ESSAYS ON TRANSPORTATION ECONOMICS AND POLICY

by

Yulai Wan

B.B.A., City University of Hong Kong, 2005

M.Phil., The Chinese University of Hong Kong, 2007

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Business Administration)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2013

Abstract

This dissertation focuses on two major topics in transportation. The first one is related to hinterland accessibility of seaports. We analytically examine the interaction between urban road congestion and competition of two seaports on a common hinterland. An increase in road capacity by a chain will improve its port's profit while reducing the rival port's profit. The impact of levying road tolls and the optimal road pricing rules are also discussed. Given the above theoretical predictions, we empirically investigate the impacts of hinterland access conditions, especially urban road congestion, on the competition between as well as efficiency of major container ports in the United States. We find that more delays on urban roads may cause shippers to switch to competing rival ports. Adding local roads tends to benefit the port and harm its rival (in terms of throughput) by reducing road congestion. In general, there is a negative association between road congestion around the port and port productivity. However, this relationship tends to be negligible for primary ports of entry which enjoy substantially larger container throughput volume. We further investigate the strategic investment decisions of local governments on inland transportation infrastructure in the context of seaport competition. The two seaports and the common hinterland belong to three independent local governments, each determining the level of investment for its own inland transportation system. We examine the non-cooperative optimal investment decisions made by local governments, as well as the equilibrium investment levels under various coalitions of local governments.

The second topic is related to airport congestion pricing. We study airport pricing with aeronautical and non-aeronautical services, incorporating the connection between congestion delay and consumption of non-aeronautical services (an aspect that has been neglected by previous studies), and the effect of passenger types. The resulting pricing rule includes two extra terms missing in the literature. Then, we model terminal congestion and runway congestion separately to accommodate their respective characteristics and identify a number of aspects in the optimal pricing rule which differ from those in the literature.

Preface

A version of chapter 2 has been published. Wan, Y. and Zhang, A. (2013) Urban Road Congestion and Seaport Competition, *Journal of Transport Economics and Policy*, 47(1), 55-70. Anming Zhang came up with the problem and some initial modeling analysis. I made major extensions to the modeling and conducted all the mathematical proofs and wrote the first draft of the manuscript. Anming Zhang made some major revisions to the manuscript.

Chapter 3 has been split into two papers. One has been published: Wan, Y., Zhang, A. and Yuen, A. (2013) Urban road congestion, capacity expansion and port competition: empirical analysis of U.S. container ports, *Maritime Policy and Management*, 40(5), 417-438. The other one has been accepted for publication: Wan, Y., Yuen, A. and Zhang, A. (2013) Effects of hinterland accessibility on U.S. container port efficiency, *International Journal of Shipping and Transport Logistics*. I conducted most of the work, including data collection, econometric modeling, data and empirical results analysis and I also wrote the first drafts of the manuscripts. Anming Zhang helped with revising the manuscripts of both papers. Andrew Yuen helped with revising the first paper and assisted in DEA score calculation and collecting container port input variable data for the second paper.

Chapter 4 is based on a model built by Anming Zhang and Leonardo Basso. Anming Zhang and Leonardo Basso did preliminary analysis for the pricing decision and non-cooperative investment decision in the case of welfare-maximizing ports. I conducted the analysis in strategic investment decisions with various forms of coalitions, developed the model further for profit-maximizing ports, conducted corresponding analysis for pricing as well as non-cooperative and cooperative investment decisions, and compared those results with the case of welfare-maximizing ports. I also wrote up the whole draft of the manuscript.

A version of chapter 5 has been published. D'Alfonso, T., Jiang, C. and Wan, Y. (2013) Airport pricing, concession revenues and passenger types, *Journal of Transport Economics and Policy*, 47(1), 71-89. I came up with the key component of the mathematical model which is the major innovation of this chapter. Tiziana D'Alfonso made substantial contribution on literature review. The rest of the work, such as modeling, mathematical proof, and writing, was conducted by the three co-authors together.

Chapter 6 is original, unpublished, independent work by the author, Yulai Wan.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to express my sincere appreciation to Prof Anming Zhang, my thesis supervisor, for patient guidance and various supports, morally and financially, throughout my entire doctoral study. It is Prof Zhang's enlightenment and encouragement that helped me get through difficulties at every stage of my program.

Special thanks are owed to Prof David Gillen, Prof Ralph Winter and Prof Tae Oum, for spending their valuable time on serving as my thesis committee members.

I thank Changmin Jiang and Tiziana D'Alfonso who assisted my work and provided advice on some of the chapters, as well as Elaine Cho who has provided a wide range of timely and detailed administrative supports.

My enduring gratitude also goes to my parents who always support me unconditionally.

Dedication

To my parents, Shuwen Wan and Yuan Xu.

