete Choice Analysis: Theory and Application to Travel Demand, the MIT Press, Cambridge, Massachusetts.

Filon, P., Bunting, T., and Warriner, K. (1999), The entrenchment of urban dispersion: residential preferences and location patterns in the dispersed city, Urban Study, 36(8), pp.1317-1347.

Gaubatz, P. (1999), China's urban transformation: patterns and process of morphological change in Beijing Shanghai and Guangzhou, Urban Study, 36(9), pp.1495-1521.

Guo, Y.J. (2002), Comprehensive Assessment Theory and Method, Sciences Press, Beijing.

Helly, B.(1975), Urban System Models, Academic Press, New York.

Ishikawa, Y. and Fielding, A.J. (1998), Explaining the recent migration trends of the Tokyo metropolitan area, Envionmental Planning A, 30, pp.1797-1814.

McDonald, J.F. and Osuji, C.I. (1995), The effect of anticipated transportation improvement, Regional Science and Urban Economics, 25, pp.261-278.

Kim, D.S., Mizuno, K., and Kobayashi, S.(2002), Modeling urbanization by population potential considering greenbelt effect and various accessibility measurement methods, Journal of Rural Planning Association, 20(4), pp.253-264.

Kim, D.S., Mizuno, K., and Kobayashi, S. (2003), Modeling urbanization by accessibility in rapid growth areas, Journal of Urban Planning, Development, 129(1), pp.45–63.

Kim, D.S. (2007), Location modeling of population and land-use change in rural area by new expressway, Journal of Urban Planning and Development, ASCE, September, 201-208.

Ohno, E. (1999), Population distribution change due to the Shinkansen project in Japan, Journal of the Eastern Asia Society for Transportation Studies, 3(4), pp.165-180.

Yam Richard, C.M., Whitfield Richard, C., and Chung Raymond, W.F. (2000), Forecasting traffic generation in public housing estates, Journal of Transportation Engineering, July/August, pp.358-361.

Rustiadi, E. and Kitamura, T. (1998), Analysis of land use changes in city suburbs: a case study on some subdistricts of the Bekasi area of west Java, Indonesia, Journal of Rural Planning Association, 17(1), pp.20-31.

Sim, L.L., Malone-Lee, L.C. and Chin, K.H.L. (2001), Integrating land use and transport planning to reduce work-related travel: a case study of Tampines regional centre in Singapore, Habitat International, 25(3), pp. 399-414.

Wu, F. (1998), Polycentric urban development and land-use change in a transitional economy: the case of Guangzhou, Environmental Planning A, 30, pp.1077-1100.

Modal splits in rapid passenger transportation corridor

Zhongzhen Yang, Ying Li

College of Transportation Management, Dalian Maritime University, Dalian, Liaoning, 116026, China.

Abstract

To analyze the modal splits and to make good use of the resources of rapid passenger transportation facilities which consists of airline (Air) and high speed railway (HSR), this research makes a state preference (SP) survey on the passengers who use flight for travel in a corridor, and with the collected data a Logit model is developed to estimate the modal splits of the rapid travel modes in the corridor. Moreover, sensitivities of high-speed travel demand to attributes of travel modes, such as needed time, ticket price and service frequency of Air and HSR, are analyzed. Then, by taking the being constructed HSR between Dalian and Harbin in China's North-East area as an example, operational strategies are given to the rapid transportation operators in the context of competition-cooperation theory. The given strategies may optimize the relationship of competition and cooperation of the two kinds of high speed transportation facilities.

Keywords: High-speed railway, Air transport, Logit model, Sensitivity analysis

1. Introduction

With the fast social and economic development in China, the demand for high-speed passenger transportation increases rapidly, which results in the rapid development of civil aviation, expressway and HSR in China. During the period, large amount of flight-equipments are purchased by airline companies and many airports and airlines are opened. After the six times increments, the operation speed of national railway in some lines have reached 200*km/h* and many high-speed passenger railways are being built or planned. It is can be predicted that in the near future there will be several alternative modes in one corridor as the development of airway, expressway and HSR. For example, there will be flight, expressway and HSR (designed speed: 350 *km/h*) in the corridor between Harbin and Dalian after the construction of Har-Da HSR. As the increment of the alternatives, the original market scale of flight and expressway may decrease (He et al., 2006). This kind of situation has happened in several other countries (Table 1). Thus, It is necessary to study the competition and modal splits between the modes in the same high speed passenger transportation corridor, which may support the operators to take competition-cooperation measures to improve the service and increase market scale to obtain the balanced and sustainable development of high speed transportation (Milan, 1996; Nuzzolo et al., 2000; Daniel, 2002; Si & Gao, 2005).

Here, under the condition of supposing a HSR is opened between Dalian and Harbin, we make SP questionnaire survey on flight passengers and with the data analyze the utilities of HSR and the flight and the relationship between them. Further, modal splits are estimated and sensitivities of the needed time, ticket price and service frequency of the HSR on the modal splits are analyzed. Finally, the obtained learning is used to optimize the transportation resources through integration and cooperation method.

Table 1: Effects of the HSR on the Flight Mode

HSR (O-D)	Distance	Changed Share of Flight Mode
France TGV (Paris-Lyons)	450 km	From 30% down to 15%
France TGV (Paris - Marseilles)	700 km	From 45-55% down to 35-45%
France TGV (Paris -Nice)	900 km	From 55-65% down to 50-60%
Span AVE (Madrid - Seville)	470 km	From 40% down to 13%
Korea KTX (Seoul - Taego)	293 km	Decreased by 72%

2. Model and Methodology

SP survey is wildly used in modal split studies because it can simulate the choice behavior of decision makers based on the relative utility of each alternative (Piao,1993). For example, it can be used to analyze the choice behavior between car and transit, choice behavior among different paths.

2.1. Design of SP Survey

SP survey here contains two parts: basic information of samplers and their choice behaviors of travel modes. The former means sex, age, trip purpose, burden style of the payment, modes and time for access. For the survey, first a set of attributes of flight and HSR, which serve the same corridor, are given, then change some attributes of the being constructed HSR and let the samplers to choose the travel modes again. Because the study topics are the high speed transportation, the passengers waiting at the airport are selected for the survey. During the survey, three factors of each mode are used. They are the needed time, ticket price and service frequency, while the needed time contains the access/egress time and the travel time.

2.2. Stochastic Utility Model

Stochastic utility model (Eq. (1)) is widely used to analyze the sensitivities of the attributes of travel modes on the modal splits of travelers in a fixed pair of sites (Milan, 1996; Park & Ha, 2006).

$$U_i = \sum_{im} \beta_{im} A_{im} + \beta_a R_{Air} + \beta_h R_{Highway} + \beta_r R_{HSR} + \varepsilon_i \qquad (1)$$

Where, U_i = stochastic utility; $A_\bullet$ = attributes of modes; $\beta_\bullet$ = weights of the attributes; $R_{Air}, R_{Highway}$ and R_{HSR} = dummy variables to express the specific preference of travelers for air and HSR modes respectively; ε_i = stochastic factor; *i* and *m* represent travel modes and mode attributes.

Here we try to study the modal splits in a corridor rather than a fixed pair of sites to analyze the common mode choice behaviors in several pair of sites along the corridor. Therefore, a detailed stochastic utility model is built to obtain Eq.(2). And Eq.(2) will be calibrated with data collected by the SP survey.

$$\begin{aligned} U_i &= \beta_1 \frac{L_{in}}{V_{in}} + \beta_2 \frac{M_{in}}{L_{in}} + \beta_3 F_{in} + \beta_4 R_{Air} + \beta_5 R_{Highway} + \beta_6 R_{HSR} + \varepsilon_i \\ &= \overline{\beta_1} V_i^{-1} + \overline{\beta_2} C_i + \overline{\beta_3} F_i + \overline{\beta_4} R_{Air} + \overline{\beta_5} R_{Highway} + \overline{\beta_6} R_{HSR} + \varepsilon_i \end{aligned} \qquad (2)$$

Where, $L_\bullet$ = distance between a pair of sites (*km*); $V_\bullet$ = running speed *(km/m)*; $M_\bullet$ = ticket price,

$C_\bullet$ = fare (*yuan·km*$^{-1}$), i.e. $C_\bullet = M_\bullet / L_\bullet$. Because different pair of sites has different needed time and cost, $V_\bullet$ and $C_\bullet$ are used to replace the needed time and cost in Eq.(2); $F_\bullet$ = service frequency of a mode (*h*); n= number of the pair.

If ε_i obeys Gumbel distribution, the probabilities of modes being selected can be calculated with Eq.(3) (Gerken,1991):

$$P_i = \frac{\exp(U_I)}{\sum_j \exp(U_J)} \tag{3}$$

Putting Eq.(2) into Eq.(3), we can get a more detailed equation of the probability.

$$P_i = \frac{\exp(\overline{\beta_1} V_i^{-1} + \overline{\beta_2} C_i + \overline{\beta_3} F_i + \overline{\beta_4} R_{Air} + \overline{\beta_5} R_{Highway} + \overline{\beta_6} R_{HSR})}{\sum_j \exp(\overline{\beta_1} V_j^{-1} + \overline{\beta_2} C_j + \overline{\beta_3} F_j + \overline{\beta_4} R_{Air} + \overline{\beta_5} R_{Highway} + \overline{\beta_6} R_{HSR})} \tag{4}$$

The model splits represent the competing relationship among alternative modes in the same corridor. By understanding the modal splits, we can make an operation plan to optimize the resource of the whole high-speed transportation system. Moreover, in order to make the high speed mode obtain a maximum market share, it is also necessary to arrange the operation of the high speed mode optimally. Thus, the interaction among the high speed modes in the same corridor is studied with elasticity demand analysis method as follows.

In elasticity demand analysis, the changes of modal splits are analyzed in the context of changing the running speed, ticket price and service frequency by 1%. For a single passenger *k*, the demand elasticity of mode *i* is defined as follows:

$$E_{x_{jmk}}^{p_k(i)} = -\Delta P_k(j) x_{jmk} \beta_m / \Delta x_{jmk} \tag{5}$$

Where, $\Delta P_k(j)$ = change of choice probability of mode *j*; x_{jmk} = attribute m of mode *j*. If $i = j$, Eq.(5) means the direct elasticity; otherwise, it means the cross one.

The change of demand ΔM can be calculated with the *E* and the total demand *M*. Then the impact of attribute m of mode *j* on demand is as follows:

$$\Delta M_{jm} = \sum_j M P_j E_{x_{jm}}^{p(i)} \tag{6}$$

Eq.(4) represents the competing relationship among several modes, while Eq.(6) represents their cooperating relationship. With them, we can optimize the modal splits in a high speed corridor through adjusting some mode attributes, and further to make full use of the total resources.

3. Case Study

Figure 1 shows the spatial locations of Harbin-Dalian corridor and Har-Da HSR, which are used for the case study. There are 4 major cities along the corridor: DaLian, ShenYang, ChangChun and Harbin. The distances between DaLian and the others are 400km, 700km and 946km. It is believed that the influence of HSR on high-speed transportation demand changes according to the travel distance (Park & Ha, 2006), i.e. for travel between 200-500km, the HSR has an absolute advantage; while for travel between 500-1200km, the flight and HSR compete each other fiercely.

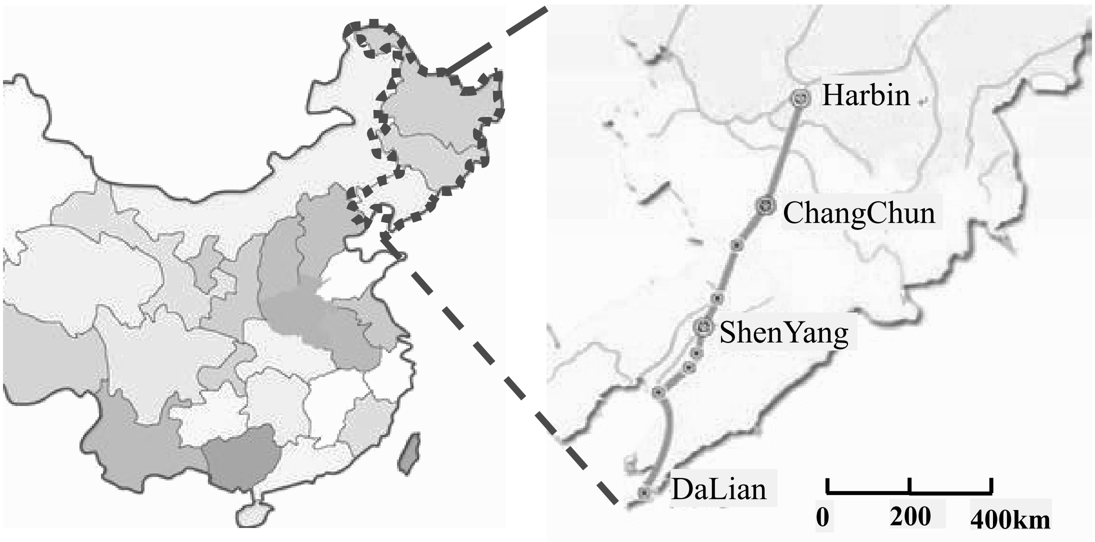


Figure 1: Spatial Locations of Har-Da HSR

Because the railway and expressway services are very convenient and fast between DaLian-ShenYang, the flight has weak competing ability and the frequency is only 3 flights every two days. As the results, we just analyze the sensitivities of passengers on the mode attributes in terms of trips of DaLian-ChangChun, and DaLian-Harbin. Because the access/egress time of all modes is almost same, they are omitted here. Because there are only a few passengers choosing the expressway, here we just take the flight and HSR into account in sensitivity analysis.

A SP survey was carried out on the passengers who waited at DaLian airport and went to Harbin and ChangChun respectively. In the questionnaire sheet, several kinds of ticket price and service frequency of HSR are given. For example, for Dalian-ChanChun, the picket prices of 350, 400 and 450 RMB are supposed, while the service frequency of 1 train per 2 hours and 1 train per 3 hours are supposed. First, the passengers are asked to answer the questions facing a given set of attributes shown in Table 2. Then attributes of HSR are changed as stated above sequentially for passengers to answer in the new contexts. Totally, there 65 passengers form DaLian to ChangChun and 78 passengers from DaLian to Harbin are interviewed.

Table 2: Initial Mode Attributes of the Survey (DaLian-ChangChun)

Mode	Needed Time	Ticket Price	Frequency
Air	1.25hr.	580RMB	1time/9hr.
HSR	3.5hr.	***350 RMB***	***1 time/2hr.***
Highway	8.5hr.	200 RMB	1 time/2hr.

3.1. Data Analysis

For survey for DaLian-ChangChun, 65% of the surveyed passengers are male, and for survey for DaLian-Harbin, 59% of the surveyed passengers are male. The situations of ages, trip purposes and burden styles of the payment are shown in Table 3, Table 4 and Table 5. The modes and access time are listed in Table 6 and Figure 2 respectively.

Table 3: Ages of the surveyed Passengers

Ages	21-30	31-40	41-50	51-60	Others
DaLian-ChangChun	30.77%	27.69%	30.77%	7.69%	3.08%
DaLian-Harbin	23.07%	41.03%	21.80%	14.10%	0

Table 4: Trip Purposes of the surveyed Passengers

Purpose	Business	Leisure	Others
DaLian-ChangChun	55.38%	24.62%	2.00%
DaLian-Harbin	47.44%	24.36%	28.20%

Table 5: Pay Styles of the surveyed Passengers

Employer Pay	Government	Public Institution	Super-sized Enterprise	Large-sized Enterprise	Medium-sized Enterprise	Small-sized Enterprise
DaLian-ChangChun	8.12%	14.96%	15.38%	13.85%	6.15%	1.54%
DaLian-Harbin	8.15%	12.36%	6.41%	11.54%	8.97%	1.28%
Private Pay (Income: RMB/m)	<2000	2001-3000	3001-4000	4001-5000	5001-6000	>6001
DaLian-ChangChun	12.31%	4.62%	4.62%	3.08%	6.15%	9.23%
DaLian-Harbin	8.97%	15.38%	1.28%	3.85%	7.69%	14.10%

Table 6: Access Modes of the surveyed Passengers

Access Modes	Private Car	Taxi	Bus	Business Car	Airport Bus	Flight
DaLian-ChangChun	19.05%	65.08%	3.17%	6.35%	4.76%	1.59%
DaLian-Harbin	41.33%	37.33%	2.67%	2.67%	4.00%	12.00%

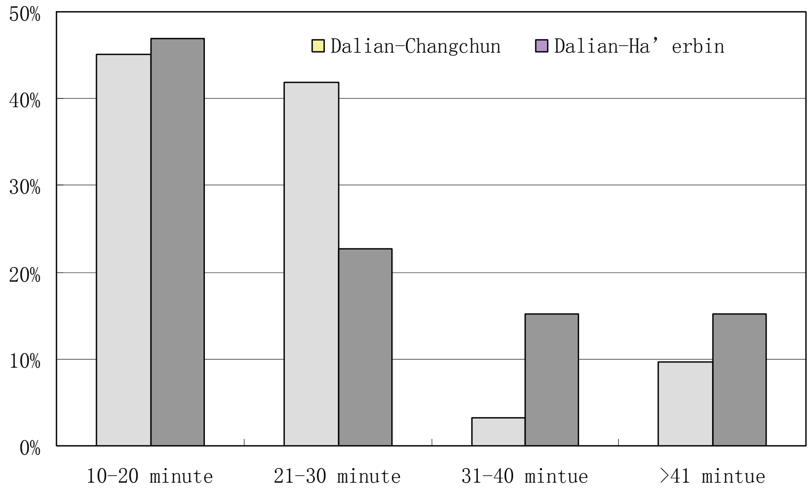


Figure 2: Access Time of the surveyed Passengers

3.2. Calibration of the Utility Model

With the surveyed data, the utility model is calibrated to get Table 7 and Eq.(7). It shows that at 95% significance level the statistical indicators are satisfied.

Table 7: Calibrated Results of the Utility Model

	Coefficient	***t*-Value**	**R^2 and Adjust R^2**
Running Speed	-1.31298	-4.62	
Unit Fare	-4.097011	-11.64	
Service Frequency	-0.101786	-2.82	$R^2 = 0.9494$
Dummy Variable (R_{Air})	0.436151	5.23	$\overline{R}^2 = 0.9305$
Dummy Variable (R_{HSR})	1.535649	6.36	

$$U_i = -1.31298 V_i^{-1} - 4.097011 C_i - 0.101786 F_i + 0.436151 R_{Air} + 1.535649 R_{HSR} \quad (7)$$

It can be seen that the utility of a mode will increase if running speed is increased or fair and service frequency are decreased, due to the coefficients of them are negative ones. The utilities of air and HSR (U_i) can be calculated with a given set of attributes of modes shown in Table 8, and then the probability (P_i) can be obtained with Eq.(4). We also analyzed the demand change in the context of changing the mode attributes by 1% to get Table 9. It can be sent that if flight speed increases 1%, flight demand is increased 2.88% and HSR demand is decreased 0.89%. Generally, the elasticity of

F_i is much less than that of V_i and C_i, it means that the passengers are more sensitive to ticket price and needed time and the travel but are less influenced by service frequency.