1 Introduction

This dissertation includes two major research topics in transportation economics: (1) hinterland accessibility of seaports and its implications on port competition; and (2) airport congestion pricing with the incorporation of the connection between congestion and concession activities. Both issues are related to congestible transportation facilities and measures to mitigate congestions. Basically, we are interested in the impacts of congestion pricing and capacity expansion on various parties in the transportation system, the structure of optimal pricing rules and investment rules for the transportation infrastructure and the corresponding policy implications. In terms of the model structure, the first topic concentrates on the competition between two congestible facilities while the second topic mainly investigates stand-alone facility with a focus on the strategic behavior of various users of the facility.

1.1 Container port competition and hinterland access conditions

Chapters 2-4 relate to the first topic, i.e. interaction between port competition and hinterland accessibility. Since containers became widely accepted for ocean shipping, the cost of shipping has been substantially reduced, which has become one of the driving forces for globalization and assisted the adoption of just-in-time production strategy. As a result, a large proportion of marine cargoes are now originated from or destined to inland regions farther away from the seaports. Competition between gateway ports has been intensified due to increased overlapping of hinterland. As inland transportation and logistics account for a significant share of container shipping costs, seaport competition is no longer just between individual ports but between alternate intermodal chains and the connectivity between ports and hinterland has been a major influential factor of seaport competition (e.g. Fleming and Baird, 1999; Kreukels and Wever, 1998; Heaver, 2006; Notteboom, 1997).

Being an essential part of the intermodal chain which connects ports and the hinterland, urban roads have been identified as the new bottleneck of container shipping since the improvement in landside accessibility seems to lag behind the growth in container shipping. According to a report written by Texas Transportation Institute (2010), road congestion costs trucking firms USD $216 million a year in an average large urban area and USD $1,273 million in a typical very large urban area, while this amount rises to more than USD $3,000 million in major port regions such as Los Angeles-Long Beach area and New York-New Jersey area. To compete with other port cities, over the past few decades, governments of those port cities have made significant investment to improve hinterland access conditions and relieve local road congestion. As road congestion has been acknowledged as one major constraint for port development in the literature (e.g. Maloni and Jackson, 2005), formal theoretical studies on the interaction between roads and ports are emerging but still limited (see chapter 2 for detailed discussion on the related literature), while we fail to locate any empirical study. Early theoretical works do not provide a complete picture: Yuen *et al.* (2008) ignores the competition between ports; Zhang (2008) does not take into account the interaction between different types of road users; and De Borger *et al.* (2008) focuses on road capacity expansion and has little to say about road tolls. Thus, it remains a question whether various measures to mitigate road congestion can effectively improve a port's competitiveness and helps the port win more business over its rivals. From the policy makers' point of view, as an intermodal transportation chain or system involves a variety of players who may interact with one another or even impose externality on others, it is also necessary to evaluate the impact of those measures on the well-being of individual stakeholders and design for policies which appropriately internalize those externalities.

To study the aforementioned issues, we begin with an analytical model in chapter 2 which assumes duopoly ports compete for a common hinterland. Each port is linked to the hinterland by a local road shared by local commuters and trucks moving cargos between the port and the hinterland. Based on this model, a number of research questions are addressed. First of all, will policy interventions favor the local port over the rival port while mitigating road congestion? Which players in the port-road transportation system will be better-off or worse-off? In terms of policy interventions, we examine road capacity expansion and two types of road toll systems: discriminative congestion tolls which price trucks and local commuters independently and fixed-ratio toll schedule which keeps the ratio between commuter toll and truck toll to be constant. Further, we compare our answers to the above questions for ports with different objective functions, i.e. maximizing profit versus maximizing local welfare, as well as different modes of competition, i.e. quantity competition versus price competition. We also derive the optimal road capacity investment rule and optimal road tolls from local governments' point of view and compare them with those in the literature. The amount of congestion externality internalized by the optimal tolls is examined and the underlying reasons of not pricing at the marginal external cost are discussed.

Chapter 3 further empirically explores the relationship between road congestion and port competition based on theoretical predictions obtained in chapter 2. There are a few empirical evidences suggesting that hinterland connectivity does play a crucial role in shippers' port selection process and ports' competitiveness (e.g. Fan *et al.* 2012; Slack, 1985; Lirn *et al.*, 2004; Ugbonma *et al.*, 2006; Yuen *et al.*, 2012), but studies focusing on road congestion are nonexistent. Noting that road congestion toll has seldom been implemented in the United State where our sample is based on, chapter 3 studies the impacts of road congestion delay and urban road supply on the throughput of major container ports and their respective rivals. In addition, as productivity is another major port performance indicator, we discuss the potential influence of urban road congestion on port productivity from both the output and input sides and test whether road congestion adversely affects port productivity scores calculated by the data envelopment analysis. Other potential indicators of hinterland connectivity such as provision of on-dock rail facility and Class I rail services are also included in the regression models as explanatory variables.