Table 8: Utilities and Probabilities of Modes

	V_i (*km/min*)	C_i (*RMB/km*)	F_i (*service/h*)	U_i	P_i (%)
Air	9.3333	0.9857	9	-3.5596	0.2330
HSR	2.9167	0.5000	3	-2.3679	0.7670

Table 9: Results of Sensitivity Analysis of Travel Demand Elasticity

	V_{Air} (+1%)	V_{Air} (-1%)	V_{HSR} (+1%)	V_{HSR} (-1%)	C_{Air} (+1%)	C_{Air} (-1%)
E_{Air}	2.8761%	-2.8739%	-2.8432%	2.9076%	-2.9538%	2.9584%
E_{HTR}	-0.8899%	0.9072%	0.8974%	-0.8995%	1.4835%	-1.5158%
	C_{HSR} (+1%)	C_{HSR} (-1%)	F_{Air} (+1%)	F_{Air} (-1%)	F_{HSR} (+1%)	F_{HSR} (-1%)
E_{Air}	2.9726%	-2.9403%	-0.01679%	0.01654%	0.01667%	-0.01665%
E_{HTR}	-1.5229%	1.4765%	0.00554%	-0.00557%	-0.00561%	0.00549%

Further, we calculated the first partial derivatives of probability of HSR to find that the first partial derivatives in term of fare is the largest, it means that the impacts of the fare on the demand is most evident. Eq.(8) is the second partial derivative of probability to the fare, and when c=0.68 *RMB/km* the second partial derivative equals *0*. It means that the fare of HSR should be less than 0.68 *RMB/km*, otherwise, HSR is less competitive in the high-speed transportation market.

$$\frac{\partial^2 P_{HST}(V_i, C_i, F_i)}{\partial C_{HST}^2} = \left(\frac{\partial U_{HST}}{\partial C_{HST}}\right)^2 \exp(U_{Air})\exp(U_{HST})\frac{\exp(U_{Air} - U_{HST})}{\exp(U_{Air} + U_{HST})^3} \quad (8)$$

Figure 3 shows the changes of the ratio between the modal splits of the flight and HSR in the corridor corresponding to the different C_{Air} and C_{HSR} in the context of the given (V_{Air} ,V_{HSR} ,F_{Air} ,F_{HSR}) in Table 7. It can be seen that the ratios of P_{HSR} to P_{Air} differ a lot corresponding to different C_{Air} and C_{HSR}. The line $C_{Air} = C_{HSR} +0.3$ in undersurface is a projection of the sites where the ratios begin to change. The camber on the left-top of the line is steep, which means that the ratio P_{HSR}/P_{Air} increases quickly and if C_{Air} and C_{HSR} are within this area the substitution potential of HSR for the flight increases dramatically. The camber on the opposite side is smooth, it means that if C_{Air} and C_{HSR} are within this area the substitution potential of HSR for the flight is not obviously. Figure 3 provides a theoretical principle for the two kinds of suppliers to make pricing strategy.

It can be known that open of HSR in Har-Da corridor will reduce the market scale of flight by 76.70% in the context of situation given in Table 7. Therefore, in order to battle this, the flight supplier may take some counter-measures, such as shortening the needed time through access improvement or reduction of waiting time; decreasing the ticket price; or increasing service frequency. However, this kind of methods only reflects the competition between the two modes and does not take the cooperation into account. We think that as a whole system the suppliers of the high speed transportation should compete and cooperate simultaneously. For example, due to its high cost, serious pollution and high energy consumption, the flight supplier should share the same passenger resource to form a linked timetable and combined frequency.

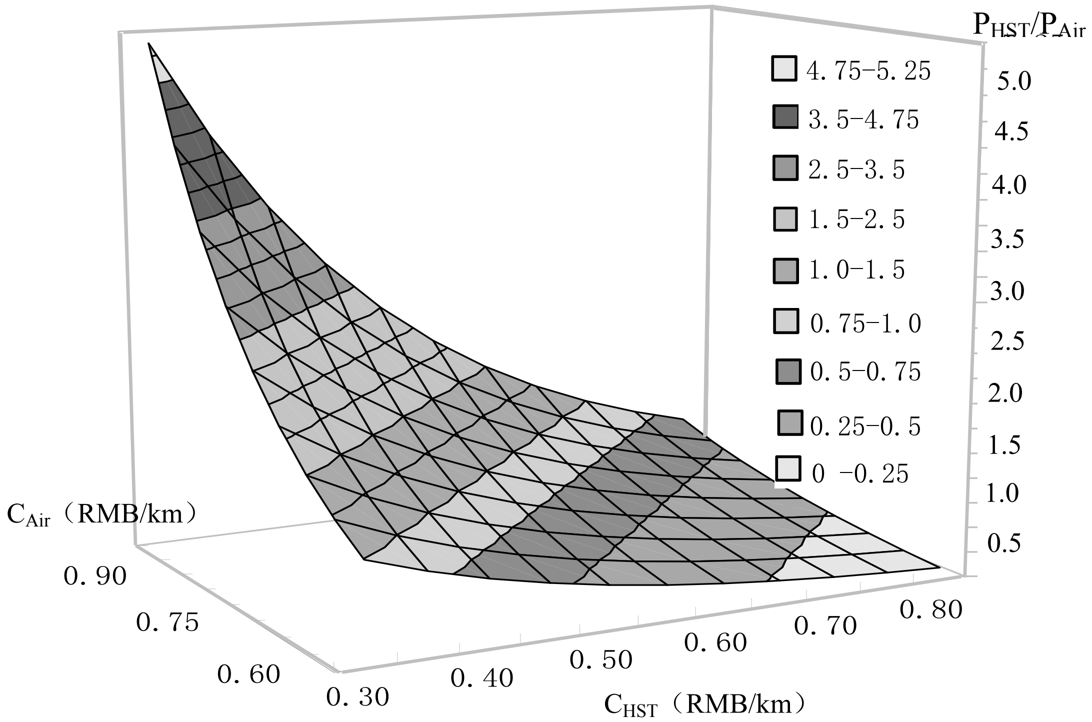


Figure 3: Ratio of Modal splits of DaLian-Harbin based on Fares

Under the situation given in Table 8, the optimal strategy of competition-cooperation can be calculated with Eq.(4) and Eq.(6) as (C_{Air} ,C_{HSR} ,V_{Air} ,V_{HSR} ,F_{Air} ,F_{HSR})=(0.87*RMB/km*, 0.48 *RMB/km*, 9.62 *km/min*, 3.69 *km/min*, 8.2*h*, 2.7*h*). This operation strategy can reduce pollution, save energy and make full use of the flight and HSR jointly. This principle can be used to allocation the resources in Har-Da transportation corridor. On one side it can keep the competition between two kinds of high-speed modes, while on the other side it can help the high-speed modes to get maximum market scale in the competition with conventional-speed modes.

4. Conclusion

The aim of this research is to forecast the influence of HSR on existing transport modes before the open of HSR. With data from SP survey, the influence of HSR attributes (running speed, rate per km and vehicle service frequency) on air transport is analyzed to support the optimization of resource allocation and network of high-speed passenger transport. Har-Da HSR is taken as an example for case study in this paper, but the method can be expanded to high-speed passenger transport corridor.

However, there are some insufficiencies in this study, for example: in modal split analysis, it is lack of short distance in case study, due to the much small quantity of air passengers in Dalian-Shenyang range; it is lack of change running speed of HSR in the survey to analyze the sensitivity of time cost in the corridor, due to the degree of difficulty for carrying out survey and the relatively unchangeable running speed of HSR after it completed, which is negative for optimization of high-speed transport supply in corridor; the questionnaires need more detail analysis based on classification by basic information of respondents to study policy to different kind of passengers.

References

Daniel, V. (2002), Optimal pricing in railway passenger transport: theory and practice in Netherlands, Transport Policy, 9, pp.95-106,

Gerken, J. (1991), Generalized Logit model, Transportation Research, 25B, pp.75-88.

He, Y.Q., Mao, B.H., Chen, T.S. and Yang, J. (2006), The mode share model of the high speed passenger railway Line and its application, Journal of the China Railway Society, 28(3), pp.18-21.

Milan, J. (1996), The trans European railway network: three levels of services for the passengers, Transport Policy, 3(3), pp. 99-104.

Nuzzolo, A., Crisalli, U. and Gangemi, F. (2000), A behavioural choice model for the evaluation of railway supply and pricing policies, Transportation Research Part A: Policy and Practice, 34(5), pp. 395-404.

Park, Y. and Ha, H.K. (2006), Analysis of the impact of high-speed railroad service on air transport demand, Transportation Research Part E, 42(2), pp.95-104.

Piao, J.N. (1993), A study of intercity transport mode choice model, China Journal of Highway and transport, 6 (Sup. 1), pp.51-56.

Si, B.F. and Gao, Z.Y.(2005), Optimization model for rail passenger transportation under the condition of integrated transportation system, Journal of the China Railway Society, 27(5), pp.6-10

Study on the location of land-port in Guizhou

Zhongzhen Yang, Yonglei Jiang and Jing Lu

College of Transport Management, Dalian Maritime University, Dalian, 116026, China.

Abstract

In order to help the inland area to build the land-port scientifically and reasonably, this paper studies the location of land-port. First, the location principle is analyzed, which divides the land-port location factors into two parts, the economic factors and the non-economic ones. Second, a land-port location model is developed based on the flow method, which takes the traffic assignment model as the main part and aims to maximize the volume of the containers transferred from the land-port to the seaport. Moreover, in order to solve the model, a method of building the integrated transport network is put forth. Finally, by taking the Southwest of China as the case study area, a numerical test is done, which gives an optimal location pattern of land-ports in Guizhou Province.

Keywords: Land-port, container transport, location, and traffic assignment

1. Introduction

As important logistics nodes, land-ports have been developing rapidly in China. The main reason is that seaports need to enhances the competitiveness through constructing their own logistics platform in inland areas to enlarge their market share, and on the other hand, in order to develop the economy, promote the geographical superiority and save logistics cost, the inland areas need to build a logistics node that can connect the overseas efficiently (Notteboom and Rodrigue, 2005). The land-port basically has the same functions with the seaports except for the vessel loading and uploading operations. Importers and exporters can complete the procedures of booking, customs declaration, and commodity inspection in a land-port, and then pass the cargoes to shipping companies and obtain bills of lading, or return the empty containers to the shipping company after the custom clearance and picking up the cargoes (Ye, 2005).

For many years the Southwest in China has been working hard to build the transport network to the seaports, and the southwest road corridor to the sea that was opened in 2005 has greatly improved the accessibility from the Southwest to the ports in Beibu Gulf. However, the transport corridors from some areas in the Southwest to seaports are still not good enough and can not satisfy the demands from its own economic zone and the China-ASEAN Free Trade Area. The main problems are short of comprehensive logistics terminals and high costs of the time and money during the transport. Therefore, it is necessary to construct an international land-port in the Southwest to cooperate the corridor construction to form an effective logistics network to the seaports.

Based on the import and export freight volume and the existing transport corridors, this paper studies the location of land-port in Guizhou province. A location model of land-port is developed and the number and spatial distribution of the land-ports in Guizhou province are determined. At last, directions and the related volumes of the container freight from and to the seaports are calculated and analyzed in the designed logistics network.

2. Location Principle and Candidate Locations of the Land-Ports

In order to make full use of the land-port, the best location for the land-port should be a city or a transport terminal, with big transport volume, wide development potential and great radiation and distribution effect on the neighboring region. Secondly, the land-port should be near to an economic central city to take the advantage of the good infrastructures and city economy for its survival and development. Moreover, the land-port should be located near to the railway freight station or the road freight terminal for the purpose of using the railway or highway to link to the domestic seaport and further to the oversea ports.

The location of the land-port involves the aspects such as the economic development, regional policy, natural environment, labor resource, transport infrastructure, and the shipper location. According to the location preference theory of Weber, the location factors can be divided into economic factors and non-economic ones (Nakamura et al., 1997). Table 1 shows the details of the location factors.

Table 1: Location Factors of Land-port

Economic Factors	transport cost, land cost, construction expense, fuel price, resource costs like water and electricity payment, salary level, other social service costs
Non-economic Factors	national and regional laws and regulations, foreign trade development level, transport accessibility, supporting facility, natural condition

At present, transport corridor from the Southwest to the seaports, which starts from Chongqing and ends at Zhanjiang in Guangdong, is about 1314 km long and takes the seaports in Beibu Gulf of China as the key point. However, because the port condition in Beibu Gulf is quite poor, Guizhou province (in the middle of the Southwest) is constructing second transport corridor to the seaport, i.e. corridor from Guizhou to Guangdong, with the aim of improving the accessibility from the Southwest to Pear River Delta. On both corridors, Guizhou is located in the middle and serves as a connecting point between the up-stream and down-stream. Thus it is necessary to locate the land-port in Guizhou to form an efficient logistics network based on highways and railways to realize the economies of scale of the transport system. Then, by taking the Southwest as the study area, the location alternatives of land-ports in Guizhou are given as follows.

In Guizhou, the non-economic factors differ a lot from place to place, while the cost elements of the economic factors differ slightly. Therefore when determining the alternative locations, it is necessary to consider the income element in the economic factors as well as the non-economic factors such as economies of scale and market demand, ect.. According to the economic situation, transport facilities and condition of the supporting facilities, we determine five location alternatives in Guizhou for the land-port. The spatial locations are shown in Figure 1. It should be noticed that all alternatives are the occurrences of the provincial logistics terminal (Communication Department of Guizhou, 2004).

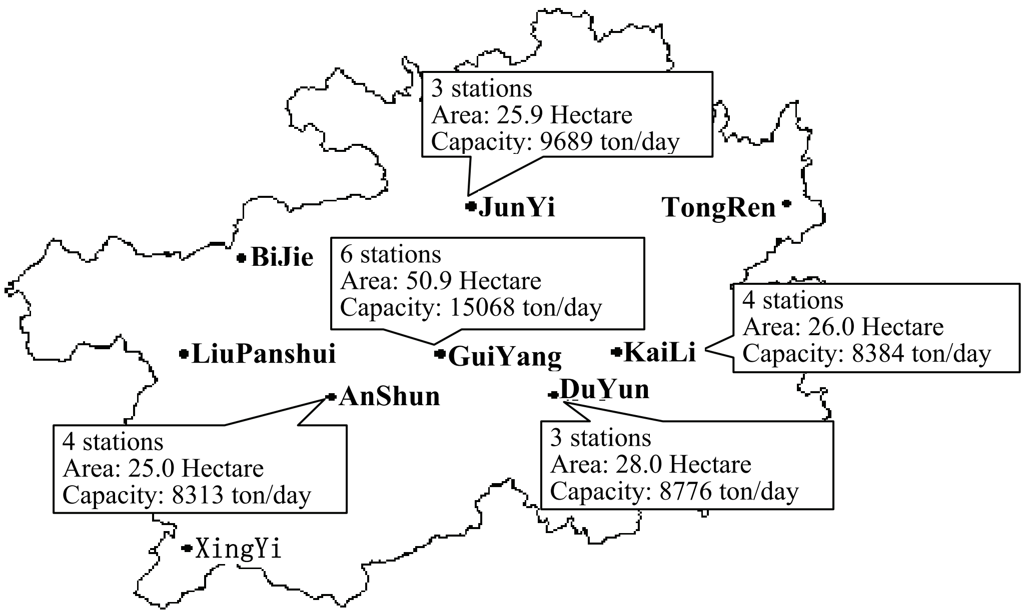


Figure 1: Alternative Locations of the Land-port in Guizhou

The above qualitative analysis describes the location alternatives in the Southwest quite roughly. It is obviously impossible to build five large-scale land-ports simultaneously in one province since the capacity of five land-ports exceeds the demand. Therefore, it is necessary to further study with quantitative method.

3. Location Choice Based on Transferred Volume

We need to determine the optimal scheme in the context of the five alternatives according to the actual demand. It means that if the transferred volume in an alternative could not achieve a certain threshold, there are no necessity and feasibility to build land-port there. Therefore, to determine the number and spatial distribution of the land-ports, it is necessary to analyze the path choice behavior of the served freights that are from or to the seaports, and to calculate the transferred volume in each alternative, and then to give the final construction plan. Figure 2 shows the structure of the quantitative analysis. Firstly, an integrated transport network model, which covers the study area and includes highway and railway links and nodes, is constructed. In the network, the alternative locations are assumed to be the terminals where freights can be transshipped from one mode to another. Secondly, O-D matrix of the exported and imported freight (container based) of the study area is estimated. Next, the O-D matrix is assigned onto the integrated transport network to obtain the being transferred volume and flow direction in each alternative, and finally the number and corresponding locations of the land-port are determined.

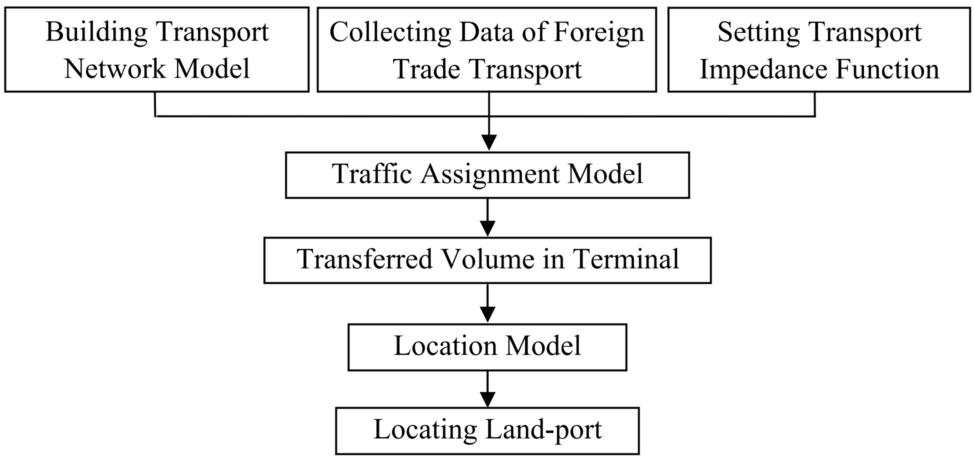


Figure 2: Structure of Locating Land-port based on transferred Volumes

3.1. Construction of the Integrated Transport Network Model

3.1.1. Structure of the Network

Here the integrated transport network means a network composed by the links and nodes of highways, railways and waterways. The nodes represent the terminals for transfer between two modes, or intersections of two corridors of the same mode (Southworth amd Peterson, 2000), while the links, which represent the lines of railway, highway or waterway have the attributes of length, speed and capacity. The speed changes with the flows of the link. The most important work in constructing an integrated network is to classify the links and set up the performance function. With the performance function, the O-D matrix can be assigned onto the network and then the optimized path between each O-D pair can be analyzed and the flow passing each node can be obtained.

Links of the integrated transport network can be divided into two kinds, namely real links and dummy ones. The former represents the real transport corridor, such as roadway, railway and waterway. The latter represents the virtual transport corridor, for example, the transfer terminal should be represented into nodes and dummy links, the dummy links display the freight transfer process between the highway and railway, and between the railway and railway as well as the railway and waterway. Transport impedances of dummy links includes both money cost and time delay of the transfer.

3.1.2. Attributes of the Links

In order to differentiate and manage link elements when building the integrated network model, they need to be coded type by type. The coding method is shown in Table 2, and the attributes of the real links are shown in Table 3.

Table 2: Types of the Links and their Codes

Types	Real Link				Dummy Link			
					Land-port			Seaport
	Express	National/ Provincial Highway	Railway	Waterway	Road to Road	Road to Rail	Rail to Rail	Port to Port
Codes	1001	1003	1002	1004	01	02	03	04

Table 3: Attributes of the Real Links

Attribute \ Mode	Highway		Railway	Waterway
Speed	Express	100 km/h	120 km/h	20 kn
	National / Provincial	60 km/h		
Length	Length of the Links		Length of the Links	Length of the Links
Cost	3 Yuan/Container · km		0.7 Yuan/Container · km	0.1 Yuan/Container · km
Capacity	C_{road}		C_{rail}	maximum

Within the integrated network model, the dummy link may be further divided into two categories: 1) the internal dummy links in land-port that display the road-to-rail and rail-to-rail transfer; and 2) the internal dummy links in seaport that display the road-to-water and rail-to-water transfer. The land-port is only a conceptual one in the period of location analysis, its capacity can be assumed without limitation, while the transfer cost can be set a suitable value based on its scale and the reference of some existing land-ports. The transfer time and cost in seaports are decided based on real situation in the seaports. The designed capacity of the seaport is used as its throughput capacity. The transport costs (including time and money) of seaports can be obtained through multiplying their attraction by the corresponding weights (Wang, 2008). Here, the attributes of the dummy links are set as shown in Table 4.

Table 4: Attributes of the Dummy Links

Type	Transfer in Land-port	Transfer in Seaport
Code	01, 02, 03	04
Time	-	Transfer time in seaport
Cost	Transfer cost in land-port	Transfer cost in seaport
Capacity	Very large value	Designed capacity

3.1.3. Artificial Network for Structure of Land-port

The land-port is a regional logistics terminal, which has highly effective transfer equipments and connects the inland highway and railway. If we want to build the land-port into an international commercial port in the integrated network, dummy links must be used comprehensively and accurately to reflect the transfer relationship between different modes. Figure 3 illustrates the artificial network that represents the transfer structure in a land-port. It can be seen that in the integrated network, a land-port is expressed as a local network consisting of several nodes and dummy links. The local network of the land-port is in an octagon structure, whose eight nodes are linked to eight channels that connect the land-port to the outside respectively (three railways, four expressways, and a national highway), and the internal dummy links connect the channel with each other to form an organized transport terminal unit.