Chapter 2 only includes the land side infrastructure of the port regions into consideration while ignoring the potential interaction between hinterland and the ports as well as the shippers in the captive catchment areas around ports. In fact, inland shippers' well-being is affected not only by the transportation cost incurred in port area but the transportation infrastructure in the hinterland; meanwhile, the competition intensity among rival ports could be affected by the transportation cost in the common hinterland. Such interplay between the hinterland and the competing ports may have strong influence on the local governments' infrastructure investment decisions. For example, in chapter 2 we assume local governments of the port regions make independent decisions on road capacity and consequently each government has incentive to overinvest. However, coordination among governments in different regions, especially between port cities and hinterland, is not uncommon. The Hong Kong-Zhuhai-Macau Bridge is an example directly relevant to the port industry. Being under construction since late 2009 after almost 20 years of studies and debates, the bridge project relies on the collaborative efforts of three local governments – Hong Kong, Guangdong and Macau. Although the bridge helps port of Hong Kong to extend its hinterland into the west part of Guangdong province and is supposed to have major contribution on the development of the whole west Pearl River Delta, it may substantially change the relationship between port of Hong Kong and port of Zhuhai, another major port in South China.

Thus, it is of great importance at both policy level and academic interest to have a systematic understanding on the implication of inter-governmental collaboration on transport infrastructure investment decisions. Chapter 4 is therefore motivated to study the strategic investment decisions of local governments on local as well as inland transportation infrastructure in the context of seaport competition and various inter-regional coalitions.

It addresses three questions: (1) how do inland infrastructure investment decisions affect port competition and regional welfare? (2) How do optimal investment decisions look like under various forms of coordination (coalitions) among local governments? (3) How do port ownership structures play a role in answering the above questions? In particular, we modify the linear city model and consider two seaports with their respective captive catchment areas and a common hinterland for which the seaports compete. The two seaports and the common hinterland belong to three independent local governments, each determining the level of investment for its own land transportation system. Two representative port ownership forms are compared: public ports which maximize regional welfare and private ports which maximize profit. One important feature which is omitted in chapter 2 for analytical tractability but included in chapter 4 is the inclusion of local captive catchment area for each competing seaport, which brings about interesting results when comparing ports under different ownership.

1.2 Airport pricing with connection between congestion and concession activities

Air travel delays have become a serious problem as the air transport market has experienced continuous growth during the first decade of the 21st century. In the United States, the largest air travel market in the world, about 20% of flights were delayed. The world's second largest air travel market, China has experienced even more delays in recent years: over 30% of its domestic flights were delayed. The US Department of Transport has identified congestion reduction as one of the top management challenges. One major cause or catalyst of air travel delays is airports congestion during the peak hours. Scholars have advocated the use of price mechanism to resolve airport congestion. Since 2008 US airports have been allowed to charge peak-period landing fees in addition to weight-based fees. Airport congestion and pricing has increasingly received attention from academia since the recognition of airlines market as in the form of imperfect competition and the heated debate on congestion internalization by airlines with market power (for a comprehensive review, see Zhang and Czerny, 2012).

In early airport pricing literature, airport congestion is modeled in the same way as for road congestion by assuming that airlines are atomistic and hence each airline operates only one flight. Consequently, an individual airline ignores the marginal external congestion cost (MEC) its flight imposes on other flights and the socially optimal airport charge should be set at MEC to internalize the congestion externality. Later studies however recognize that many airlines operate a large number of flights in an airport and thus do have market power. Those airlines will internalize the self-imposed congestion externality and raise their airfare by the amount equal to their own share of contribution to congestion. This self-internalization view suggests that the socially optimal airport charge on a certain flight should equal to the share of MEC not internalized by the airline plus a downward correction on airlines market power (Brueckner, 2002; Zhang and Zhang, 2006; Basso and Zhang, 2007).

Another feature of recent airport research is the inclusion of airport concession[1] activities. As

[1] The word 'concession' refers to to all non-aeronautical activities. As a significant part of non-aeronautical

many airports have been commercialized or privatized and required to be financially self-sufficient, concession revenues have become a major source of income for airports. Concession or non-aeronautical activities now contribute 45-80% of revenue for most major airports. The complementarily between aeronautical services and concession services predicts a negative component on the optimal aeronautical charge of an airport (Starkie 2002, 2008; Zhang and Zhang, 2003, 2010; Oum *et al*., 2004; Yang and Zhang, 2011).

Despite the relatively large literature on airport pricing, one missing piece is the fact that passengers' airport dwell time, which may be affected by airport congestion, can affect concession revenue. Empirical evidence suggests that the expenditure on concession goods increase as passengers' waiting time in the airport increases (Rowley and Slack, 1999; Geuens *et al.*, 2004; Torres *et al.*, 2005; Entwistle, 2007; Castillo-Manzana, 2010). On the other hand, according to Graham (2008), the purpose of the trip influences expenditure in the commercial area with leisure travelers spending, on average, more than business travelers. Therefore, chapters 5 and 6 study optimal airport pricing by incorporating the above empirical findings and linking airport congestion and concession activities together with different passenger types.

Chapter 5 employs a conventional approach in modeling airport congestion delay when runway congestion is the focus: that is, congestion delay is considered as a function of total traffic volume. The link between delay and concession activity is based on a positive relationship between delay and consumption of concession goods. Following Czerny and Zhang (2010) we model passenger heterogeneity based on passengers' differentiated values of time.