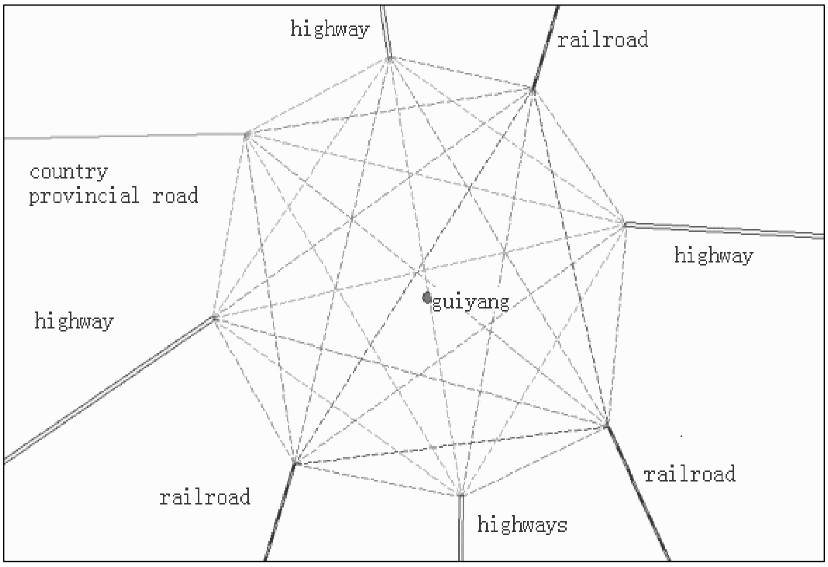


Figure 3: Artificial Network for the Transfer Structure of Land-port

3.1.4. Link Performance Function

In order to reflect the total transport costs of the corridor, it is necessary to take both distance and time costs into account when setting link performance function. Generally, the distance cost is determined by the fixed cost and the corridor-length based operating costs, the travel time is composed of the free-flow time and congestion time (Wang et al., 2004), the finally obtained performance function for real link is as Eq. 1shown.

$$c_i(x) = (k_i + op \cdot L_i) + t_i(1 + \alpha_i (x_i / C_i)^{\beta_i})vot \tag{1}$$

Here, $c_i(x)$ = link transport impedance; k_i = fixed cost of link i; op = operation cost per unit length; L_i = length of link i; t_i = free-flow travel time of link i; C_i = capacity of link i; vot = Value of time; α_i, β_i = parameters.

In order to reflect the total transport cost in a land-port, it is necessary to consider the time and money costs used for the transfer in the performance function of dummy link. As supposing the transfer can be done only in the land-port, we set the performance function for dummy link as Eq. 2.

$$d_i(T) = T_{r-\cos t} + \frac{1}{2}(24/F)vot \tag{2}$$

Here, $d_i(T)$ = total transfer cost; $T_{r-\cos t}$ = money term of transfer cost, F_i = the operation frequency of railways container liner to direction i.

3.2. Location Model

The principle of the flow-based location method is that transshipped volume decides a node whether or not to be a land-port. The transfer volume at a node is the sum of freights flowing through the node owing to transport demand between O-D pairs. Therefore, under the condition of the given alternative locations, land-port location model based on transfer flow method is as Eq. 3.

$$\begin{aligned} &Max \quad \sum_i \sum_j Q_{ij} y_i \\ &s.t. \quad \begin{cases} \sum_i y_i = n \\ Q_{ij} = \sum_k q_{ij,k} \end{cases} \end{aligned} \tag{3}$$

Here, n =optimized number of land-ports, Q_{ij} =container flow of railway j in alternative i. If the alternative is designed as land-port, $y_i = 1$, otherwise $y_i = 0$. $q_{ij,k}$ =container flow on dummy link k that connecting to railway j in alternative i, it can be calculated with traffic assignment model shown in Eq. 4.

$$\begin{aligned} &min \quad \sum \int_0^{x_i} c_i(x)dx \\ &s.t. \quad \sum_k f_k^{rs} = q_{rs} \quad \forall r,s \\ &\quad x_i = \sum_r \sum_s \sum_k f_k^{rs} \delta_{i,k}^{rs} \quad \forall i \\ &\quad f_k^{rs} \geq 0 \quad \forall r,s \quad \forall k \end{aligned} \tag{4}$$

Here, x_i =container flow of link i; $c_i(x)$ =transport impedance of link i; f_k^{rs} =container flow on path k, which starts from origin r and ends at destination s; $\delta_{i,k}^{rs}$=1 if link i is within path k from origin r to destination s, otherwise $\delta_{i,k}^{rs} = 0$; *R*=set of origins; *S*=set of destinations.

It should be noted that in the model the optimized n is a variable and not known until we solve the model. However, basically, the number of land-ports in a region should have the upper and lower limits, for example, from necessity analysis it can be known that at least one land-port should be built in Guizhou, while in terms of feasibility there should be no more than three land-ports. As the result, in our case, we can set $1 \leq n \leq 3$.

Since n is a variable, we cannot solve Eq. 3 directly, but have to solve it in the cases of $n = 1$, $n = 2$ and $n = 3$ respectively. When $n = 1$, it is an enumeration problem, where in terms of each alternative we assign the O-D matrix to the network, and then compare the transfer quantities among the n scenarios (here n is the number of the alternatives) to find the biggest one, which will be the land-port location plan if one land-port is constructed. When $n = 2$ and $n = 3$, we can also use enumeration method to set a set of scenarios, and then test each scenario to get the best results for the cases of $n = 2$ and $n = 3$. Finally comparing the three optimized plans that are corresponding to the cases of $n = 1$, $n = 2$ and $n = 3$ respectively, we can find the general best one. As the increment of the number of the alternatives, the genetic algorithm may be used for the optimization.

4. Calculation And Analyses

We use the Southwest in China to carry out the case study in the context of the area doing import and export businesses through the seaports in Beibu Gulf and Pear rive Delta. First, we build an integrated network model that covers the Guizhou, Sichuan, Chongqing, Hunan, Guangxi, Guangdong and Yunnan provinces and consists of railway, expressway and national/provincial highway. Second, we estimate regional foreign trade volume in terms of container O-D matrix with the data of Guizhou, Sichuan provinces and Chongqing city. The estimated results are shown in Table 5 (Wang, 2008).

Thus we can solve the model based on the integrated network and the foreign trade container O-D matrix in the context of $n=1$, $n=2$ and $n=3$ respectively. The calculated results are as follows.

Table 5: Containers O-D Matrix between the Southwest and Foreign Countries (Unit: TEU)

O \ D	10	11	12	13	14	15	16	17	18	21
1	184	240	286	224	494	530	881	516	558	539
2	10	13	16	12	27	29	48	28	32	29
3	288	375	447	350	773	828	1377	807	907	842
4	0	0	0	0	1	1	1	1	1	1
5	18	23	28	22	46	51	85	50	56	52
6	1	1	2	1	3	3	5	3	3	3
7	379	494	589	460	1,080	1,091	1,814	1,063	1,195	1,109
8	1,638	2,131	2,543	1,988	4,394	4,708	7,833	4,589	5,160	4,789
9	129	167	200	156	345	370	615	360	405	376
19	60,030	13,452	15,313	5397	36,171	38,777	134,458	20,559	16,147	105,500
20	8143	4519	6995	3989	7181	16838	38483	3038	12265	11061

Remark: 1=ZunYi Area, 2=KaiLi Area, 3=LiuPanshui Area, 4=BiJie Area, 5=TongRen Area, 6=XingYi Area, 7=Duyun Area, 8=GuiYang Area, 9=AnShun Area, 19=Sichuan Province, 20=ChongQing; 10=India, 11=Thailand, 12=Indonesia, 13=Philippines, 14=Korea, 15=Japan, 16=America, 17=Taiwan, 18=Viet Nam, 21=Hong Kong.

4.1. Result in the Case of Only One Land-port

Table 6 illustrates the transfer volumes corresponding to the five scenarios of n=1. It can be seen that locating land-port in Guiyang has the largest transfer volume, which is 180,600 TEU/year, and the transfer ratio is 99.5%. Therefore, if only one land-port is built, it should be located at Guiyang. In this case, the containers leaving Guiyang land-port for Beibu Gulf and Pear River Delta through railway are 69,000 TEU /year and 111,500 TEU/year respectively. In other words, the number of the export containers from the Southwest to the seaports in Pear River Delta is approximately twice of that to the Beibu Gulf. It means that if land-port was constructed in Guiyang to transfer the export containers for the Southwest, seaports in Pear River Delta will become more attractive than the seaports in Beibu Gulf. Thus, the Southwest should strengthen the relationship with Pear River Delta positively to encourage the port authorities and shipping companies to involve in the construction. On the other hand, the port authorities and shipping companies in Pear River Delta should get permission from Guiyang government as fast as possible to construct the facility and service in the land-port.

Table 6: Annual Transferred Containers with one Land-port

Alternatives	Guiyang	Zunyi	Duyun	Kaili	Anshun
In-bound Amount (TEU)	181,500	290,400	310,300	154,200	149,500
Amount of Transfer to Railway (TEU)	**180,600**	155,400	153,800	154,200	148,600
Ratio of Railway	99.5%	53.5%	49.6%	100.0%	99.4%

4.2. Results in the Cases of Two and Three Land-ports

Table 7 shows the transferred volumes of the top three scenarios in the case of two land-ports, it can be seen that the transferred volume of the second one (locating a land-port in Guiyang and Anshun respectively) is the largest, which is 236,000TEU. In this case, 69,000TEU and 112,000TEU containers will be transferred to Beibu Gulf and Pear Rive Delta respectively through the land-ports.

Table 8 shows the transferred volumes of the top three scenarios in the case of three land-ports, it can be seen that the transferred volume of the third one (locating a land-port in ZunYi, GuiYang and AnShun respectively) is the largest, which is 294,000TEU. In this case, 72,000TEU and 110,000TEU containers will be transferred to Beibu Gulf and Pear Rive Delta respectively through the land-ports.

Table 7: Annual transferred Containers in two Land-ports (Unit: 10000TEU)

	Pattern 1		Pattern 2		Pattern 3	
	Guiyang	**Duyun**	**Guiyang**	**Anshun**	**Guiyang**	**Kaili**
In-bound Amount	15.1	20.3	18.1	5.5	14.9	5.1
Amount of Transfer to Railway	15.0	7.7	**18.0**	**5.5**	14.8	5.1

Table 8: Annual Transferred Containers in three Land-ports (Unit: 10000TEU)

	Pattern 1			Pattern 2			Pattern 3		
	Zunyi	**Guiyang**	**Duyun**	**Zunyi**	**Guiyang**	**Kaili**	**Zunyi**	**Guiyang**	**Anshun**
In-bound Amount	23.9	13.3	17.4	24.2	12.4	4.5	23.5	13.2	5.4
Amount of Transfer to Railway	11.2	12.3	5.2	11.0	11.6	4.5	**10.9**	**13.1**	**5.4**

Although total container volume transferred to railway increases with the increment of the number of the land-ports (transfer volumes are 180,600TEU, 236,000TEU and 294,000 TEU respectively corresponding to the cases of one, two and three land-ports), it is worthy to note that no matter how many land-ports are constructed, the transferred volumes from highway to railway in all land-ports is about 180,000TEU, among them 70,000TEU are transferred to seaports in Beibu Gulf, and 110,000TEU are transferred to seaports in Pear River Delta. The transferred volumes of 236,000TEU and 294,000TEU in the cases of two and three land-ports include rail-to-rail transfer, therefore, it can be concluded that construction of one land-port in Guiyang is the best choice.

5. Conclusion

Land-port is the extension of seaports. It is an important terminal in container transport network and an integrated logistics center. It should be located in the economic central city and at the intersection of the highways and railways. The logistics cost of export and import from inland will be greatly reduced through the operation of the land-port.

The location of the land-port is an important problem in logistics system. The optimized land-port location pattern may realize an integrated transport of the export and import containers for inland areas with lower cost, and improve the accessibility of the inland area to foreign ports. The flow-based method of land-port location model in this paper can help us find a suitable site to build land-port. However, for adoption of the model, an integrated transport network model that corresponds to the study area should firstly be built with a geography information platform.

References

Communication Department of Guizhou (2004), “Report of Highway Terminal in Guizhou”, Guizhou, China, 2004.10.

Nakamura, H. *et al*. (1997), New Systematic Civil Engineering 50: Survey – Regional Survey and Analyses, GihoDo Press Company, Tokyo, 1997.

Notteboom, T.E. and Rodrigue, J.P. (2005). Port regionalization: towards a new phase in port development. Maritime Policy and Management, V32, N3, 297–313.

Southworth, F. and Peterson, B.E. (2000), International freight network modeling. Transportation Research Part C 8 (2000) 147-166.

Wang, H. (2008), Research on the Feasibility and Effect of Constructing Dry Port in Guizhou Province, Master Thesis of Dalian Maritime University, Dalian, China.

Wang, Y., Zhou, W., Lianen, L.V. (2004), Theory and Application Study of the Road Traffic Impedance Function, Journal of Highway and Transportation Research and Development, 2004.9，21(9): 82-85.

Ye, L. (2005), Construction of Land-port, China Water Transport, 2005.4: 54-55.

Unilateral GHG control measure and aviation industry: a theoretical analysis

Andrew C. L. Yuen[1] *and Anming Zhang*[2]

[1] Dept of Decision Sciences and Managerial Economics, The Chinese University of Hong Kong, Hong Kong.
E-mail: andrewyuen@baf.msmail.cuhk.edu.hk
[2] Sauder School of Business, University of British Columbia, Vancouver, BC, Canada V6T 1Z2.
E-mail: anming.zhang@sauder.ubc.ca

Abstract

This paper investigates the effects of unilateral GHG control measures on the aviation industry and total GHG emissions. The result suggests that if a country unilaterally takes actions to control aviation GHG emissions, the charges at its hub and spoke airports may increase. This increase in the airport charges will lead to a shift of domestic connecting flights between hub and spoke airports outside the country. It may also place its home airlines at a strategic disadvantage, as the home airlines usually operate their hubs at the home country. Furthermore, the unilateral control measure will reduce GHG emissions in the country implementing the measure, while it may lead to an increase or a decrease in world emissions. It is because as mentioned, the domestic connecting flights may increase at another country, in which the airline network may be very inefficient in terms of GHG emissions.

Keywords: Aircraft emissions; Airline Competition; Environmental Economics; Antitrust

1. Introduction

Public concerns over climate change have pressured both the aviation industry and regulators to mitigate aviation greenhouse gas (GHG) emissions in hope of avoiding their adverse impacts on global climate. As a result, while a number of airlines set up their own voluntary schemes to reduce their GHG emissions,[1] some governments also consider unilaterally mandating the implementation of some measures to control emissions from the industry. Among these control measures, the one of the largest scale would be the inclusion of the aviation sector in the European Emission Trading System (EU ETS) in 2012.[2] This kind of unilateral action has been strongly opposed by the industry and various countries, based partly on economic concerns, such as competitive issues.[3] For example, Air France argued that the EU ETS would give a competitive edge to those airlines with a hub outside of Europe.[4] It is because connecting flights between hub and spoke airports in the EU are subject to related GHG charges resulting from the EU GHG emission control measure, while those flights outside the EU are not.

[1] Currently, a number of airlines, including Air Canada, American Airlines, Cathay Pacific, Continental Airlines, Delta Airlines, Northwest Airlines and Qantas, have introduced their carbon offset programs for their customers. More discussion about voluntary emission control measures for aviation can be found in ICAO (2007).

[2] In July 2008, the European Parliament passed the second reading vote of a directive which includes GHG emissions from flights to, from and within the EU in the EU ETS from 2012. Under the new legislation, all airlines will be covered (irrespective of nationality). They will be able to sell surplus allowances if they reduce their emissions and will need to buy additional allowances if their emissions grow.

[3] For example, in the ICAO's meeting in July 2007, the US representative objected to the inclusion of international flights in the EU ETS without their consent, and pointed out that it runs contrary to EU Member State international legal obligations under the Chicago Convention on International Civil Aviation, and under numerous bilateral air services agreements, including those with the U.S.

[4] Reuters, June 9, 2008.

Due to the potentially significant impacts of unilateral GHG control measures on GHG emissions, national welfare and the aviation industry, the economic arguments for and against the unilateral GHG control measures ought to be examined more rigorously. Although there is a growing empirical literature (e.g., Anderson, 2001; Delft, 2005, 2007; Scheelhaase and Grimme, 2007; Scheelhaase, et al., 2008) examining the effects of GHG control mechanisms in the aviation industry, discussion of the issue so far lacks much relevant theoretical work. Thus this paper aims to provide an analytical framework for investigating GHG emission control measures in the aviation industry. In particular, we will first examine how the unilateral measures affect the competition of airlines with different networks. For example, to serve the same market, airlines may face different GHG charges if their hub locations differ (as argued by Air France mentioned above). Thus the measures may give some airlines strategic advantages in the competition. The empirical results in Scheelhaase, et al. (2008) showed that network carriers based outside the EU are likely to gain a significant competitive advantage, compared to EU network carriers, if the aviation industry is included in the EU ETS. Our second objective is to look at how the unilateral GHG emission control measures affect world emissions and its distribution among countries. In particular, the unilateral measures may affect airlines' pricing and network utilization, which, in turn, could affect world emissions.[5]

In the literature, a few theoretical works have been published on pollution control measures on the aviation industry; for example, Brueckner and Girvin (2008) explored the impacts of airport noise regulation on flight frequency and airfares; Brueckner and Zhang (2009) explores the effect of airline emissions charges on airfares, airline service quality, aircraft design features, and network structure. One of the major differences between this paper with them is that we consider asymmetric airlines. In our analysis, two airlines operate flights between two countries, and belong to each country, respectively. Under such duopoly competition, the two airlines use different networks to serve the market. As shown below, the introduction of asymmetry in the analysis is necessary for us to examine the competitive issues and the change in world emissions with the unilateral action to control GHG emissions – the main objectives of this paper. However, due to the complexity after considering the asymmetric case, we cannot take into account flight frequency in our model as the studies in the literature. Yet an important extension to this paper could examine the impacts of GHG control measures on air service quality, such as flight frequency: more frequent flights will reduce scheduled flight delays (Douglas and Miller, 1974). To derive the airlines' demand, we consider an infinite linear city model, where a consumer's utility depends on their brand preference. A number of studies in the literature (e.g., Brueckner and Zhang, 2001; Brueckner, 2004; Brueckner and Flores-Fillol, 2007) consider similar spatial models to examine airlines' network choice between fully-connected (FC) and hub-and-spoke (HS) networks.

In the analysis, we compare the case in which one of the countries unilaterally imposes GHG charges on both airlines with the no-action base case in which both countries have no action regarding GHG emissions. We find that if a country price discriminates between domestic and international flights, the GHG charges may be positive or negative, while they will be positive when its home airline's profit is not taken into account in the country's airport pricings. On the other hand, under uniform pricing, when a country moves toward GHG pricing, the airport charge will increase. Note that the adverse impacts of the positive GHG charges on the home airline may be larger relative to their foreign counterpart, as the home airline will operate connecting flights between the hub and spoke airports within its home country, where the GHG pricing is implemented. Thus, such an initiative may raise competitive concerns from those airlines. We also show that the positive GHG charges will reduce the total output in the airline market, while domestic flights connecting hub and spoke airports may be shifted from the home country to the foreign counterpart. Nevertheless, the consumer surplus of the air travel market will decrease. Finally, we find that world emissions may increase when a country takes the GHG control actions unilaterally. This is due to the fact that the domestic connecting flights may be shifted to a country with a less efficient network. This is a very interesting

[5] From a different perspective, Hoel (1991) showed that the unilateral reductions of GHG emissions may affect the outcome of international negotiations about reduced emissions, which may imply higher total emissions than if no action is taken.

result, with potentially important implications for governments deciding to implement unilateral GHG control measures.

The paper is organized as follows: Section 2 discusses the relationship between global climate change and the aviation industry, including GHG control measures implemented in the industry in practice. Section 3 sets up the model, and Section 4 examines airline behavior. Section 5 derives the unilateral GHG pricing, and Section 6 investigates its effects on airline competition, market output, consumer welfare and GHG emissions. Section 7 concludes.