Being aware that both terminals and runways can be congested, we model terminal and runway congestions separately in chapter 6. This is motivated by the observation that passengers are most likely atomistic when determining their airport arrival time and the

---

activities is concession and this term is widely used in the related literature, we keep using 'concession' to refer to non-aeronautical activities in this dissertation.

length of dwell time in the terminal; besides, the arrival patterns may differ among passenger types. However, airlines, as mentioned earlier, have market power when deciding the number of flights to operate which is the driving force for runway congestion. In addition, as passengers cannot purchase when being on board and waiting for the flights to take-off, in chapter 6, we assume that concession consumption is irrelevant to runway congestion but directly linked to passengers' dwell time and the relationship between traffic volume and dwell time is elicited by the bottleneck model (Arnott *et al.*, 1990; 1993; 1994).

2 Urban Road Congestion and Seaport Competition: An Analytical Model[2]

2.1 Introduction

Since the 1950s, containerization has dramatically increased competition among seaports (e.g. Cullinane and Song, 2006; Luo and Grigalunas, 2003). As a node in the global supply 'chain' (Heaver, 2002), a gateway port connects its hinterland – both the local and interior (inland) regions – to the rest of the world by an *intermodal transport* network. Consider the cargo flow to the hinterland (the reverse flow can be similarly analyzed). Goods from the rest of the world (imports) are first shipped to a seaport, and then are transported to the inland by truck, rail, inland waterway, or a combination of these modes. The intermodal movement of freight by containers has reduced port and other intermodal handling costs and enlarged the reach of markets served by a given seaport. A hinterland that used to be served exclusively by a certain seaport may now be reached through another seaport. As argued by van Klink and van den Berg (1998), gateway ports are in a unique position to, on the one hand, stimulate intermodal transport and, on the other hand, use the intermodal systems to enlarge their hinterlands. In the commercially famous Le Havre-Hamburg port range, for instance, major seaports vigorously vie with one another for interior hinterland shipments that have alternative intermodal routing possibilities.

Containerization has also stimulated the demand for sea shipping (Heaver, 2002; Levinson, 2006; Notteboom, 2006b). With containers it becomes cheaper, faster and more reliable to move cargos around the world, resulting in enormous growth in global sourcing and sea-bound trade. The increased demand has been stressing seaports and their inland transportation systems. Tremendous efforts have thus been extended to the resolution of congestion at seaports at both the policy and research levels (e.g. De Borger *et al.*, 2008; Heaver, 2006; Yuen *et al.*, 2008; Zhang, 2008). Since the 1970s, the bottleneck of port

[2] A version of chapter 2 has been published. Wan, Y. and Zhang, A. (2013) Urban Road Congestion and Seaport Competition, *Journal of Transport Economics and Policy*, 47(1), 55-70.

congestion has shifted from the ship/port interface (e.g. terminal/berth investment, crane and yard productivity) to the port/inland interface (e.g. port access/egress, road transportation, hinterland connections) (Heaver, 2006). A survey conducted by Maloni and Jackson (2005) suggests that U.S. port managers' greatest concern in port capacity expansion planning is the capacity constraint imposed by local roads. More generally, inland logistics which accounts for 40-80% of total container shipping costs has become the primary source of potential savings for shipping lines (Czerny, 2007; Notteboom, 2004), and inland access in particular has been considered one of the most influential factors of seaport competition (e.g. Fleming and Baird, 1999; Kreukels and Wever, 1998; Heaver, 2006; Notteboom, 1997).

In this paper, we link port competition with road congestion on the hinterland. On the one hand, congested roads, being an essential part of the intermodal chain, may inhibit port throughput growth and reduce port competitive strength, as it is the chains rather than individual ports that compete (Suykens and Van De Voorde, 1998; Robinson, 2002). Options such as road capacity expansion and congestion tolls have been actively discussed in both the academic and policy circles. A few papers consider the competition between two parallel roads (e.g. De Borger *et al.*, 2005; De Borger and Van Dender, 2006) or two serial roads (e.g. De Borger *et al.*, 2007; Ubbels and Verhoef, 2008), but none of them explicitly incorporates port competition. On the other hand, port-related freight traffic can contribute to urban road congestion. Berechman (2007) finds that the additional road traffic due to a (modest) 6.4% container throughput increase at the Port of New York would induce annual 'social costs' ranging from $0.66 billion to $1.62 billion, over 60% of which is from road congestion costs (the time-loss due to traffic conditions and drivers' discomfort). In Vancouver, BC, truck traffic generated primarily by the port-related activities is becoming a conspicuous contributor to road congestion (Lindsey, 2007, 2008).[3] Taking the interaction between road and port traffic into account, will the policy interventions (road capacity expansion and congestion tolls) favor the local port while mitigating road congestion? Will the policies

[3] In the greater Vancouver area, truck traffic is anticipated to increase by 50% between now and 2021. In the United States, from 1993 to 2001 truck traffic on urban highways increased more than twice as much as passenger traffic, implying that freight traffic was contributing to worsening congestion at a faster rate than passenger traffic (US GAO, 2003).

affect a profit-maximizing port and a local welfare-maximizing port[4] differently? We investigate these (and other) questions in the present paper.