2. Climate Change and Aviation

An increasing amount of scientific evidence indicates that the global climate change is predominantly a result of increases in GHGs caused by human activities (see Stern, 2006, for a comprehensive review). Transportation, in particular, is one of the most important sources of GHG emissions. As with other modes of transportation, air transport produces GHGs during operation. The aviation GHG of most concern is carbon dioxide (CO_2), which has long residence time in the atmosphere. Other aircraft emissions affecting climate include water vapor (H_2O), nitrous oxide (NO_X), sulphate (SO_4) and soot particles. Currently, the aviation industry contributes 3 percent of the total man-made contribution to climate change (IPCC, 2007). Given the fact that the direct contribution of the industry to the global gross domestic product (GDP) is only about 1 percent, the industry share of the man-made contribution to climate change is thus significant. Furthermore, air travel grew significantly in the past decade and will continue growing at a rapid pace in the coming years. This rapid growth in air traffic may lead to a significant increase in GHG emissions from the sector, even with improvements in aircraft fuel efficiency.[6] IPCC (2007) found that CO_2 emissions from global aviation increased by a factor of about 1.5 from 1990 to 2000, and predicted that the emissions will continue to grow by around 3-4 percent each year, unless additional measures are taken; Olsthoorn (2001) estimated that between 1995 and 2050, aviation emissions of CO_2 may increase by a factor of 3-6.

Given its global character, it is sensible for countries to act collectively to control GHG emissions. As of May 2008, more than 180 countries have signed and ratified the Kyoto Protocol,[7] an international agreement that sets binding targets for countries and regions to reduce GHG emissions. Yet, while the Protocol accounts for domestic aviation emissions, emissions from international aviation are not included.[8] It may be because allocating international aviation emissions to specific countries may be difficult. Instead, the Protocol calls for countries to pursue limitation or reduction of GHG emissions from international flights through the International Civil Aviation Organization (ICAO).[9] Currently, the ICAO focuses mainly on technological and operational measures in GHG emission control, whereas the progress of market-based measures through the organization is still limited.[10] Instead, a number of airlines set up their own voluntary schemes, in which passengers can pay extra to offset their emissions. The revenues from the passengers are then spent on various offset programs. Hodgkinson, et al. (2007) argued that the schemes do serve a useful purpose in enabling passengers who are concerned about their emissions to make an essentially carbon neutral flight. On the other hand, British Airways voluntarily participated in the UK ETS to reduce their emissions of CO_2 equivalent to below set targets. In return, it received an incentive payment from the UK Government. However, the effectiveness of those voluntary measures is still questionable.

As a result, some governments are also considering unilaterally mandating some measures to control emissions from the aviation industry. A growing number of empirical studies have been conducted to

---

[6] IPCC (2007) forecasted that technology developments might offer only a 40-50 percent improvement in fuel efficiency over 1997 level by 2050, while air traffic growth is estimated to be around 5 percent each year.
[7] The participating countries now contribute to about 64 percent of global GHG emissions in total.
[8] Annex I Parties of the Protocol.
[9] Article 2.2 of the Protocol.
[10] In particular, the ICAO has decided not to work towards a new global legal market-based instrument under the organization in 2004.

evaluate the proposed measures. Scheelhaase and Grimme (2007) provided an empirical estimation of the impacts of the inclusion of the aviation sector in the EU ETS on operating costs and transport demand for full service and low cost airlines, and found that the financial impact on low cost carriers and regional airlines is likely to be significantly greater than for network carriers. Furthermore, Scheelhaase, et al. (2008) found that EU network carriers are likely to encounter competitive disadvantages compared to airlines from non-EU countries on long-haul flights. Benito (2008) estimated the aviation CO_2 emissions reduction with the proposed EU ETS by taking into account the substitution between air travel and high-speed railways in the EU. In addition to the EU ETS mentioned above, the UK government levies an "Air Passenger Duty," which is designed to reduce CO_2, on all airline tickets sold in the country. However, IATA (2006) opposed the levy because the climate benefits due to the reduction in emissions is estimated to be £53 million only, while the GDP losses would be £400 million plus losses to air travelers. Hofer, et al. (2008) investigated the effects of an air travel carbon emissions tax on total travel-related carbon emissions in the U.S. They highlighted the fact that emissions taxes on air travel may cause potentially significant air-to-automobile diversion effects, which may substantially reduce the environmental benefits of air travel carbon emissions taxes. Forsyth (2008a), and Forsyth and Ho (2008) discussed the Australian ETS by taking direct emissions from aviation and indirect emissions from other industries, such as tourism, into account.

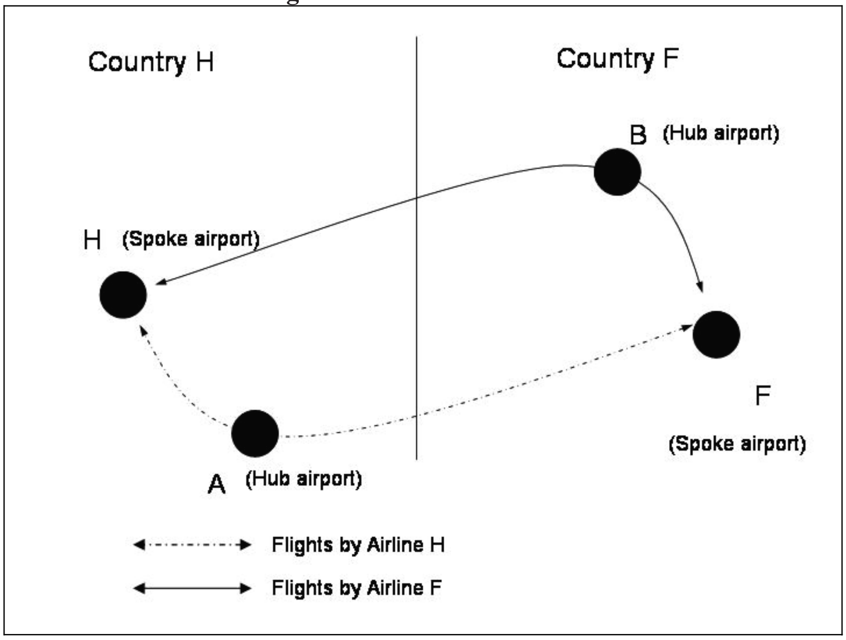
3. The Model

This analysis considers a model with an airline in country H (i.e., airline H) and an airline in country F (i.e., airline F) serving a same origin-destination (OD) market. As depicted in Figure 1, a simple air transport network is considered, as this is likely the simplest structure allowing us to address the main questions concerning us. To simplify the analysis, we focus on the OD market between airports H and F (i.e., the spoke airports in countries H and F, respectively). To serve the market, each airline will bring their passengers from the origin to its hub, which is located at its respective country (i.e., airports A and B for airlines H and F, respectively), and then from the hub to the destination. In practice, it is very common for airlines operating HS networks in both their domestic and international markets after market deregulation in the aviation industry (see Zhang, et al., 2008, for a review). One example relevant for our analysis is the air travel market between Vancouver and Munich. In this market, Air Canada will take up passengers from Vancouver and fly to Toronto, which is Air Canada's hub for its transatlantic market, then from Toronto to Munich. On the other hand, Lufthansa serves the market by connecting at Frankfurt. The assumption that airlines choose (or are only allowed to use) airports at their home country as their hubs can be justified for at least two reasons. First, although the international aviation market has been moving toward greater liberalization, cabotage is usually prohibited, and thus airlines are restricted to operate flights within the domestic borders of another country.[11] Second, most major airports around the world are suffering from congestion. Home carriers usually dominate those airports, implying that foreign carriers with limited access to the airport facilities may have difficulties operating their own flights at those airports.[12]

[11] A few exceptions include the creation of single aviation markets in the EU, and between Australia and New Zealand, in which cabotage rights are granted to airlines in the markets.

[12] Ciliberto and Williams (2007) measured the importance of operating barriers to entry, including limited access to airport facilities, as determinants of the hub premium; they found that exclusive access to and dominance of gates at the market endpoint airports are key determinants of the hub premium.

Figure 1: Network structure



In the model, aircrafts emit GHGs during the flight, where the emission intensity (i.e., emissions per flight between two airports) depends on the route they fly. On the other hand, for simplicity, we will abstract from other possible OD markets in the network (e.g., H to B, A to H) in our analysis.[13] The stylized problem that we want to analyze is the following: when one of the countries adopts (unilateral) GHG pricing, what are its effects on airlines' operating decisions, competition, market output, consumer benefits and world emissions? We model the problem as a two-stage game. In the first stage, countries determine the charges, with part of them owing to GHG pricing, at their respective hub and spoke airports. In the second stage, the two airlines compete in Cournot fashion.[14]

To derive the demand faced by the two airlines, we specify and solve the consumers' problem as follows. We consider an infinite linear city, where potential consumers are distributed uniformly with a density of one consumer per unit of length. Basso and Zhang (2007a) use a similar spatial model to investigate the rivalry between congestible facilities. Airlines F and H are located at 0 and 1, respectively, and the locations of the airlines are exogenous. If a consumer located at z chooses to travel by F, then she derives a net benefit (utility):

$$U_F = V - p_F - 4\tau|z|, \qquad (1)$$

where V is the gross travel benefit; p_F is the (equilibrium) airline F's airfare; 4τ is a parameter capturing consumers' transportation cost (i.e., the intensity of consumers' brand preference to airlines: consumers located closer to an airline's location will enjoy a higher benefit than others if they fly by that airline, given the same airfare), and τ is assumed to be positive.[15] Similarly, if the consumer chooses to travel by H, then she derives a net benefit (utility):

---

[13] Under this simplification, our model will not take the advantages of HS network into account, including those due to demand and cost complementarities, which have been widely discussed in the literature (e.g., Oum, et al., 1995; Hendricks, et al., 1995, 1999; Brueckner and Zhang, 2001; Brueckner, 2004; Brueckner and Flores-Fillol, 2007).

[14] Cournot competition between air carriers has been a common assumption in the theoretical airport pricing literature, which in fact has some empirical backing (e.g., Brander and Zhang, 1990; Oum, et al., 1993). See Basso and Zhang (2007b) for a review of the airport pricing literature.

[15] This transportation cost is multiplied by 4 in order to simplify most of the equations in the paper (see, e.g., equations (6)).

$$U_H = V - p_H - 4\tau|1 - z|, \qquad (2)$$

where p_H is the (equilibrium) airline H's airfare.

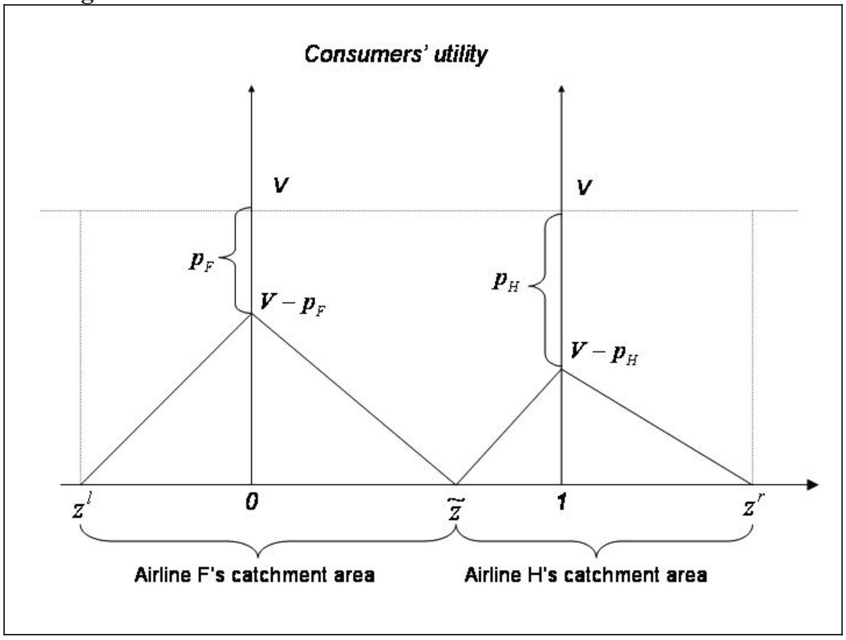
Assuming that everyone with $z \in [0,1]$ travels and both airlines capture some of the travelers,[16] then the consumer with $\tilde{z}$ will be indifferent between the two airlines when (also see Figure 2):

$$V - p_F - 4\tau\tilde{z} = V - p_H - 4\tau(1 - \tilde{z}),$$

or when

$$\tilde{z} = \frac{1}{2} + \frac{p_H - p_F}{8\tau}. \qquad (3)$$

Figure 2: Consumers' distribution and airlines' catchment areas



Condition (3) suggests that, when the airline H's (airline F's) fare decreases, more (less) passengers will choose airline H rather than airline F. Furthermore, airlines F and H serve consumers with $z < 0$ and $z > 1$, respectively. The cutoff points z^l and z^r for the customers between travel with airlines F and H, respectively, and not travel by air (which utility is normalized to zero) can be derived as follows:[17]

$$z^l = -\frac{V - p_F}{4\tau}, \qquad z^r = 1 + \frac{V - p_H}{4\tau}. \qquad (4)$$

By conditions (3) and (4), we can derive the demand system as follows:

[16] To have everyone with $z \in [0,1]$ travel, we assume $2V \geq p_H + p_F + 4\tau$, and to have both airlines capture some of the travelers, we assume $|p_H - p_F| < 4\tau$. Both of which are our maintained assumptions in the following analysis.

[17] The "no travel by air" option may also include traveling by other modes of transportation, such as automobile and high-speed railways. In such cases, we may need to further consider the consumer benefits and emissions from other modes of transportation in the following analysis. This kind of diversion effects are empirically examined in Benito (2008) and Hofer, et al. (2008).

$$q_F = \frac{V+2\tau}{4\tau} + \frac{p_H - 3p_F}{8\tau},$$
$$q_H = \frac{V+2\tau}{4\tau} + \frac{p_F - 3p_H}{8\tau}. \qquad (5)$$

For simplicity, q_i is measured by the number of flights. This measurement is equivalent to the number of passengers if each flight has an equal number of passengers, which holds when all flights use identical aircraft and have the same load factor. Note from (5) that an airline loses traffic when its fare rises, while it gains traffic when another airline's fare rises. In other words, the service provided by an airline is a substitute to another. From the demand system (5), we can obtain the following inverse demand functions faced by the two airlines:

$$p_F(q_H, q_F) = 2\tau + V - 3\tau q_F - \tau q_H,$$
$$p_H(q_F, q_H) = 2\tau + V - 3\tau q_H - \tau q_F. \qquad (6)$$

The inverse demand system (6) will be used to solve the Cournot-Nash equilibrium in Section 4.

4. Airline Competition

We shall examine the subgame perfect equilibrium of the two-stage game, starting with the analysis of the airlines' subgame in the second stage. Having characterized the demands in Section 3, we now specify airlines' cost. The carriers are symmetric in the sense that they have the same cost function. The airline *i*'s total cost function is given as:

$$C_i(q_i) = (c + t_i^S + t_i^U + t_{-i}^S)q_i, \qquad (7)$$

where *c* is the (constant) marginal cost per flight.[18] Recall that an airline will fly from a spoke airport to another spoke airport through the hub located at its home country. As a result, airline *i* will incur per-flight charges at the home country's spoke and hub airports, t_i^S and t_i^U, respectively, and a per-flight charge at the foreign country's spoke airport, t_{-i}^S. The airport charges will be determined in Section 5. In our analysis, we assume that the per-flight aircraft emissions in each route are exogenous.[19] Furthermore, the GHG taxes considered in Section 5 depend only on airlines' output. Thus, there is no incentive for airlines to improve their efficiency in GHG emissions. As a result, the GHG abatement cost is not included in (7). Yet an interesting extension to the present analysis would be considering incentive mechanisms for airlines to improve their emission intensity.

In the second stage, airline *i* chooses the number of flights to maximize its own profit. Thus its optimization problem is:

$$\underset{q_i}{Max}\ \pi_i(q_i; q_{-i}) = p_i q_i - C_i(q_i) \qquad (\textit{for } i = \text{F, H}) \qquad (8)$$
$$= (2\tau + V - 3\tau q_i - \tau q_{-i})q_i - (c + t_i^S + t_i^U + t_{-i}^S)q_i.$$

[18] Here, we assume that the airlines are operating under constant returns to scale. By using different cost functions, Kumbhakar (1990, 1992) and Hansen, et al. (2001) empirically found that airlines operate under increasing returns to scale, while Caves, et al. (1984) and Chua, et al. (2005) found constant returns to scale.

[19] In practice, some airlines have made some progress in GHG emission control. For example, Air Canada reduced the emission intensity of its flight operations by 28 percent since 1990 by investing in energy efficient new aircrafts, including Embraer regional jets and Boeing 777s.

The first-order condition for the maximization problem is:

$$\frac{\partial \pi_i}{\partial q_i} = (2\tau + V - 6\tau q_i - \tau q_{-i}) - (c + t_i^S + t_i^U + t_{-i}^S) = 0\,, \qquad (\textit{for } i = \text{F, H}) \qquad (9)$$

where $\partial^2 \pi_i / \partial q_i^2 = -6\tau < 0$. Thus, the second-order condition holds.

Solving equations (9), the airline *i*'s equilibrium output is:

$$q_i = \frac{2\tau + V - c - t_i^S - t_{-i}^S}{7} - \frac{6t_i^U - t_{-i}^U}{35}. \qquad (\textit{for } i = \text{F, H}) \qquad (10)$$

Conditions (10) suggest that airline *i*'s output decreases in the charges at the home country's airports (both the hub and spoke airports) and the foreign country's spoke airport, while it increases in the foreign country's hub airport charge. Furthermore, the total market output, $Q = q_i + q_{-i} = [2(2\tau + V - c - t_i^S - t_{-i}^S) - (t_i^U + t_{-i}^U)]/7$, decreases in the hub and spoke airport charges at both home and foreign countries.

Substituting (10) into (6), the equilibrium prices are:

$$p_H(q_F, q_H) = (2\tau + V)(1 - \frac{4}{7}\tau) + \frac{\tau}{35}[20(c + t_H^S + t_F^S) + 17t_H^U + 3t_F^U], \qquad (11)$$

$$p_F(q_H, q_F) = (2\tau + V)(1 - \frac{4}{7}\tau) + \frac{\tau}{35}[20(c + t_H^S + t_F^S) + 17t_F^U + 3t_S^U]. \qquad (12)$$

5. Country Stage

To examine the effects of the unilateral GHG emission control measure, we first derive the optimal airport charges, with part of them owing to GHG pricing. In this paper, we focus on the unilateral action from a country to control GHG emissions. Thus, in the country stage, we consider country H chooses charges t_H^U and t_H^S at its hub and spoke airports, respectively, taking the subsequent carriers behavior into account. The airport charges at country F are assumed to be exogenous.[20] Country H will maximize the following welfare function:

$$\underset{t_H^S, t_H^U}{Max}\, SW_H = \alpha CS + \beta \pi_H + TR_H - \gamma E_H\,, \qquad (13)$$

where CS is total consumer surplus of the air travel market. α and β, which are positive, capture the importance (relative to tax revenue) of consumer surplus and home airline's profit, respectively, in country H's policy-making decision.[21] TR_H is country H's tax revenue, which is the sum of tax revenues from its spoke and hub airports (i.e., $t_H^S(q_H + q_F)$ and $t_H^U q_H$, respectively). Recall airline H will use both the hub and spoke airports at country H, while airline F will only use the spoke airport at the country (see Figure 1). E_H is the GHG emissions related to country H, the positive γ is the

[20] If both countries' airport pricings are treated as endogenous, we need to consider the strategic behaviors between the two countries (see Yuen, Basso and Zhang, 2008, Small and Verhoef, 2007, and Ubbels and Verhoef, 2008, for recent reviews of the studies on the strategic behaviors of transport facilities).

[21] The magnitudes of α and β depend on a number of factors, including the percentage of foreign ownership stake in the home airline, the proportion of consumers from country H relative to the whole air travel market, costs of taxation from other sectors, etc. For example, in case of France, which supports national "champion" such as Air France, β tends to be large.

country H's GHG (per-unit) abatement cost or the cost to meet a GHG control target, for example, those set under the Kyoto Protocol.[22] Thus, the last term in (13) is the total cost on GHG emission reduction in order to meet a particular emission target. Note that the last term in (13) may also be interpreted as the benefits of emission reduction to the country and its consumers. In particular, for $\gamma = 0$, country H does not account for any GHG emission costs to the country. Thus, the pricing derived from the maximization problem in (13) will not include any GHG component – the base case considered in this paper. In the following, we will compare the equilibrium in the base case with that in the case of $\gamma > 0$.

We further specify the GHG emissions related to country H as follows:

$$E_H = e_H q_H + \delta e_K (q_H + q_F), \tag{14}$$

where e_H and e_K are the (per-flight) emissions of domestic travel in country H (i.e., between A and H as shown in Figure 1) and international travel (i.e., between A and F or between B and H), respectively. In practice, the emissions depend on a number of factors, including aircraft type, load factor and flight distance. In our present analysis, those factors are treated as given, thus the per-flight emissions are exogenous. Yet it is an important extension for this paper to consider the factors endogenously in the analysis. More details about this extension will be discussed in the concluding remarks of this paper. On the other hand, the international emissions are discounted by a factor $\delta \in [0,1]$ in counting the emissions related to country H. The different weight of domestic and international emissions in counting the emissions related to country H may be due to different treatments between the two kinds of flights in the existing GHG control mechanisms. For example, as mentioned, emissions from domestic flights are included in the Kyoto Protocol, while those from international flights are not.

An important assumption here is that country H will not take the emissions of domestic flights of another country into account. Note that although GHG emissions may have global effects, GHG emission control measures considered by most countries in practice aim to limit the emissions within its own territory, rather than global emissions. For example, under the EU ETS, the EU Member States are required to limit their emissions below *national* emission caps. On the other hand, in evaluating the effectiveness of the inclusion of the aviation sector in the EU ETS, the Full Assessment Report (EC, 2006) submitted by the European Commission only took into account the emissions from flights to, from and within the EU, but not from the flights outside the EU. Similarly, Delft (2005), a report commissioned by the European Commission, also evaluated different GHG control options by comparing the emissions from the flights within the European airspace.

In the following, we will explore both the price discrimination case and the "uniform pricing" case, in which country H charges a uniform (per-unit) charge to flights at its hub and spoke airports. Note that the uniform pricing case does not necessarily mean flights at the hub and spoke airports are charged exactly the same taxes. Instead, it should be understood as the existence of some a *priori* rule which says that flights at the hub and spoke airports are charged according to some fixed ratio (for example, the hub airport charges three times what the spoke airport charges). To save on notation – but without losing insight – we set the ratio to one in the uniform pricing case, that is, $t_H^S = t_H^U = \bar{t}_H$. In practice, the uniform pricing case in our analysis may be more relevant. For example, the proposal for including the aviation sector in the EU ETS requires domestic and international flights to participate in the same ETS. As a result, the flights are required to buy emission permits from the ETS at a same

[22] For example, a country may offset its GHG emissions through some project-based transactions, in which the buyer purchases emission credits from a project that can verifiably demonstrate GHG emission reductions compared with what would have happened otherwise. The most notable examples of such activities are under the Clean Development Mechanism (CDM) and the Joint Implementation (JI) mechanisms of the Kyoto Protocol. In 2006, the CDM and JI grew sharply to a value of about US$5 billion (World Bank, 2007).

price. Nevertheless, both the price discrimination case and the uniform pricing case will be examined in the following.

In the price discrimination case, country H chooses airport taxes at its hub and spoke airports to maximize its social welfare given in (13). The first-order conditions with respect to the charges give rise to the welfare-maximizing pricing rules. We obtain (the derivation is straightforward but long, and is hence given in Appendix):

$$t_H^U = \{\frac{5[14-\beta+\tau(\beta-4\alpha)]}{\Omega_1}e_H + \frac{30(2-\tau)\beta\delta}{\Omega_1}e_K\}\gamma + K_1, \quad (15)$$

$$t_H^S = \{\frac{6\beta+17\alpha\tau-6\beta\tau}{\Omega_1}e_H + \frac{[70-65\beta+(29\beta+14\alpha)\tau]\delta}{\Omega_1}e_K\}\gamma + K_2, \quad (16)$$

where $\Omega_1 = -(\tau-1)^2\beta^2 + 2(34\tau+3\alpha\tau-70)\beta - 4(2\alpha\tau+3)\alpha\tau + 140 > 0$ by the second-order conditions. K_1 and K_2 are constants.

In the base case, country H does not take the GHG emission cost into account, i.e., $\gamma = 0$. Thus, equations (15) and (16) suggest that in the base case, the equilibrium charges at the hub and spoke airports at country H are equal to K_1 and K_2, respectively. In particular, K_1 and K_2 may be considered as trade taxes (or subsidies) under imperfect international competition, which have been widely discussed in the strategic trade policy literature (e.g., see Brander, 1995, for a comprehensive review). When country H moves toward (unilateral) GHG pricing (i.e., $\gamma > 0$), the bracketed terms in (15) and (16) (multiplied by a positive γ) are those additional components to the base case. An interesting catch here is that, once GHG emissions are taken into account, the airport charges may increase or decrease. This result seems contradictory to the traditional wisdom that a positive charge should be imposed to internalize the negative externality (e.g., pollution or GHG emissions in our case). To understand why the signs of the additional terms are undetermined, we make note of three points. First, an increase in airport charges may decrease the airlines' output, thereby reducing GHG emissions. Second, an increase in t_H^U, which is only imposed on airline H, may shift customers to airline F. Thus domestic connecting flights in country H decrease, while those in country F increase. Recall that only the emissions from the former are counted in country H's emissions, but not that from the latter. As a result, the emissions related to country H may decrease. Third, a decrease in airport charges may confer a strategic advantage to its home airline. Note that the home airline will use both the hub and spoke airports, while its foreign counterpart will only use the spoke airport. Thus, the decrease in the airport charges will benefit the home airline more than its foreign counterpart. While the first two have positive effects on the airport charges after GHG emissions are taken into account, the last effect implies that negative GHG charges are possible.

To see how the airline H's profit in (13) affects the airport charges after the unilateral GHG taxes are imposed, we consider a particular case that country H does not take airline H's profit into account (i.e., $\beta = 0$). As a result, the last effect discussed above vanishes. In this case, equations (15) and (16) reduce to:

$$t_H^U = \frac{5e_H}{2(\alpha\tau+5)}\gamma + K_4 > 0, \quad (17)$$

$$t_H^S = \frac{1}{(7-2\alpha\tau)}[\frac{17\alpha\tau}{(\alpha\tau+5)}e_H + 14\delta e_K]\gamma + K_3 > 0, \quad (18)$$

where the denominators are positive by the second-order conditions. K_3 and K_4 are constants. Equations (17) and (18) suggest that, when country H does not take airline H's profit into account, both the airport charges at the hub and spoke airports will increase, after country H implements GHG pricing. Note that the GHG taxes are increasing in γ. In other words, as the GHG abatement cost increases, country H will charge higher taxes on flights. Equations (17) and (18) show that the charge at the spoke airport increases in both e_H and e_K, while the charge at the hub airport increases only in e_H. This observation implies that the two charges serve different roles in the GHG control. Recall that only airline H will be subject to t_H^U. Thus, an increase in t_H^U will decrease the output by airline H, thereby reducing the emissions from the domestic flights in country H. As a result, if the domestic flights in country H are less efficient in GHG emission (a large e_H), a higher t_H^U is necessary. Note that an increase in t_H^U will also increase the output by airline F, thereby increasing the emissions from the domestic flights in country F, though country H does not take this into account. On the other hand, an increase in t_H^S will reduce the output of both airlines, thereby reducing both the emissions from the domestic flights in country H and international flights operated by the two airlines. Thus, t_H^S increases in both e_H and e_K. The above discussion leads to:

Proposition 1 In the price discrimination case, when a country moves toward GHG pricing, its airport charges may increase or decrease, while the charges will increase if the country does not take its airline's profit into account.

We next consider the uniform pricing case, in which country H chooses a single airport tax at the hub and spoke airports (i.e., $t_H^S = t_H^U = \bar{t}_H$) to maximize its social welfare given in (13). The first-order condition with respect to the charge gives rise to the welfare-maximizing pricing rules. We obtain (see the derivation in Appendix):

$$\bar{t}_H = \frac{35}{\Omega_2}(11e_H + 15\delta e_K)\gamma + K_5, \tag{19}$$

where $\Omega_2 = -401\alpha\tau + 814\beta\tau - 1540\beta + 1820 > 0$ by the second-order condition. K_5 is a constant.

Equation (19) suggests that when country H moves toward GHG pricing (i.e., from $\gamma = 0$ to $\gamma > 0$), the (single) airport charge will increase. The airport charge is a weighted average of e_H and e_K. The weight of e_K depends on δ, the discount factor of the international emissions in counting country H's GHG emissions. Note that for $\delta = 1$, the weight of e_H is less than that of e_K in the airport charge given in (19). This result may be due to the existence of a trade-off in country H's airport pricing between domestic emission reduction in country H and airline H's profit. In particular, a measure to reduce the domestic emissions, which are only from the flights operated by airline H, will increase airline H's operating cost, but have no effect on that for airline F. As a result, this will confer a strategic advantage to airline F. Yet this tradeoff does not exist in a measure to reduce the international emissions, as it affects both airlines' operating cost by the same amount. As a result, for $\delta = 1$, country H may have more incentives to reduce the international emissions than domestic ones. Finally, it is important to note that, with uniform pricing, the effects of the GHG measure on airline H will be larger than that on airline F. This is due to the fact that airline H will be subject to the airport charge at both the hub and spoke airports in country H, while airline F will only incur such charge at the spoke airport but not at the hub airport in country H. As equation (19) suggests that the airport charge will increase after the imposition of GHG pricing, the pricing may confer a strategic advantage to airline F in the competition. As a result, the GHG pricing may raise competitive concerns. This leads to the following proposition:

Proposition 2 In the uniform pricing case, when a country moves toward GHG pricing, its airport charge will increase, which will place its airline at a strategic disadvantage.

Before analyzing how the unilateral GHG pricing affects the airline competition and total emissions, it is important to note that the results in this section may have significant implications for empirical studies on the issue. We show that the implementation of the GHG pricing may have different impacts on home and foreign airlines. As a result, the pricing may alter the strategic relationship between the home and foreign airlines. If this effect is not taken into account in the empirical studies on the issue, the estimation in those studies may underestimate the effects of the GHG pricing on the welfare of the country implementing the measures and its airlines' profits. On the other hand, the empirical studies may overestimate the effects on consumers, as it does not account for their option to choose an airline less affected by the GHG control measures.

6. Effects of Unilateral GHG Pricing

In this section, we further investigate the effects of unilateral GHG pricing examined in the previous section on market outputs, consumer surplus and GHG emissions. We focus on two particular cases: the price discrimination case in which airline H's profit is not taken into account in country H's airport pricings (i.e., $\beta = 0$) and the uniform pricing case. Note that as the change of airport charges in the general price discrimination case is undetermined as shown in the previous section, its effects on market outputs and emissions are also generally undetermined.

6.1. Market Outputs and Consumer Surplus

For the price discrimination case with $\beta = 0$, substituting equations (17) and (18) into (10) gives the following expressions for the changes in market outputs under unilateral GHG pricing (superscripts *A* and *B* for variables after and before the imposition of the unilateral GHG pricing, respectively):

$$q_H^A - q_H^B = -\frac{\gamma[(12-\alpha\tau)e_H + 2(5+\alpha\tau)\delta e_K]}{4(\alpha\tau+5)(7-2\alpha\tau)} < 0, \tag{20}$$

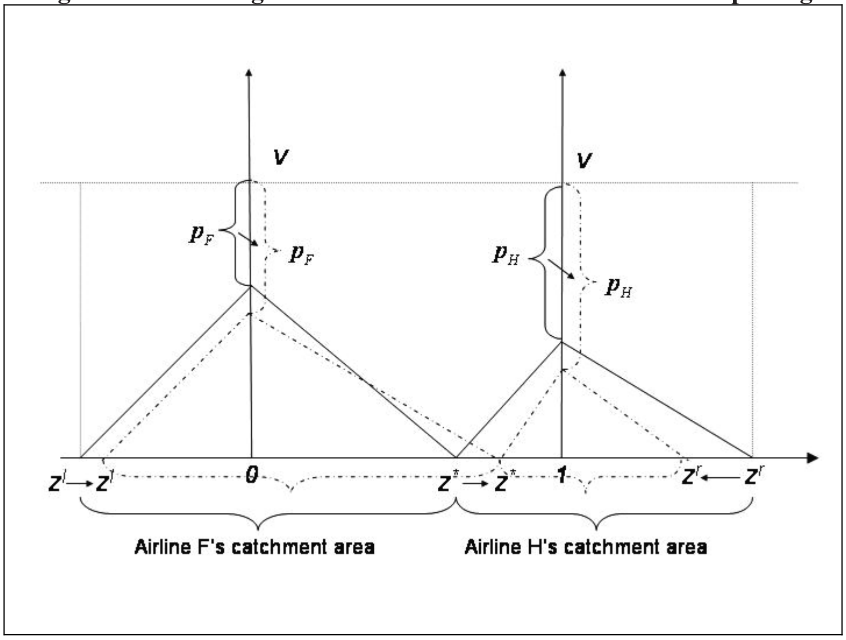
$$q_F^A - q_F^B = -\frac{\gamma[(2-3\alpha\tau)e_H + 2(5+\alpha\tau)\delta e_K]}{4(\alpha\tau+5)(7-2\alpha\tau)} < or > 0, \tag{21}$$

$$Q^A - Q^B = -\frac{\gamma(e_H + 2\delta e_K)}{2(7-2\alpha\tau)} < 0, \tag{22}$$

where the denominators are positive by the second-order conditions for country H's welfare maximization problem.

Equation (20) suggests that airline H's output will decrease after the imposition of unilateral GHG pricing. After the policy change, both airline H and F's (equilibrium) fares will increase, while the increase for the former will be more than that for the latter (see equations (11) and (12)). Given this, we can break down the decrease in airline H's output into two parts (also see Figure 3): (i) an increase in airport charges will increase the fare charged by airline H, and lead to a contraction of airline H's market at the region where consumers located at $z > 1$ (i.e., z^r decreases); (ii) the airline H's market share in the market for consumers with z between 0 and 1 will decrease (i.e., $\tilde{z}$ increases). As mentioned, an increase in airport charges will have a stronger (positive) effect on airline H's fare than that on airline F's fare. Thus, more consumers in the region of $0 \le z \le 1$ will choose airline F over airline H. Equation (21) suggests that the airline F's output may increase or decrease. The result is undetermined because airline F's market at the region where consumers located at $z < 0$ will contract (i.e., $|z^l|$ decreases), while its market share in the market for $0 \le z \le 1$ will increase.

Figure 3: The change in market size and its share after GHG pricing



Nevertheless, equation (22) suggests that the total market output will decrease after the imposition of the unilateral GHG pricing. Note that as the market size in the region of $0 \le z \le 1$ does not change, the decrease in the total market output is due to the decrease in the market at the two extremes (i.e., $z < 0$ and $z > 1$). On the other hand, we can conclude that consumer surplus of the air travel market will decrease after the imposition of the unilateral GHG pricing. It is because, on one hand, fewer consumers will choose to travel by air (i.e., Q decreases). On the other hand, passengers traveling on both airlines H and F will face higher fares. This result suggests that it is important to consider the adverse impact on consumer benefits, when a country plans to implement GHG control measures.

Proposition 3 In the price discrimination case, when a country moves toward GHG pricing without taking its airline's profit into account, its airlines' and total market outputs will decrease, while that of the foreign airline may increase or decrease. Consumer surplus will decrease.

For the uniform pricing case, by substituting equation (19) into (10), the changes in market outputs under the unilateral GHG pricing are given below:

$$q_H^A - q_H^B = -\frac{11(11e_H + 15\delta e_K)\gamma}{\Omega_2} < 0, \tag{23}$$

$$q_F^A - q_F^B = -\frac{4(11e_H + 15\delta e_K)\gamma}{\Omega_2} < 0, \tag{24}$$

$$Q^A - Q^B = -\frac{15(11e_H + 15\delta e_K)\gamma}{\Omega_2} < 0. \tag{25}$$

Equations (23) to (25) suggest that the total market output, and airline H and F's outputs will decrease after country H implements GHG pricing. Comparing (23) and (24), we can see that the decrease in airline H's output will be greater than that for airline F's output. This is due to fact that airline H will be charged twice (at the hub and spoke airports in country H) by the expression given in (18), while airline F will be charged once (at the spoke airport in country H). Thus, the airline H's fare will increase more than that for airline F's fare. As a result, airline H's market share in the market for $0 \le z \le 1$ will decrease (i.e., $\tilde{z}$ increases). On the other hand, as in the price discrimination case, the decrease in total market output in (25) is due to the decrease in the market at the two extremes (i.e.,

$z<0$ and $z>1$). For the consumer surplus, since airfares increase and output decreases, consumers will be worse off after the imposition of unilateral GHG pricing.

Proposition 4 In the uniform pricing case, when a country moves toward GHG pricing, both its home and foreign airlines' outputs will decrease, and consumer surplus will decrease.

6.2. GHG Emission

In this section, we examine how the unilateral GHG control measure will affect world emissions. This section may be related to the pollution haven hypothesis discussed in the environmental economics literature (see, e.g., Copeland and Taylor, 1994), which suggests that industries that are highly (local) pollution-intensive may migrate from high-income countries to its low-income counterparts, causing world pollution to increase.[23] Yet note that the effect of GHG emissions is global: its impact is independent of where in the world the emissions occur. In the following, we will explore a possibility that the aviation GHG emissions may be shifted from a country to another, which may lead to an increase in world emissions. As in the previous section, we focus particularly on the price discrimination case, in which airline H's profit is not taken into account in country H's airport pricings (i.e., $\beta=0$) and on the uniform pricing case. The world emissions are the sum of the emissions from airlines H and F, thus:

$$E=q_H(e_H+e_K)+q_F(e_F+e_K), \tag{26}$$

where e_F is the (per-unit) emissions due to the connecting flights operated by airline F between the hub and spoke airports in country F (i.e., the route between B and F as shown in Figure 1). Thus, the first term on the RHS of (26) is the total emissions from airline H's flights, and the second term is those from airline F's flights. Note that domestic flights in a country with a lower e_i are more efficient in terms of GHG emission than that in another country. Thus, the difference between e_H and e_F may capture the difference of factors affecting emission intensity of domestic routes in countries H and F, including aircraft fuel efficiency, route structure and load factor.

For the price discrimination case with $\beta=0$, by substituting equations (20) and (21) into (26), and rearranging the equation gives:

$$E^A-E^B$$
$$=-\frac{\gamma[(12-\alpha\tau)e_H+2(5+\alpha\tau)\delta e_K]}{4(\alpha\tau+5)(7-2\alpha\tau)}e_H-\frac{\gamma(e_H+2\delta e_K)}{2(7-2\alpha\tau)}e_K-\frac{\gamma[(2-3\alpha\tau)e_H+2(5+\alpha\tau)\delta e_K]}{4(\alpha\tau+5)(7-2\alpha\tau)}e_F \tag{27}$$

Equation (27) suggests that the total GHG emissions may increase or decrease after the imposition of unilateral GHG pricing. It is a very interesting result. To deal with the GHG problem, many environmental groups advocate that given the absence of an international agreement, a country ought to take unilateral actions to reduce its environmentally harmful emissions, as they at least contribute in the right direction (Hoel, 1991). Yet here we show that a country's unilateral action may possibly worsen the problem rather than improve it. To understand the intuition behind the result, we need to further examine the factors affecting the sign of world emission change. From equation (27), it can be shown that:

[23] Empirical studies of the pollution haven hypothesis can be found in, for example, Grossman and Krueger (1991), Xu (2000), Eskeland and Harrison (2003), Kahn (2003), Levinson and Taylor (2004), and Ederington, et al. (2005).

$$E^A - E^B > 0$$

if and only if (i) $e_F > \frac{(12-\alpha\tau)e_H^2 + (5+\alpha\tau)[2e_H(\delta+1)+4e_K\gamma]e_K}{(3\alpha\tau-2)e_H - 2(5+\alpha\tau)\delta e_K}$,

and (ii) $(2-3\alpha\tau)e_H + 2(5+\alpha\tau)\delta e_K < 0$. (28)

Condition (28) suggests that total GHG emissions will increase after the imposition of the unilateral GHG pricing, if and only if the (per-flight) emissions for the domestic flights in country F (i.e., e_F) is larger than a threshold (i.e., (i) in (28) holds) and the domestic flights in country F increases after country H implements the GHG pricing (it would be the case when (ii) in (28) holds, see equation (21)). In Proposition 3, we find that domestic flights in country H and total output will decrease after imposing the GHG pricing, while the domestic flights in country F may increase. Thus, the emissions from domestic flights in country H and international flights will decrease, while that in country F may increase. These two opposing effects on total emissions depend on the emission intensity of the three routes. When e_F is very large, the latter effect may dominate. In other words, the unilateral GHG pricing may result in a shift of domestic flights from one country to another, which may have a less efficient domestic network. As a result, the unilateral GHG pricing may increase world emissions. Although this paper does not explicitly consider global cooperation in control measures, the potential adverse impact of the unilateral GHG control action on world emissions control may provide an argument to support international cooperation on the issue.