More specifically, we develop an analytical model in which we treat port competition as part of the rivalry between alternative intermodal transportation chains. There are two competing chains, each consisting of a seaport and its urban roads which connect the port to the (common) inland market. The urban roads are shared by both port-related trucks and local commuting cars and are congestible. Changes in road capacity or road tolls will influence congestion on urban roads, which in turn will affect ports' outputs and profits.[5] The ports maximize their profits and engage in Cournot (quantity) competition, taking road capacity and road tolls as given. Both the discriminative tolls, in which port-related trucks and commuter cars are charged separately, and the fixed-ratio toll schedule, in which trucks and cars are charged according to a fixed ratio are examined. We find that, opposite to the case of price competition, an increase in road capacity by a chain will improve its port's profit while reducing the rival port's profit, and road congestion of both chains will fall as a result. Increasing fixed-ratio tolls by a chain may increase its port's profit and reduce the rival port's profit. As a consequence, roads are tolled above the marginal external congestion costs, provided that the value-of-time of shippers is sufficiently large relative to that of commuters. When a discriminative toll system is implemented, however, commuters are tolled at the marginal external congestion costs while truck tolls are much lower. Furthermore, the case of ports' maximizing local welfare is examined, and the results under our assumption of ports competing in quantities are compared with those of price competition.

There is a limited literature on the interaction between port competition and road congestion. Our work is most closely related to a comprehensive paper by De Borger *et al.* (2008) who investigate a two-stage game in which local governments decide on the port and hinterland

[4] The case of local welfare-maximizing ports can be considered as having local public control over port's decision variables.

[5] The interaction among other major players across the chain (shipping companies, port authorities, terminal operators and hinterland operators) is abstracted away from the present analysis. As discussed in Van De Voorde and Vanelslander (2009), coordination and competition among these parties have manifested in forms of great variety and complexity and the issue may deserve further research in the future.

capacities in the first stage (both the port and inland are congestible) and in the second stage, the duopoly ports engage in price competition. The hinterland road tolls are taken as given in their analysis. Our analysis abstracts away port congestion and related port capacity decision, whilst both the hinterland capacity and road tolls are considered as the first-stage decision variables. This distinctive formulation allows us to link port competition not only with road capacity, but also with road pricing. Thus, unlike De Borger *et al.* who assume road tolls are exogenously given and are identical for both road users, trucks and commuter cars, we study in more detail the role of tolls and look at both the fixed-ratio and the discriminative tolls.[6] Furthermore, we assume ports compete in quantities. We will, in Section 2.4, argue that quantity competition may be more realistic than price competition in the case of ports, and will further compare our results with De Borger *et al*.'s results under the assumption of price competition.[7]

This paper is also related to Zhang (2008) and Yuen *et al.* (2008). Assuming a single intermodal chain, Yuen *et al.* investigates the effects of congestion pricing implemented at a gateway port on its hinterland's optimal road pricing, road congestion and social welfare. Zhang focuses on the corridor congestion and capacity investment rather than urban road congestion *per se*. In addition, a number of new features of the present paper were excluded in Zhang (2008), including differential shipper vs. commuter values-of-time, welfare-maximizing ports, explicit comparison between quantity competition and price competition, both the fixed-ratio and discriminative road tolls. Therefore, the present paper offers a more complete analysis on the interaction between road congestion and port competition. To a lesser extent, our work is related to De Borger and De Bruyne (2011) who examine the impact of *vertical integration* between terminal operators and trucking firms on optimal road

[6] The uniform toll system in which trucks and cars are charged identically is equivalent to the fixed-ratio toll schedule when the ratio equals one. Therefore, the present paper is comparable to some works on uniform tolls in the literature, for example, De Borger et al. (2005, 2007, 2008) and Ubbels and Verhoef (2008). The 'uniform pricing' mentioned by Yuen et al. (2008) is exactly the fixed-ratio toll schedule in this paper. Throughout the paper, we use the term 'fixed-ratio toll' when referring to those earlier studies on uniform tolls.

[7] There are several other differences between the two papers. For instance, we are also interested in which parties will be better off (worse off) and whether road congestion in each region will be alleviated if certain policy intervention is imposed, which are not extensively discussed in De Borger *et al.* but nevertheless are major concerns in many port cities.

toll and port charge.[8] The major differences with the present study are: first, De Borger and De Bruyne do not consider the competition between ports; second, they do not consider road capacity investment; and third, they allow market power among trucking firms, but to leave issues of double marginalization and double internalization of congestion out of the picture we assume perfect competition in this sector. If ports are monopolists, our model will produce optimal road toll and port charge consistent to De Borger and De Bruyne's work. Finally, the discriminative/fixed-ratio tolls have been examined in De Borger *et al.* (2005, 2007) and Ubbels and Verhoef (2008) who consider competition between two parallel roads or two serial roads. Each region imposes road tolls on commuter cars and 'transit' truck traffic. Port competition is not examined in these papers and so truck traffic is not port-related. Furthermore, transit truck traffic is generated from jurisdictions outside of the regions under consideration. We will provide a comparison between our results and these studies' in Section 4.[9]

The paper is organized as follows. Section 2.2 sets up the basic model. Section 2.3 examines how road capacity expansion and road tolls affect port rivalry when ports maximize their profits. Section 2.4 further discusses the impacts on rivalry between social welfare-maximizing ports and compares our findings with those in the literature. Finally, Section 2.5 contains concluding remarks.

2.2 Basic model

We consider an intermodal network that is likely the simplest structure in which our questions can be addressed. There are two seaports, labeled 1 and 2, competing for traffic

[8] The "vertical integration" issue is also at the center of two recent analytical papers by Czerny (2007) and Cantos-Sanchez *et al.* (2010). Czerny (2007) modeled port competition, which is indirectly through the competition of vertically-integrated firms that provide both inland and seaside services. Since he focused on transportation firms and their investment in port terminals, strategic rivalry between ports themselves was abstracted away by simply assuming constant (and exogenously given) port charges, nor was inland road congestion analyzed. Cantos-Sanchez *et al.* investigated the question: Once a shipping line invests in a container terminal, should the terminal be dedicated to itself or shared with the rival shipping lines for an access fee?

[9] For useful overviews of the general literature on urban road pricing and capacity investment, see Small and Verhoef (2007) and Ubbels and Verhoef (2008).

into and out of a common inland market. Each port is connected to the inland through roads in the urban area near the port. By selecting a port, shippers choose between two intermodal chains, with each offering a package of port operation and urban road transport (assuming no difference between transportation conditions further into the hinterland). A central point of the modeling is that port competition is viewed as competition between alternate intermodal networks in which ports form an important component. Consequently, when choosing the gateway port for their cargo, shippers and consignees (sometimes represented by forwarders and third-party logistics providers) will apply the generalized cost or ‘full price’ approach in the sense that all members of the transport supply chain, including the port, would contribute to the ‘cost’ of cargo shipments. Basically, other things being equal, the intermodal chain that imposes the lowest generalized cost and its corresponding gateway port will win a customer’s business. Another important critieria for shippers’ port choice decision is the vessal sailing schedule. In particular, shippers prefer keeping the number of ports called between the ports of origin and departure as few as possible so that time spending travelling between ports during a sequence of port calls can be kepted at minimum. Thus, outbound cargoes tend to select the port which is located at the end of the port rotation list before heading to the destination port while inbound cargo tend to select the first port of call right after departing the origin port. As port choice based on sailing schedule involves the choice of shipping lines which determine the port rotations while interacting with shippers, adding this feature into our current model will not only complicate the analysis but distract our focus away from the port-hinterland interaction. Thus, sailing schedule is not explicitly modeled in this paper, but this feature can be partly captured by the level of differentiation between competing ports which is built into the demand function.

Specifically, the demand function for port i depends on the full prices of cargo movement, ρ_i and ρ_j ($j \neq i$):

$$X_1 = d_1(\rho_1, \rho_2), \qquad X_2 = d_2(\rho_1, \rho_2). \tag{2.1}$$

Since in each chain the shipments flow through the port and its urban roads and these congestible roads are shared by both port-related traffic and local commuter traffic, the shippers' full price per cargo unit (e.g. twenty-foot equivalent unit) includes port charge p_i, road toll T_i, and delay cost of road congestion D_i:[10]

$$\rho_i \equiv p_i + D_i(V_i, K_i) + T_i, \qquad i = 1, 2, \tag{2.2}$$

where D_i depends on total road traffic volume (per unit of time) V_i and road capacity K_i.[11] This road congestion cost satisfies:

$$\frac{\partial D_i}{\partial V_i} > 0, \quad \frac{\partial D_i}{\partial K_i} < 0, \quad \frac{\partial^2 D_i}{\partial V_i \partial K_i} \leq 0, \qquad i = 1, 2. \tag{2.3}$$

Increasing traffic volume will increase road congestion while adding capacity will reduce road congestion. Conditions (2.3) are quite general and hold for several widely used delay functions, including a 'linear' delay function in that D_i is a linear function of the volume-capacity ratio (e.g. De Borger and Van Dender, 2006; De Borger *et al.*, 2005, 2007, 2008).

Since the road is shared by cargo shipments X_i and local commuters, we have $V_i = X_i + Y_i$, with Y_i denoting commuter traffic volume.[12] Equilibrium in the commuter traffic market implies equalization of the (inverse) demand for local commuter traffic, $\rho_{Li}(Y_i)$, and the generalized cost:

$$\rho_{Li}(Y_i) = g_{Li} \equiv \alpha D_i(X_i + Y_i, K_i) + t_i, \qquad i = 1, 2. \tag{2.4}$$

---

[10] For simplicity we assume all the charges/tolls are paid by shippers (or consignees), the final customers. In practice, these charges may not be levied directly on the final customers but on other users, such as shipping lines, forwarders and third-party logistics providers, who represent/serve the final customers and pass these costs to them.