Proposition 5 In the price discrimination case, when a country moves toward GHG pricing, total emissions may increase or decrease, depending on aircrafts' emission efficiency in that country and its foreign counterparts.

Proposition 5 may have significant implications for implementing GHG control measures unilaterally, like the inclusion of the aviation industry in the EU ETS. As EU airlines operating flights between EU airports may face a stricter regulation on GHG emissions (i.e., need to buy permits for their emissions), they may be required to charge a higher fare to their customers. As a result, more consumers may choose non-EU airlines, and thereby increasing hubbing activities outside the EU. For example, after the inclusion of the aviation sector in the EU ETS, it is expected that more domestic connecting flights in the EU for hubbing will be shifted to the U.S., as US airlines, which usually operate their hubs in the U.S., may gain a larger market share of the transatlantic market (Scheelhaase, et al. 2008). On the other hand, Table 1 shows the emission intensity of domestic flights in some selected routes in the U.S. and the EU. It suggests that the US domestic flights are less efficient than that in the EU in terms of GHG emissions. Thus, to evaluate the effectiveness of the inclusion of the aviation industry in the EU ETS, it is important to take the change in GHG emissions outside the EU into account, as the unilateral GHG control measure may shift flights from a more efficient network (e.g., in the EU) to a less efficient one (e.g., in the U.S.).[24] However, the change in GHG emissions outside the EU due to the EU ETS is usually ignored in the present empirical studies (e.g., Delft, 2005; EC, 2006).

[24] Note that for the U.S. and the EU moving towards more liberalizing air market, EU airlines may allow operating their hubs in the U.S. Given a stricter regulation on GHG emissions in the EU, EU airlines may operate more flights through their U.S. hubs.

Table 1a: CO_2 Emissions of selected routes in US domestic markets[a]

From New York to	Per-passenger CO_2 emission (kg)*	Distance (km)	Emission per distance (kg/km)
Columbus	130.3	774	0.168
Detroit	131.4	816	0.161
Louisville	154.6	1061	0.146
Chicago	157.6	1186	0.133
New Orleans	197.4	1889	0.105
Denver	206.7	2607	0.079
Average	163.0	1389	0.132

Table 1b: CO_2 Emissions of selected routes in EU domestic markets

From London to	Per-passenger CO_2 emission (kg)*	Distance (km)	Emission per distance (kg/km)
Berlin	97.0	946	0.103
Milan	114.2	978	0.117
Barcelona	110.7	1146	0.097
Rome	134.8	1442	0.093
Lisbon	141.7	1564	0.091
Athens	217.6	2242	0.097
Average	136.0	1386	0.099

*Economy class

Source: ICAO Carbon Emissions Calculator

[a.] It is shown from the two tables that emission per km decline strongly in the U.S., but not in the EU. Note that the emissions per passenger tend to fall with distance because a significant portion of flight emissions occur during taking-off and landing, and also because larger aircraft are used for long-distance flight which produce fewer emissions per passenger (e.g., Schipper, 2004).

For the uniform pricing case, by substituting equations (23) and (24) into (26), the change in total GHG emissions with the unilateral GHG pricing is given below:

$$E^A - E^B = -\frac{\gamma}{\Omega_2}(11e_H + 15e_K)(15e_K + 11e_H + 4e_F) < 0. \qquad (29)$$

Equation (29) suggests that the total GHG emissions will decrease if a country unilaterally imposes GHG pricing. As shown in Proposition 4, domestic flights in countries H and F as well as international flights will decrease after imposing unilateral GHG pricing. Thus, the reduction in GHG emissions by (29) is the sum of the emissions reduction from the three routes. As shown in (29), the reduction is increasing in the (per-unit) emissions in the three routes (i.e., e_H, e_F and e_K).

Proposition 6 In the uniform pricing case, when a country moves toward GHG pricing, total GHG emissions will decrease.

7. Concluding Remarks

The main objective of this paper is to investigate the effects of unilateral GHG pricing implemented at a country's airports on the airline market competition and world emissions. We explored both the price discrimination case and the uniform pricing case, in which a country imposes a single charge at both hub and spoke airports. In the price discrimination case, we found that the unilateral GHG pricing may be positive or negative. This is due to the fact that positive charges are imposed to internalize the negative externality, while the country may subsidize airlines, which may confer a strategic advantage to its home airlines. Yet if the home airline's profit is not taken into account or in the uniform pricing case, the airport charges will increase when the country unilaterally implements GHG pricing. In such cases, a strategic advantage is conferred to foreign airlines. It is because the effects of such (positive) GHG charges have a less adverse impact on the foreign airlines than on the home airlines. Furthermore, the increase in airport charges due to the unilateral GHG pricing leads to a reduction in total market output, and the home airlines' market share will decrease. Consumer surplus of the air travel market will decrease. Finally, in the price discrimination case, total GHG emissions may increase after the unilateral GHG pricing is imposed. This may be due to the fact that domestic flights connecting hub and spoke airports may be shifted from a more efficient flight network to a less efficient one.

The paper has also raised a number of other issues and avenues for future research. First, our analysis assumes that the airlines' network structure is fixed. Given the airline industry moves toward more liberalization, airlines will have more flexibility to choose different networks for their flight operations. From a forward-looking perspective, it might be important to take the response of airlines network choice into account, when we examine the effects of GHG control measures on the airline industry. For example, note that the emission intensity is increasing in flight distance and the number of taking-offs and landings. As a result, given GHG pricing, airlines may be less likely to use HS networks than FC networks, since the HS networks usually involve longer flying distances and more taking-offs and landings, which would incur higher costs due to the GHG pricing. On the other hand, the unilateral GHG control measure may also lead to inefficient airlines network choice: For example, Albers, et al. (2008) argue that airlines may add an intermediate stop outside of the EU on long international flights, so that the EU emissions charge only applies to the final short leg. As a result, the flying distance and the number of taking-offs and landings may increase, thereby increasing world emissions (see, e.g., Morrell and Lu, 2007, for more discussion about the relationship between airlines network choice and environmental costs).

Second, we take the emission technology of aircrafts to be exogenous. In the long run, airlines may control its aircraft emissions by changing, for example, its speed of aircraft replacement, aircraft size and load factor. Thus, it is also practically relevant to take the airlines' decisions which affect emission technology into account, when we examine the effects of GHG control measures. Finally, this paper considers a GHG pricing in a single market – the aviation market. Several empirical studies have been conducted to examine the side effects of the pricing on other industries, for example, tourism (Forsyth, 2008; Forsyth and Ho, 2008), automobile (Hofer, et al., 2008) and high-speed railways (Benito, 2008). Thus, it is interesting to take other related industries into account in our theoretical analysis.

Finally, this paper only considered GHG emissions, while other externalities due to air transportation, such as congestion, noise and local air pollution, are omitted in the analysis. Note that contemporary estimates (e.g., Quinet and Vickerman, 2004; Schipper, 2004) indicate that the environmental costs of noise and air pollution are at least as large as the costs of GHG emissions. The presence of these other externalities would not be a concern if they were independent of GHG emissions. However, they are likely to be interdependent to a degree. Congestion probably contributes to air pollution and GHG emissions through greater fuel consumption as aircraft fly longer routes, queue on taxiways, and circle airports waiting to land. For similar reasons, congestion contributes to noise. More generally, efforts to reduce one externality are likely to affect the magnitude and spatial distribution of other externalities. It would be interesting to extend this paper to consider the interdependence of GHG emissions and other externalities into account in the analysis.

8. Acknowledgement

The authors would like to thank David Gillen, Robin Lindsey, and Tae Oum for valuable comments on earlier drafts of this paper. All remaining errors are the sole responsibility of the authors.

References

Albers, S., Buhne, J. and Peters, H. (2008) Will the EU-ETS Instigate Airline Network Reconfigurations? Journal of Air Transport Management, forthcoming.

Andersen, A. (2001) Emissions Trading for Aviation, a study commissioned by the International Air Transport Association, London.

Basso, L.J. and Zhang, A. (2007a) Congestible Facility Rivalry in Vertical Structures, Journal of Urban Economics, 61, 218-237.

Basso, L.J. and Zhang, A. (2007b) An Interpretative Survey of Analytical Models of Airport Pricing, in: Lee, D. (Ed.), Advances in Airline Economics, Vol.2, Elsevier, 89-124.

Benito, A. (2008) Climate Change Impact as a Constraint to the Air Transport Growth: The European Case, Chapter presented at World Conference on Transport Research, Berkeley, California.

Brander, J.A. (1995) Strategic Trade Policy, in: Grossman, G. and Rogoff, K. (Eds.), Handbook of International Economics, Vol. III, Elsevier Science, Amsterdam.

Brander, J.A. and Zhang, A. (1990) Market Conduct in the Airline Industry: An Empirical Investigation, RAND Journal of Economics, 21, 567–583.

Brueckner, J.K. (2004) Network Structure and Airline Scheduling, Journal of Industrial Economics, 52, 291-312.

Brueckner, J.K. and Flores-Fillol, R. (2007) Airline Schedule Competition, Review of Industrial Organization, 30, 161-177.

Brueckner, J.K. and Girvin, R. (2008) Airport Noise Regulation, Airline Service Quality, and Social Welfare, Transportation Research Part B, 42, 19-37.

Brueckner, J.K. and Zhang, Y. (2001) A Model of Scheduling in Airline Networks: How a Hub-and-Spoke System Affects Flight Frequency, Fares and Welfare, Journal of Transport Economics and Policy, 35, 195-222.

Brueckner, J.K. and Zhang, A. (2009) Airline Emission Charges: Effects on Airfares, Service Quality, and Aircraft Design, manuscript.

Caves, D.W., Christensen, L.R. and Tretheway, M.W. (1984) "Economies of Density versus Economics of Scale: Why Trunk and Local Service Airline Costs Differ", RAND Journal of Economics, 15(4), 471–489.

Chua, C.L., Kew, H. and Yong, J. (2005) Airlines Code-Share Alliances and Costs: Imposing Concavity on Translog Cost Function Estimation, Review of Industrial Organization, 26(4), 461-487.

Ciliberto, F. and Williams, J.W. (2007) Limited Access to Airport Facilities and Market Power in the Airline Industry, Working paper, Department of Economics, University of Virginia.

Copeland, B.R. and Taylor, M.S. (1994) North-South Trade and the Environment, Quarterly Journal of Economics, 109, 755–787.

Delft, C.E. (2005) Giving Wings to Emission Trading – Inclusion of Aviation under the European Emission Trading System (ETS): Design and Impacts, study commissioned by the European Commission, Delft.

Delft, C.E. (2007) Allocation of Allowances for Aviation in the EU ETS: The Impact on the Profitability of the Aviation Sector under High Levels of Auctioning. Final report, Delft.

Douglas, G.W. and Miller, J.C. (1974) "Quality Competition, Industry Equilibrium, and Efficiency in the Price-constrained Airline Market", American Economic Review, 64 (4), 657-669.

EC (2006) Full Assessment Report, accompanying document to the Proposal for a Directive of the European Parliament and of the Council, European Commission.

Ederington, J., Levinson, A. and Minier, J. (2005) Footloose and Pollution-free, Review of Economics and Statistics, 87, 92-100.

Eskeland, G.S. and Harrison, A.E. (2003) Moving To Greener Pastures? Multinationals and The Pollution Haven Hypothesis, Journal of Development Economics, 70, 1-23.

Forsyth, P. (2008a) Emissions Trading Schemes – Should International Aviation be Excluded?, manuscript.

Forsyth, P. (2008b) The Impact of Climate Change Policy on Competition in the Air Transport Industry Discussion Paper No. 2008-18, Organization for Economic Co-operation and Development (OECD), International Transport Forum.

Forsyth, P. and Ho, T.V. (2008) Aviation's Indirect Carbon Emissions: Measurement and Implications, Paper for the German Aviation Research Society Workshop Aviation and Climate Change, Amsterdam, July.

Grossman, G.M. and Krueger, A.B. (1991) Environmental Impacts of a North American Free Trade Agreement, NBER Working Paper No. 3914, National Bureau of Economic Research, Cambridge, Massachusetts.

Hansen, M.M., Gillen, D., Djafarian-Tehrani, R. (2001) "Aviation Infrastructure Performance and Airline Cost: A Statistical Cost Estimation Approach", Transportation Research Part E, 37, 1-23.

Hendricks, K., Piccione, M. and Tan, G. (1995) The Economics of Hubs: The Case of Monopoly, Review of Economic Studies, 62, 83-99.

Hendricks, K., Piccione, M. and Tan, G. (1999) Equilibria in Networks, Econometrica, 67, 1407-1434.

Hodgkinson, D., Coram, A. and Garner, R. (2007) Strategies for Airlines and Aircraft Emissions and Climate Change: Sustainable, Long-Term Solutions, Hodgkinson Group, Working Paper No 2, June.

Hoel, M. (1991) Global Environmental Problems: The Effects of Unilateral Actions Taken by One Country, Journal of Environmental Economics and Management, 20, 55-70.

Hofer, C., Dresner, M. and Windle, R. (2008) The Environmental Effects of Airline Carbon Emissions Taxation in the U.S., Paper presented at Air Transport Research Society Conference, Athens.

IATA (2006) Economics Briefing Impact of the Rise in UK Air Passenger Duty, International Air Transport Association.

ICAO (2007) Report on Voluntary Emissions Trading for Aviation, Preliminary Edition, International Civil Aviation Organization.

IPCC (2007) Fourth Assessment Report, Technical Summary, Intergovernmental Panel on Climate Change.

Kahn, M.E. (2003) “The Geography of US Pollution Intensive Trade: Evidence from 1958 to 1994”, Regional Science and Urban Economics, 33, 383-400.

Kumbhakar, S.C. (1990) “A Reexamination of Returns to Scale, Density and Technical Progress in U.S. Airlines”, Southern Economic Journal, 57, 428-442.

Kumbhakar, S.C. (1992) “Allocative Distortions, Technical Progress, and Input Demand in U.S. Airlines: 1970–1984”, International Economic Review, 33, 723-737.

Levinson, A. and Taylor, M.S. (2004) “Unmasking the Pollution Haven Effect”, NBER Working Paper No.10629, National Bureau of Economic Research, Cambridge, Massachusetts.

Morrell, P. and Lu, C. (2007) The Environmental Cost Implication of Hub-hub versus Hub By-pass Flight Networks, Transportation Research Part D, 12, 143-157.

Olsthoorn, X. (2001) Carbon Dioxide Emissions from International Aviation 1905-2050, Journal of Air Transport Management, 7, 87-93.

Oum, T.H., Zhang, A. and Zhang, Y. (1993) Inter-firm Rivalry and Firm-specific Price Elasticities in Deregulated Airline Markets, Journal of Transport Economics and Policy, 27, 171-192.

Oum, T.H., Zhang, A. and Zhang, Y. (1995) Airline Network Rivalry, Canadian Journal of Economics, 28, 836-857.

Quinet, E. and Vickerman, R. (2004), Principles of Transport Economics, Cheltenham and Northampton, Mass: Edward Elgar.

Scheelhaase, J.D. and Grimme, W.G. (2007) Emissions Trading for International Aviation – An Estimation of the Economic Impact on Selected European Airlines, Journal of Air Transport Management, 13, 253-263.

Scheelhaase, J.D., Grimme, W.G. and Schaefer, M. (2008) The Inclusion of Aviation into the EU Emission Trading Scheme – Impacts on Competition between European and non-European Network Airlines, manuscript.

Schipper, Y. (2004) Environmental Costs in European Aviation, Transport Policy, 11, 141-154.

Small, K.A. and Verhoef, E.T. (2007) The Economics of Urban Transportation, Routledge, London.

Stern, N. (2006) Stern Review Report on the Economics of Climate Change, a report to the Prime Minister and the Chancellor of the Exchequer on the Economics of Climate Change, United Kingdom.

Ubbels, B., and Verhoef, E.T. (2008) Governmental Competition in Road Charging, Regional Science and Urban Economics, 38, 174-190.

World Bank (2007) State and the Trends of Carbon Market 2007, The World Bank Institute, Washington, D.C.

Xu, X. (2000) International Trade and Environmental Regulation: Time Series Evidence and Cross Section Test, Environmental and Resource Economics, 17, 233-257.

Yuen, A., Basso, L. and Zhang, A. (2008) Effects of Gateway Congestion Pricing on Optimal Road Pricing and Hinterland, Journal of Transport Economics and Policy, 42(3), 495-526.

Zhang, A., Zhang, Y. and Clougherty, J.A. (2008) Competition and Regulation in Air Transport, Handbook of Transport Economics, Edward Edgar, forthcoming.

Appendix

Derivation of GHG pricing for the price discrimination case

To solve the welfare maximization problem in (13), we first derive consumer surplus CS:

$$CS = \int_0^{|z_l|}[V - p_H(q_H, q_F) - 4\tau z]dz + \int_0^{\tilde{z}}[V - p_H(q_H, q_F) - 4\tau z]dz$$
$$+ \int_0^{1-\tilde{z}}[V - p_F(q_H, q_F) - 4\tau z]dz + \int_0^{z_F}[V - p_F(q_H, q_F) - 4\tau z]dz$$

Using (4), (5) and (6), we can solve the integrals to obtain:

$$CS = \tau(q_F^2 + 6q_F q_H + q_H^2 - 4)/2\,. \quad \text{(A.1)}$$

Substitute the inverse demand system (6) into airline H's profit function, we have:

$$\pi_H = (2\tau + V - c)q_H - (t_H^S + t_H^U + t_F^S)q_H - 3\tau q_H^2 - \tau q_H q_F\,. \quad \text{(A.2)}$$

With (A.1) and (A.2), the welfare maximization problem (13) can be written as:

$$\begin{aligned} SW_H &= \alpha[\tau(q_F^2 + 6q_F q_H + q_H^2 - 4)/2] + \beta[(2\tau + V - c)q_H \\ &\quad - t_F^S q_H - 3\tau q_H^2 - \tau q_H q_F + t_H^S q_F - \alpha[e_H q_H + \delta e_K(q_H + q_F)] \end{aligned} \quad \text{(A.3)}$$

The first-order conditions for the maximization problem with respect to t_H^S and t_H^U are,

$$\frac{dSW_H}{dt_H^S} = \frac{\partial SW_H}{\partial q_H}\frac{\partial q_H}{\partial t_H^S} + \frac{\partial SW_H}{\partial q_F}\frac{\partial q_F}{\partial t_H^S} + \frac{\partial SW_H}{\partial t_H^S} = 0\,, \quad \text{(A.4)}$$
$$\frac{dSW_H}{dt_H^U} = \frac{\partial SW_H}{\partial q_H}\frac{\partial q_H}{\partial t_H^U} + \frac{\partial SW_H}{\partial q_F}\frac{\partial q_F}{\partial t_H^U} + \frac{\partial SW_H}{\partial t_H^U} = 0\,. \quad \text{(A.5)}$$

and assuming the second-order conditions are satisfied.

By solving conditions (A.4) and (A.5),

$$t_H^U = \{\frac{5[14 - \beta + \tau(\beta - 4\alpha)]}{\Omega_1}e_H + \frac{30(2-\tau)\beta\delta}{\Omega_1}e_K\}\gamma + K_1\,,$$
$$t_H^S = \{\frac{6\beta + 17\alpha\tau - 6\beta\tau}{\Omega_1}e_H + \frac{[70 - 65\beta + (29\beta + 14\alpha)\tau]\delta}{\Omega_1}e_K\}\gamma + K_2\,.$$

where $\Omega_1 = -(\tau-1)^2\beta^2 + 2(34\tau + 3\alpha\tau - 70)\beta - 4(2\alpha\tau + 3)\alpha\tau + 140 > 0$ by the second-order conditions; and

$$K_1 = \frac{1}{\Omega_1}\{[(\tau-1)^2\beta^2 + [69 - 3\tau(2\alpha+11)]\beta + \alpha\tau(8\alpha\tau+9) - 70]t_F^U$$
$$+[(\tau-1)^2\beta^2 - [75 + 3\tau(2\alpha-13)]\beta + \alpha\tau(8\alpha\tau+26) - 70]t_F^S$$
$$-[2\tau^3 + (V-c-4)\tau^2 + 2(c-V+1)\tau + V - c]\beta^2$$
$$-[-2(6\alpha+39)\tau^2 + 3[(2\alpha+13)(c-V)+50]\tau + 75(V-c)]\beta$$
$$-16\alpha^2\tau^3 - 4\alpha\tau^2[13+2\alpha(V-c)] + (70-26\alpha\tau)(V-c) + 140\tau\}$$

$$K_2 = \frac{1}{\Omega_1}\{-[(\tau-1)^2\beta^2 - [75 - 3\tau(2\alpha+11)]\beta + 8\alpha\tau(\alpha\tau-1) - 70]t_F^U - 30\beta(\tau-2)t_F^S$$
$$-30\beta[2(V-c) - (4-V+c)\tau + 2\tau^2]\}$$

Derivation of GHG pricing for the uniform pricing case

The first-order condition for the maximization problem with respect to $\bar{t}_H$ is,

$$\frac{dSW_H}{d\bar{t}_H} = \frac{\partial SW_H}{\partial q_H}\frac{\partial q_H}{\partial \bar{t}_H} + \frac{\partial SW_H}{\partial q_F}\frac{\partial q_F}{\partial \bar{t}_H} + \frac{\partial SW_H}{\partial \bar{t}_H} = 0, \tag{A.6}$$

and assuming the second-order condition is satisfied.

By solving conditions (A.6),

$$\bar{t}_H = \frac{35}{\Omega_2}(11e_H + 15\delta e_K)\gamma + K_5,$$

where $\Omega_2 = -401\alpha\tau + 814\beta\tau - 1540\beta + 1820 > 0$ by the second-order condition; and

$$K_5 = \frac{1}{\Omega_2}\{[(655-401\tau)\beta + 499\alpha\tau - 655]\bar{t}_F + [(810\tau - 1470)\tau + (405\tau - 735)(V-c)]\beta$$
$$-25[24\alpha\tau^2 + 6[2\alpha(V-c) - 7]\tau - 21(V-c)]\} .$$

Research on the construction method of container yard microanalysis simulation model

Jieshu Zhang

Shanghai International Shipping Institute, Shanghai Maritime University, Shanghai, China.

Abstract

Container yard daily production organization has the characteristics of micro and detail, but in previous simulation studies, details about container yard have been described roughly, and the entities have been simplified. In this case, the paper presents the concepts about CYMSM (CYMSM: Container Yard Microanalysis Simulation Model), analyses characteristics of CYMSM, makes studies on CYMSM construction method in a general sense, and describes the four elements separately, which are Entity, Event, Activity & Process (EEAP). This paper explores the operation rule which influences the validity of CYMSM, based on this, it has made a relatively simulation mechanism. This paper has also offered the evaluation index and adjustment mechanism of container yard simulation model based on the microscopic analysis. The results of study can provide evidence for building container yard operation simulation system. And this system, for the purpose of offering strategy supports, has been applied to certain container yard in Shanghai, thus proving the effectiveness of this method.

Keywords: Container Yard Microanalysis Simulation Model, Microanalysis Simulation, simulation mechanism, evaluation index, adjustment mechanism

1. Introduction

With the rapid development of China's foreign trade, the status of container transport has become increasingly important. Under the ever-increasing traffic circumstances, many of the actual throughput of container ports has already exceeded the design throughput capacity, and the yard operation is in overload operation state. To expand container port admission ability, new-built and expanded port is for one aspect, but how to make full use of the operating capacity of the existing yard is also an important subject, and the key to tap the existing yard capacity is that the implementation of the modern port management. Compared with container ports at home and abroad, it can be found that the Chinese port in the modern production facilities does not have too much differences from foreign ports, but in the production organizations, experience is mainly carried out, therefore, the yard managers are in urgent need to increase the level of production operation, especially the level of container yard production management with huge investment.

System simulation, based on the real system model being studied, is a method making use of computers for the purpose of doing an experimental study. Owing to the complexity of the port operation, many scholars have pointed out that the simulation has become an important technology for ports plan, analysis, and decision-making (Dahal et al., 2003).

According to the purpose of simulation study, port simulation could be divided into two categories: Firstly, providing decision supports for new building, expansion, or rebuilding of the yard. In such matters, research of port handling capacity (Ward, 1995), mechanical selection (Yong, 1993), comparison of handling technology (Liu et al., 2001) and such issues are common; Secondly, providing production operation decisions for the port management after completion (Jinglei and Yizhong, 2006)(Bin et al., 2008). This type of problems is rather complicated. In recent years it began to enter the field of simulation research. This thesis will discuss the second question, which is providing managing decisions for port management. In the research literatures concerning with the second question, the establishment of container yard simulation system is mainly based on macro analysis, among these, important parameters are the distribution of ships to ports, mechanical operating time distribution, distribution of container type and container quantity, etc. The Simulation model takes the day as a step size, one year or more as an overall simulation time in general, which mainly providing long-term decision-making for port managers (for example, the suitable machinery number (Degano and Di, 2001), macro-scheduling principles (Sen et al., 1991). But the simulation model which gives a very detailed description of entity and behaviour elements based on the micro-simulation technology is rare, whose overall simulation time is in the limited time (eg. 48 hours) and the research object is single entity (such as vehicles, gantry cranes). The entities of "micro-model" are basic units and elements in port production organization such as one single gantry crane, container, truck, and these entities' micro-behaviour, such as loading, unloading, operating and waiting, which can be simulated in great detail. Considering the characteristics of container yard day-to-day production operations, the establishment of micro-simulation model has practical significance. This thesis has done a lot of research on this.

2. Concept and Characteristics of Container Yard Micro-analysis Simulation Model

This paper has put forward the concept of CYMSM (Container Yard Microanalysis Simulation Model): The CYMSM is based on the system simulation basic theory. Single entity (such as container, handling machinery, the level handling machinery, etc.) is regarded as the research object. The goal is to replicate each entity's internal changes in time and space of container yard. The establishment of this detailed and realistic simulation model, which can provide an analysis, judgment and decision-making tool for the container yard daily production operations.

This concept has the following key elements:

Firstly, CYMSM regards a single entity (such as handling machinery) as the basic research unit. The micro-behaviour of these entities, such as handling machinery's condition of loading, unloading, and waiting, etc., can be reflected in a real and detailed way.

Secondly, CYMSM gives a description for individual characteristics of each single entity, such as the operating efficiency, operating time of one handling machine, etc. Through a true reflection of all the individual characteristics of the system, the overall characteristics of the system is reflected, which is unprecedented in the container yard simulation model based on macro-analysis.

Thirdly, because the micro-simulation model is able to calculate performance characteristics of each entity at every moment, it can provide detailed information for managers. Therefore, it is very useful for CYMSM to evaluate the container yard from the operational level, and it provides container yard managers with daily operation decision support.

Table 1 compares CYMSM with the general container yard simulation model, reflecting the unique characteristics of CYMSM.

Table 1: A Comparison between CYMSM and the General Container Yard Simulation Model

model classification / characteristics	container yard simulation model based on micro-analysis technology	the general container yard simulation model
basic research unit	single entity(such as: single container）	a kind of entity（collection of many containers with the same characteristics）
a description of entity behavior	of great detail and truth	not detailed
a description of single entity's individual characteristics	detailed records	not trace the behavior of each individual entity
single entity simulation result	can be calculated in detail	can not be calculated
description of the bottleneck link	able to find and reflect the dynamic changes of the bottleneck, the description of the bottleneck is very micro (such as a blocking phenomenon occurring in the yard)	able to find and reflect the dynamic changes, the description of the bottleneck is very macro (such as the yard area link)
an evaluation of emergency solutions	suitable	not suitable
suitable decision behavior	daily operation decision（such as: aiming at a certain ship's mechanical allocation analysis）	Long-term operation decision（such as: aiming at the whole yard mechanical allocation analysis）
simulation time	usually for several days	usually for one year
simulation step size	usually for minutes	usually for days

3. EEAP (Entity, Event, Activity& Process) of CYMSM

3.1. Container Yard Handling Technology System Analysis from Micro-level

From the perspective of micro-level, Container Yard Handling Technology (CYHT) treats the ship arrival as a service object, loading and unloading containers from a ship as the operating units. And production operation plan is made before ship berthing (for example, determining the number of operating lines, mechanical type and quantities, etc.). After the ship berths, operations plans will be put into practice and problems that they encountered in the operation process are dealt with.

CYMSM treats handling technology system as a research object from micro-level. After the container ship arrives, it then displays the whole process of each container from ship-to-shore (STS) handling, horizontal transport, to yard storage (as shown in figure 1), and keeps records of the main data and parameters involved in handling process of all the machinery, such as mechanical handling efficiency, handling volume, transportation time, waiting time, so as to calculate the final simulation results.

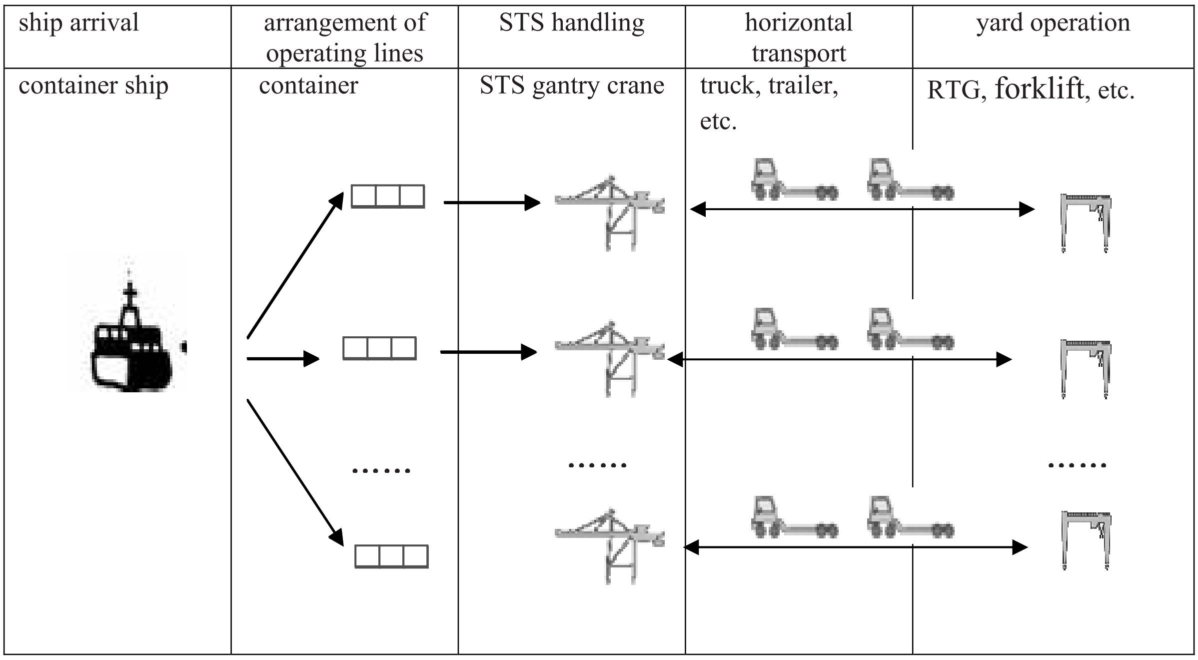


Figure 1: Container Yard Handling Process schematic diagram from micro-level

3.2. Entity and Process

Through the research of the real CYHT system from the micro-level, a complete CYMSM can be divided into six processes. The processes and the corresponding entities of the simulation model are shown in table 2.

Table 2: processes and entities of CYMSM

process	permanent entity	temporary entity
the operation sequence process of loaded (unloaded)	berth	container ship; container
the unloading process of operating lines		
the loading process of operating lines		
the operating process of STS handling	yard, the machines of the three links of STS handling, horizontal transport, and yard storage	container
the horizontal transport process		
the yard operation process		

3.3. Event, Activity and Process

Each process is made up of events and activities. Through the research of real system, the events and activities of that six processes can be summarized and construct. Take the most important process (the operating process of STS handling) as an example, events and activities are shown in Table 3.

Table3: events and activities of the operating process of STS handling

process	event/activity	event/activity description
the operating process of STS handling	event	transfer the STS machine
	activity	move the STS machine
	event	remove the hatch cover
	activity	Lift the hatch cover and put it at the appropriate locations in the dock apron area
	event	the STS machine wait for the horizontal machine
	activity	the STS machine wait for the horizontal machine's arrival
	event	unload one containers onto the horizontal machine
	activity	the STS machine unload one container onto the horizontal machine
	event	load one container onto the ship
	activity	the STS machine load one container onto the ship

We can see that as a result of the research methods of microscopic analysis, the CYMSM studies a number of events, which related to the container yard production systems. These events, according to the characteristics of production operations, give a detailed description to the variety of activities of operating lines and machines in the container yard production process. The actual operating mode of the yard is the same with these activities. Table 3 reflects that there are a large number of events. In addition, these events also have very complex logical relations. Whether many events occur or not needs to be judged, each event would constitute a subset of simulation system modules, and simulation system is driven by events.

4. The Research on the Operation and Simulation Rules of CYMSM

Container yard makes a number of different rules in the daily handling process. These rules, some are based on the security of port production, such as dangerous goods should be put in a dangerous area; others are based on the handling machinery properties , for example, RTG operation zone's layers should not exceed the number of required stack layers. In the general container yard simulation model, a description of these rules are very rough or not be considered, but in the micro-simulation model, because of the need for very real and accurate description of yard operation, therefore, it is necessary to do some research on the authenticity rules of container yard operation micro-systems, and on this basis determine the corresponding simulation mechanism.

To fully reflect the daily operating system of container yard, research must be done on the actual corresponding operation of each process, and summed up the operating rules of the actual system and the simulation rules of corresponding simulation system. Owing to space limitations, a STS gantry crane for single container (moving once, only loading or unloading one container) as an example to show the operating rules and simulation rules.

Summarizing the actual operation of container yard, unloading operation sequences of a STS gantry crane for single container are made up of two operating rules:

Rule 1: containers are unloaded simultaneously from different "operating lines". Different operating lines unload from sequence 1 of "operation sequence" which is determined by the dispatch department.

Rule 2: a container under some "operation sequence" is unloaded from land side to sea side and from top to bottom.

Simulation rules of operating sequence generated from the above operating rules are as follows:

Rule 1: a ship dispatch personnel arranges the number of unloading operating lines and the operating sequence under some operating line.

Rule 2: each operating line starts unloading at the same time.

Rule 3: the unloading sequence of some operating line is from land side to sea side and from top to bottom.

5. Simulation Results Evaluation Index

The running of the CYMSM can obtain a series of valuable output indexes, which can be used to give a technical and economic evaluation of the performance of container yard daily handling program. Such indexes mainly include:

5.1. Statistics of Mechanical handling Containers

The indexes are used for, in a simulation period, making statistics of the mechanical loading, unloading capacity. These indexes are acquired from the running of the CYMSM. Mechanical type can be further divided into the STS machinery (such as gantry crane), the horizontal machinery (such as container trucks) and yard operating machinery (such as RTG). The operating state can be divided into "loaded" or "unloaded". And container volume can be divided into unit (such as 20' or 40' container) and throughput (TEU).

For instance, the calculation method of loading throughput for some STS machine i is:

$$Q_{SLi} = 1*\sum q_{sl20i} + 2*\sum q_{sl40i} \quad (1)$$

Among them: $\sum q_{sl20i}$ indicates the quantity of 20' container loaded by STS machine i
$\sum q_{sl40i}$ indicates the quantity of 40' container loaded by STS machine i.

5.2. Mechanical Operation Efficiency Statistics

The indexes are used for making statistics of the number of loading and unloading containers per machinery per hour on the average. Among them, a dividing of machinery type, operating situation and container volume are the same with "5.1. Statistics of Mechanical handling Containers"

5.3. Machine Utilization Ratio Statistics

The indexes are used for giving statistics of proportion of working hours to the whole operation hours (i.e. the sum of working time and waiting time). Mechanical working time and waiting time in the simulation model will be respectively calculated by machine type, operating situation and container volumes.

5.4. Container Restacking Quantity and Time Statistics

In order to enhance the capacity of storage yard, containers are usually stacked for 3-7 layers. When the needed handling containers are not put on the top layer, containers restacking operation is inevitable. Although each time spending on one container restacking with respect to the whole time of handling is very short, the accumulative time is long and should not be neglected. In the running process of CYMSM, number and time of each restacking operation can be obtained, and these two indexes can help yard managers to better plan operation process and reduce the restacking number and time.

5.5. Blocking Quantity and Time Statistics

Blocking refers to that, when horizontal machines in the yard are moving on the roads, the former vehicle is handling containers blocks the latter's moving, causing the latter vehicle to wait. More often than not, blocking is not included in the scope of the China's business statistical indicators (Xuyuan, 1998) (Hong, 2003). However, as the yard operations of the container hub port are increasingly busy, the blocking phenomenon occurs frequently, which has large effect on the efficiency of container yard operations. As the entity's detailed description of the micro-simulation model, it can keep records of blocking quality and time of each block area. Through accumulation, we get index, which can provide yard personnel with operation plan adjustment and optimizing decision-making basis.

6. Study on the Production Plan Adjustment Mechanism of Micro-simulation Model

One of an important function of CYMSM is served as a container yard production "Decider". Before ship's arrival, simulation results of all kinds of plans can be analyzed in advance. We can observe every block link, adjust yard operation plan, and ultimately find an optimal operation plan.

6.1. Factors Triggering the Production Operation Plan Adjustment

There are two types of factors that can trigger production operation plan adjustment: Firstly, the simulation results of the operation plan are not satisfied; Secondly, unexpected events occur.

6.1.1. Unsatisfied with the Simulation Results

The following four kinds of situation may result in unsatisfied with operation plan, thus triggering the production adjustment mechanism:

- Operating time imbalance of loading (or unloading) ship lines

The imbalance of Ship (or ship) operating time will lead to the early ending time of certain operating line, while others are ended late. According to the principle of casks, the end of the ship's operating time is subject to the longest operating line, therefore, it is necessary to adjust the plan of operating lines to keep the operating time almost the same.

- The low efficiency of the STS machinery

The handling efficiency of STS machinery is an important indicator of competitiveness for port. High handling efficiency can decrease handling time, which can reduce transportation costs for carriers, and enhance berth through capacity for the port

- The duration of port stay is too long

If ship stays too long at port, the ship voyage time and cost will increase. Therefore, container yard attaches great importance to speed up the turnover of ship and other land vehicles.

- Too long blocking time at container yard

Container yard, in some cases, may generate block phenomenon. The occurrence of such a situation, on the one hand, may due to lack of container yard space and narrow roads, on the other hand, may result from improper operation plan, for example, operating lines interfere with each other, the import and export containers stacked at improper place and so on. For the latter cause, the operating plan should be adjusted. This CYMSM can test the operating plan in order to choose a desirable one.

6.1.2. Occurrence of Unexpected Events

The following two unexpected events also trigger the production adjustment mechanism:

- Failure of some operating machines
- Early completion of handling operation based on user request or because of some special reasons

6.2. Operation Plan Adjustment Methods

When the factors triggered operation plan adjustment occur, the container yard need to timely adjust the operation plan in order to meet the demand. As the simulation model will not disrupt the operation of real systems, and can carry out repeated experiments in a relatively short time, therefore, adjusting plans is an important function of simulation model. Take one of the above factors, "the duration of port stay is too long", as an example, adjustment methods of operation plan are shown in table 4.

Table 4: Operation Plan Adjustment Methods

factors triggering the operation plan adjustment	adjustment methods
the duration of port stay is too long	• increase operating lines; • change machine type; • change the equipment allocation of the three main operational link （STS loading/unloading, horizontal transport and yard storage）; • change the vehicle dispatching policy (for example, from a fixed policy to a flexible policy)

7. Conclusions

The research results of CYMSM can provide basis of the construction of the container yard operation simulation system. Based on this, container yard operation simulation system has been applied to container yard in Shanghai, and provides decision supports, thus testify the effectiveness of this method.

8. Acknowledgements

This research was supported by Innovation Program of Shanghai Municipal Education Commission (Grant No.06FZ033).