[11] In addition to average (expected) congestion on the road, variability in congestion might affect shipping cost as well. For example, the variation in the level of road congestion could cause containers to arrive at the port earlier or later than the scheduled vessel departure time, enlengthening the port dwell time. As this is not the focus of the present paper, we abstract this issue away for the time being.

[12] It can be assumed that a truck is equivalent to three to four passenger car units. Thus, truck traffic volume can be transformed into passenger car units.

In equation (2.4) g_{Li} is the generalized cost encountered by an average commuter, which is the sum of commuter road toll t_i and commuter congestion cost $\alpha D_i(X_i + Y_i, K_i)$. Comparing with (2.2), parameter α denotes the ratio between per unit time delay cost of local commuters and per unit time delay cost of shippers, and is referred to as the 'commuter-shipper value-of-time ratio.' The sizes of passenger cars and trucks vary a lot. Longer and heavier trucks may have more influence on traffic flow than small cars. Thus, both the traffic volume and the relative value of time could subject to the way of unit conversion between cars and trucks, which is a complicated issue itself. In our context, we can assume one unit of cargo traffic is a fixed fraction of an average truck and variable X_i is the units of truck traffic converted into equivalent number of cars. If one truck is equivalent to n cars on average, the parameter α is the ratio of values of time between a passenger car and $1/n$ units of a truck. Thus, throughout the paper, when referring to one truck or one shipment, we are considering one passenger car equivalent unit of truck traffic or cargo. According to FHWA (2000), the impact of a truck on urban highway is equivalent to 1.5-4 passenger cars, depending on road conditions and truck size. However, the unit congestion cost of a commercial truck is usually more than three folds or even four folds of that of a passenger car. For example, it is estimated that in year 2011 the commercial trucks congestion cost is $86.81 per hour per vehicle while that of a passenger car is $20.99 per hour (TTI, 2012).[13] Yuen *et al.*(2008) listed a few earlier empirical estimations on the unit cost of travel delays for passengers and trucks. These earlier estimations together with the estimation by TTI (2012) suggest that it is sensible to impose $0 < \alpha < 1$ on the basis of passenger car equivalent unit, since per passenger car equivalent unit of freight traffic tends to have higher value-of-time than local commuting traffic. Equation (2.4) implicitly determines $Y_i = Y_i(X_i, K_i, t_i)$, and hence the total volume is a function of (X_i, K_i, t_i) as well: $V_i = X_i + Y_i(X_i, K_i, t_i)$. It is straightforward, using (2.3) and $\rho_{Li}'(\cdot) < 0$ (downward-sloping demand), to show:

[13] TTI (2012) finds in the US an average passenger car holds 1.25 passengers while each passenger's value of time is estimated to be $16.79. Thus, one passenger car's value of time is calculated by 1.25 times $16.79.

$$\frac{\partial Y_i}{\partial t_i} = \frac{1}{\rho_{Li}^{'} - \alpha(\partial D_i / \partial V_i)} < 0, \qquad \frac{\partial Y_i}{\partial K_i} = \frac{\alpha(\partial D_i / \partial K_{Li})}{\rho_{Li}^{'} - \alpha(\partial D_i / \partial V_i)} > 0,$$

$$\frac{\partial Y_i}{\partial X_i} = \frac{\alpha(\partial D_i / \partial V_i)}{\rho_{Li}^{'} - \alpha(\partial D_i / \partial V_i)} < 0, \qquad \frac{\partial V_i}{\partial X_i} = 1 + \frac{\partial Y_i}{\partial X_i} = \frac{\rho_{Li}^{'}}{\rho_{Li}^{'} - \alpha(\partial D_i / \partial V_i)} > 0. \qquad (2.5)$$

for $i = 1, 2$. Inequalities (2.5) indicate that local commuter traffic will rise if commuter toll falls, or road capacity rises, or port traffic falls. Less obviously, an increase in port traffic will, while reducing commuter traffic, increase overall road traffic. This observation sheds light on the impact of port-related freight movement on urban road congestion: an increase in freight throughput leads to more congestion on urban roads.