References

Bin, L.， Wenfeng, L., and Yu Z. (2008), Agent-based modeling and simulation for vehicle dispatching at container terminals, Journal of System Simulation, 20(19):5158-5161+5198.

Dahal, K., Galloway, S., Burt, G., Mcdonald, J., and Hopkins, I.(2003), A Port System Simulation Facility With An Optimization Capability, International Journal of Computational Intelligence and Applications, 3(4):395-410.

Degano, C. and Di F.A. (2001), Modelling automated material handling in intermodal terminals, International Conference on Advanced Intelligent Mechatronics, 2:1023-1028.

Hong Z.(2003), Port Management, China Communications Press: Beijing.

Jinglei, Y. and Yizhong, D. (2006), Research on Optimal Equipment Allocation in Container Terminals by Using Closed Queueing Networks, Chinese Journal of Management Science, 14(6):56-60.

Liu, C.I., Jula, H.and Ioannou, P.A. (2001), A simulation approach for performance evaluation of proposed automated container terminals, Conference on Intelligent Transportation Systems:563-568.

Sen, A.K., Bagchi A. and Sinha B.K. (1991), Admissible search for minimum penalty sequencing of jobs with setup times on two machines, Proceeding of IJCAI-91 (International Joint Conference Artificial Intelligent):178-183.

Ward, T. (1995), Simulation analysis for planning delta port container terminal, Proceedings of the Ports '95 Conference on Port Engineering and Development for the 21st Century: 640-651.

Xuyuan C.(1998), International Container Yard Operation Management, Dalian Maritime Press: Dalian.

Yong, M.T. (1993), PortSim: simulation and animation of container port, Proceedings of the 1993 Summer Computer Simulation Conference:143-150.

Studies of the concepts, features and development strategy of the fourth generation port based on supply chain thoughts

Jieshu Zhang[1], *Hong Zhen*[1] and *Jie Gao*[2]

[1]Shanghai International Shipping Institute, Shanghai Maritime University, Shanghai, China.
[2] Department of Management, Shanghai Maritime University, Shanghai, China.

Abstract

Different from the first, second and third generation port, the fourth generation port changes from emphasizing itself a "logistic center" to one of the links in the supply chain. Therefore, further studies about the fourth-generation port based on the supply chain management ideas are necessary. Based on the definition of the fourth generation port with the feature of a supply chain management, the evolving model of the relationship between port and supply chain is put forward, pointing out that relationship has experienced three stages: separation (independent of each other), integration (the formation of the port logistics service chain providing high value-added services for manufacturing) and embedding (the port embedded in the manufacturing supply chain). The characteristics of the fourth generation port should be different, lean and agile which are distinctive from the former three generation ports to the requirements of the supply chain management. Based on the above-mentioned analysis, the development strategies of the fourth generation port should be forwardly integrate and construct its own supply chain, and cooperate with the core enterprises of supply chain.

Keywords: the Fourth Generation Port, Supply Chain, Concepts, Features, evolving model

1. Introduction

With the continuing acceleration of economic globalization and regional economic integration, more and more production, operation activities and resources allocation start at a global scale. Ports transformed from the transport hub into the basic platform of international resources, which occupies an increasingly important strategic position in a national and regional economic development.

As a focus of economic development in a new era, supply chain management has developed into an advanced business management model, which has become an important strategy for enterprises participating in global market competition. "The competition in the twenty-first century is not the competition among enterprises, but the competition among supply chains," which has reach agreements. Supply chain management thinking has very important and far-reaching significance for the fundamental shift of business management method and competition mode.

Port, the largest international trade transfer station, is the largest and most important logistics node. In the supply chain management mode, ports have an increasingly important status and functions. According to the requirements of the supply chain development, the status and functions of modern ports in the global integrated transportation system are changing. The development of economic globalization and supply chain management thinking has made new demands on the port industry. International ports are entering into an important strategic transformation period (Hong and Guiyun, 2008). We believe that the fourth generation port based on the thinking of the supply chain will be an important strategic direction for the future port development.

2. Concept of the Fourth Generation Port Based on the Supply Chain Thoughts

In 1999, the United Nations Conference on Trade and Development (UNCTAD) proposed the concept of the fourth generation port (UNCTAD Secretariat, 1999). The concept of the fourth generation port is an organization of "physically separated but linked through common operators or a common administration”. The listed examples are the combination port of Los Angeles and Long Beach as well as another combination port of Copenhagen in Denmark and Malm in Sweden. And the conference pointed out that the ports development have entered into a new stage because of alliance and combination of ports of different parts.

From the definition, UNCTAD describes an important characteristic on the region of the fourth generation port, namely, the port is not at one place, but linked with each other through a common operator or a co-management organization, and finally integrated. In recent years, as the fourth generation port becomes an important link in a supply chain, it has played an important role in integration of the supply chain involved in ports, and this characteristic has become another major research focus at present.

In the 1990s, especially at the beginning of the 21st century, with the rise of supply chain management thinking, the port is regarded as a supply chain node. Its functions and competitive strategies are suffering from the ever-changing impact. From logistics to the supply chain transformation, it means that the realization of conversion from an enterprise to the cross-enterprise. Whether port operations are smooth and efficiency directly affect the entire supply chain. Competition among the modern ports has been transformed into the competition among supply chains involved ports.

With the difference of the first, second and third generation port, the fourth generation port is stressing itself from a "center" to a greater emphasis on a link of the supply chain. This change reflects its transformation from "self-centered" point of view to focus its own role on the supply chain (Yu, 2008). "Center" often causes the logistics to gather, and thus the retention of logistics at this time is regarded as a reasonable phenomenon. However, considering the port as a link in the supply chain, it stresses the fast flow of the real things and information.

Combined with theory and practice of modern supply chain management, considering that the port is facing an increasingly complex market demands, this paper gives the following definitions of the fourth generation port: the fourth generation port is based on the third generation port which provides integrated logistics services for the general supply chain enterprises, organizes, participates and plans international economic and trade activities actively, with seamless connection of every related link of supply chain to port, and ultimately forms port supply chain with functions of economic, trade, manufacturing and logistics. The port has become the "Dispatching Center" of the port supply chain, and port services upgrade emphasizes to meet with the different, agile and refined needs of the supply chain.

3. Evolving Model of the Relationship between Port and Supply Chain

The evolving model of relationship between port and supply chain is divided into the following three stages:

The first stage: separation (figure 1). At this stage, the port and the general manufacturing supply chain are mutually independent, and the two sides, mainly through agents (such as freight forwarder or shipping agency) make connections with the supply chain. Port provides services such as handling, carrying, and processing. Port at the stage is the first generation and the second generation port proposed by UNCTAD.

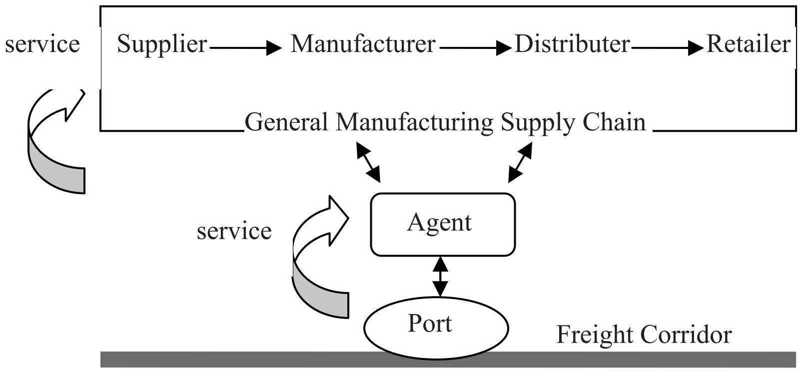


Figure 1: Port and Supply Chain's Relationship： Separation

The second stage: integration (figure 2). Ports not only start relationship with supply chain through agents, but also directly provide more and more services for the supply chain nodes enterprises. The port at this stage has become the largest and most important international trade node. As the core logistics enterprises, terminal operators integrate all related logistic services supplied by different logistics suppliers such as carriers, freight forwarders, and inland transportation companies, etc. and form port logistics service supply chain which provide highly value-added services for the general manufacturing industry supply chain. Functions of the port have been developed into the integration of logistics, trade, finance and other services. Port at this stage is the third generation port proposed by UNCTAD.

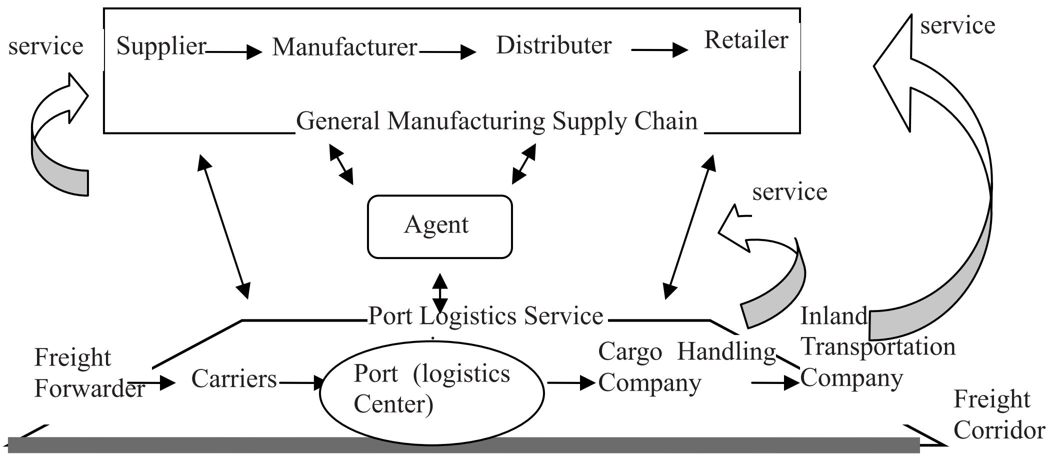


Figure 2 Port and Supply Chain's Relationship: Integration

Port logistics service supply chain, is called port logistics service chain for short, which is a kind of service supply chains. Different from general manufacturing supply chain, service supply chain does not have manufacturing sector, and it treats a service-oriented enterprise as the core enterprise which leads the supply chain.

The basic mode of port logistics service chain is: functional port logistics service providers → integrated port logistics service providers →manufacturers and retailers. Among them, the "functional port logistics service providers" is the supplier of "integrated logistics service providers", which refers to the traditional port logistics enterprises, such as carriers, freight forwarders, and inland transportation companies, and so on. Because they only provide services with a single, standard function, their business is often limited in a certain region, and they are absorbed by "integrated port logistics service providers" as suppliers in building a national or even global service network.

Port logistics service chain is a kind of new supply chain, which treats "integrated port logistics service providers" as the core enterprise, and its function is to provide a full range of logistics services for customers. The most obvious feature of "integrated port logistics service providers" is to choose appropriate "functional port logistics service providers ", combine their services to serve logistics buyers. "Integrated port logistics service providers " can be acted by many logistics operators. Liner companies and large-scale logistics operators transformed from port operators are strong competitors. When ports serve the general supply chain as the role of "integrated port logistics service provider", we think the port has already possessed the important characteristics of the third generation port.

The third stage: embedding (figure 3). Relationship between ports and the general manufacturing industry is more closely. Ports positively cooperate with important enterprises of supply chain, coordinate in business and integrate in function. We can think that the port is "embedded" in the general manufacturing industry supply chain, thus forming the special supply chain integrating a general supply chain and service supply chain, that is, the port supply chain. Port has become a control center and integrated service platform of this special supply chain, and its function has put more emphasis on the supply chain to meet with demands of service of differentiation, agility and refinement. Port at this stage is the fourth generation port proposed by UNCTAD.

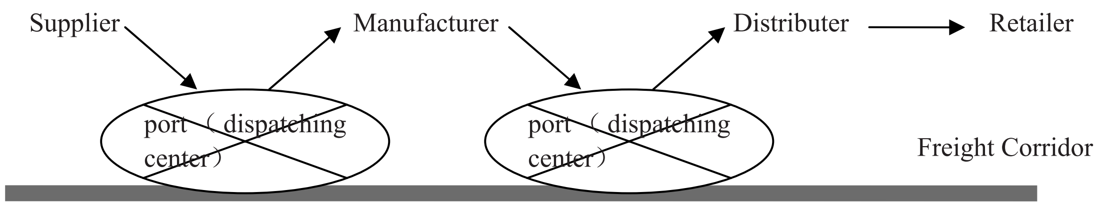


Figure 3 Relationship between Port and Supply chain: Embedding

The future competition among ports upgrades to the competition among port supply chains. Ports do not compete simply relying on geographic location and highly efficient operations, but actively participate in building port supply chains, so as to improve the overall efficiency of the supply chain, provide value-added customer services, raise the port position in the international industrial chain, and enhance its competitiveness.

4. Characteristics of the Fourth Generation Port Based on the Thinking of the Supply Chain

UNCTAD points out that characteristics of the fourth generation port include: it mainly deals with container; its development strategy is to make an alliance with shipping-related business and other ports; its production is characterized by integrated logistics; its key successful elements are decision-making, management, promotion, training and other soft factors.

This article holds that the fourth generation port, in addition to above characteristics, under the influence of supply chain management thinking, shows the following characteristics:

4.1. Become Information Control Center and Integrated Service Platform of the Port Supply Chain

The fourth generation port, through actively cooperates with important enterprises of the supply chain and collaboratively develops with other transportation providers, builds an integrated, seamless supply chain of the port, and has an in-depth participation in the port-related sub-chain in a general supply chain. The fourth generation port has become a planning and organizational control center and an integrated information service platform of sub-chains.

4.2. Emphasis is a Link of the Supply Chain

Different from the previous ports of the three generations, the fourth generation port stressed that it is transformed from a logistics center into a link of the supply chain. The nature of this change lies in the fact that the port changes from the static, node-based into a dynamic, network-type role. Based on improving the logistics center function, it emphasizes that materials and information flow quickly.

4.3. Port Network

The ports before and after the supply chain are not isolated but interrelated, thus more emphasis is placed on the coordination among these ports. At present, some yards running by port operators are forming network, which is conducive to coordination and interaction among ports.

4.4. To participate in the Cooperation and Division of the General Manufacturing Supply Chain

Besides its traditional handling, transport functions, mort importantly, the fourth generation port actively takes part in the cooperation and division of operation of the general manufacturing supply chain, in order to build an integrated, seamless supply chain network. Therefore, the port must take an initiative attitude in meeting with some production features of the manufacturing industry. Their flexible, lean, and agile operation requirements have brought new challenges to port services (Hong, 2005).

4.5. Differentiation

Different services proposed by the port supply chain put serious challenges to the port development, because till now characteristics of port production development are standard and scale. The standardization and scale of ports are necessary for port development, similar to the manufacturing industry. Port production services will ultimately move from standardization towards differentiation and personalization. And the requirement of the flexible operation mode in port was then put forward. These ports should have the abilities to meet with different customers' requirements, satisfying the fast changing market.

4.6. Lean Production

The lean production in port is a reflection of the quality of port operations. The lean production in port can enhance the operating efficiency of the supply chain. The lean production in port is realized by means of business process optimization, namely reducing the time of goods staying at port, and reducing or eliminating costs consumed by non-value-added activities, and minimizing switching time to the new business process.

4.7. Agility

Agile production in port is related to the market response ability. The port is required to be able to make rapid response to the port supply chain needs, satisfying these different, individual needs.

5. Development Strategy of the Fourth Generation Port

Taking all the definition and characteristics of the fourth generation port into account, we can see that the fourth generation port in the future port supply chain should not be limited to a supported role, but should seek for a higher position in the port supply chain in order to increase the competitiveness of the port. Therefore, it leads to the following effective development strategy of the port:

5.1. To Provide High-quality, Different services

In the global supply chain network, the form and boundary of enterprises will have a fundamental change. By integrating resources all over the world, supply chain enterprises can get lower cost, better technology and more efficient production and services. That results in cost reduction and competitiveness improvement for the whole supply chain. The key node enterprise of the supply chain chooses among the different ports according to the competitive advantage and benefits. Whether the port can be selected or not, whether it will participate in the supply chain or not, which will directly affect the competitiveness of the port (Khalid and Richard, 2004).

The former three generation port services have been characterized by standardization and mass-production. Standardization and mass-production bring a lot of convenience to port operation, reducing cost of ports. But at the same time it has added logistics links, extending the time of goods staying at port, deviating with the customer requirement of rapid response. Therefore, the port should provide high-quality, different services to make sure that it has the best and value-added effect on the supply chain, and raises its position in that chain.

5.2. Actively Integrate and Build Supply Chain

With the rapid economic development, the port cargo throughput has enormously increased, and this increases the importance of port in supply chain. As a supply-chain node, the port enterprise can make use of its unique geographical advantages, strengthening collaboration among supply chain node enterprises, promoting and improving port service performance. And ultimately it enhances the level of customer service and satisfaction, creates and enhances the core competitiveness of the port, thereby building and strengthening the central position in the supply chain port.

5.3. Establishing Network of the Port Focusing on the Demand of Supply Chain

Similar to the general supply chain, the port supply chain is also a kind of network structure. Ports in the supply chain should not be isolated, but interrelated. At present, the world port in the geographical distribution has begun to show the characteristics of network development. The port network has been or is being formed, whose major nodes are global or regional international shipping centers, and subsidiary nodes are regional hub ports and branch ports. In our view, the driving force formed by the port network, not only has the reason of ports cooperation, mutual benefit, and avoiding vicious competition. And, more importantly, because the manufacturing supply chain changes the demand for the port, an objective requirement of a seamless connection between the port and supply chain is put forward, and many port enterprises in one supply chain must realize coordination and interaction. Therefore, the cooperation between the future ports form more and more networks. In this port network, the core and dominant position will be the fourth generation port. Whether other ports can be selected and incorporated into the port network depends on whether it can operate with supply chain and service system.

5.4. Strengthening Coordination and Cooperation among Upstream and Downstream Enterprises of the Supply Chain

In a highly competitive environment and with rapid development of global economy, corporate restructuring in the port supply chain is being driven by the logistics service providers, and they not only consider reducing costs, improving efficiency as the goal, but also gain a competitive advantage in the process of transferring value to customers (Robinson, 2002). The fourth generation port should be the drivers of this reorganization, and the ultimate powerful source of the port supply chain, which drives the whole supply chain. It is necessary for the fourth generation port to consider from the overall development perspective, establishing long-term partnership with the upstream and downstream enterprises of the port supply chain, sharing risk and benefit with them. The fourth generation port should actively cooperate with the other enterprises of supply chain nodes, such as suppliers, the third-party logistics companies of port-related enterprises, shipping companies and enterprises of all types, sharing profit with them, rather than passively reach an agreement with shipping companies and establish a special yard for them. Only in this way we can enable the supply chain management of port enterprise develop effective extension and upgrade the entire supply chain port competitiveness.

6. Conclusions

The Theory and Practice of supply chain management deeply influence the development direction of modern ports. It is proposed that, the fourth generation port is not merely a logistics center, and it actively integrates the general supply chain and service supply chain, in order to form a port supply chain which is dominated by port itself. Playing a central role in the port supply chain is the strategic development trend for the fourth generation port. Adapting to that trend, characteristics such as flexibility, agility and lean production, which originally developed in manufacturing industry, have emerged in the fourth generation port. Therefore, improving the supply chain service level and strengthening the cooperation with other enterprises in the supply chain, would be an efficient development strategy for the fourth generation port.

7. Acknowledgements

This research was supported by the Applied Fundamental Research Program of Ministry of Transport of the People's Republic of China (Grant No.2007329810060) and Science & Technology Program of Shanghai Maritime University (Grant No. 2008434).

References

Hong, Z. (2005), The concept of the fourth generation port and its implementation, Journal of Traffic and Transportation Engineering,5(4): 90–95.

Hong, Z. (2005), Flexible ports: the characteristics of the fourth generation port, Journal of Chinese port, (8):27–29.

Hong, Z. and Guiyun, L. (2008), The developmental pattern of flexible ports, Shanghai Jiao Tong University Press.

Khalid, B. and Richard, G. (2004), A logistics and supply chain management approach to port performance measurement, Maritime Policy & Management, 31(1):47–67.

Robinson, R. (2002), Ports as elements in value-driven chain systems: the new paradigm, Maritime Policy & Management, 29(3): 241–255.

UNCTAD Secretariat (1999), Technical note: Fourth-generation port, Ports Newsletter, (11):9-10.

Yu, R. (2008), Building the fourth generation port based on the supply chain, China investment, (7):71.