Solving the two equations in (2.1) for ρ_1 and ρ_2 yields inverse demand functions,[14]

$$\rho_1 = \rho_1(X_1, X_2), \qquad \rho_2 = \rho_2(X_1, X_2). \qquad (2.6)$$

Equilibrium in the cargo shipment market requires $\rho_i(X_1, X_2)$ to equal shippers' full price. Using (2.6), (2.2) and $Y_i = Y_i(X_i, K_i, t_i)$ we obtain:

$$p_i = \rho_i(X_1, X_2) - D_i(V_i, K_i) - T_i \equiv p_i(X_1, X_2; K_i, t_i, T_i), \quad i = 1, 2. \qquad (2.7)$$

Equation (2.7) indicates the (inverse) demand function faced by each port. Consequently, each port's profit is:

$$\pi^i = p_i(X_1, X_2; K_i, t_i, T_i) \cdot X_i \equiv \pi^i(X_1, X_2; K_i, t_i, T_i), \qquad i = 1, 2. \qquad (2.8)$$

Notice that in (2.8) the port operating costs are, for simplicity, assumed to be zero. This assumption allows us to focus on the interaction between urban road congestion and port

[14] Since two ports are substitutes, we have $\partial d_i / \partial \rho_i < 0$, $\partial d_i / \partial \rho_j > 0$, and that the Slutsky matrix of demand functions (2.1) is negative definite. It then follows that $\partial \rho_i / \partial X_i < 0$ and $\partial \rho_i / \partial X_j < 0$.

competition through the demand side.[15] Some empirical studies find that an improvement in landside port access conditions may lead to reduction in per-unit port operating costs, and then enhance the port's competitiveness vis-à-vis its rival (e.g. Cullinane and Song, 2006; Turner *et al.*, 2004). In our model, some of the mechanisms will operate through the demand side interaction as well and hence the zero-cost assumption will not affect the basic insights of the analysis.

We consider situations in which the ports simultaneously choose their throughput quantities,[16] taking road capacity (K_i) and tolls (T_i, t_i) as given. In general, road capacity investment and road pricing are longer-term decisions as compared to the ports' quantity decisions: infrastructure investment in roads is long lasting and typically irreversible. Similarly, whether to impose road tolls, and if so, by which scheme, take a long time to decide for political and implementation reasons, and once determined, it is hard to reverse. This 'two stages' formulation is the same as the one in De Borger *et al*. (2008) except that De Borger *et al*. consider ports compete in price. Also like De Borger *et al*.'s setting, each port maximizes its profit: as they argued, port-handling operations are often privately controlled by a few operators which can, for analytical simplicity, be aggregated into one private monopoly operator per port. In reality a port consists of various service providers and hence decision makers, such as port authority, terminal operators, drayage companies, warehouse operators, etc. It is not certain whether all these entities are in line with the profit maximization objective of a port, but our basic results should hold qualitatively as long as none of these entities operate in the rival port at the same time. However, if some major service providers also operate in the rival port – for example, the global terminal operator, Hustchson Port Holdings, simultaneously operates in many ports along the coastal line and within the Pearl River Delta of China, the competitive relationship between the two ports will be substantially changed in the sense that coordination or collusion between the two

[15] Our results will continue to hold for constant (but non-zero) operating costs, however.

[16] Note that we model port competition as a one-shot static game, as it is not easy to substantially and quickly adjust port outputs. Once the facilities built and purchased roughly determines the output level as it is expensive or even impossible to increase capacity in a short notice; on the other hand as many equipments (e.g. cranes) and labors (e.g. due to labor union) are expensive, laying-off to reduce output might be possible but can be difficult and expensive as well.

intermodal chains may evolve. The present paper only focuses on the case that competition is the main theme of the relationship between alternative chains, leaving the latter case a possible future study. It is possible and interesting to incorporate various forms of organization inside a port into our model, but doing so will not only make the analysis less tractable but distract the audience from the main objective of this study which is to investigate the interaction between port and its inland road congestion. The resulting Cournot equilibrium is, using (2.8), determined by the two first-order conditions, which can be rewritten as:

$$p_i = MEC_{Xi} \frac{\partial V_i}{\partial X_i} - \frac{\partial \rho_i}{\partial X_i} X_i, \qquad i = 1,2, \tag{2.9}$$

where $MEC_{Xi} = (\partial D_i / \partial V_i) X_i$ is the marginal external congestion cost encountered by trucks in region *i*. Recall that (2.5) implies $\partial V_i / \partial X_i$ is between 0 and 1. Thus, the first term on the right-hand side (RHS) of equation (2.9) indicates that the equilibrium charge by a profit-maximizing port internalizes that part of the external congestion by its users (truck traffic), a result obtained also by De Borger *et al*. (2008) under price competition. The second term on the RHS of (2.9) is the (positive) mark-up due to port *i*'s market power. In price competition, however, there is a third term which is a mark-down due to the rival's market power and road congestion. Therefore, quantity competition results in a higher equilibrium port charge than price competition.

While some of the results (such as (2.5) above) hold for more general specifications, we will, following De Borger *et al*. (2008), confine the analysis to situations where both the demands and congestion costs are linear:

$$\begin{aligned} &\rho_{Li}(Y_i) = M_i - m_i Y_i, \text{ where } M_i > t_i, \quad i = 1, 2 \\ &D_i(V_i, K_i) = c \cdot (V_i / K_i), \quad i = 1, 2 \\ &\rho_i(X_i, X_j) = A_i - a_i X_i - b_i X_j, \forall i \neq j, \quad i = 1, 2 \end{aligned} \tag{2.10}$$

The linear specifications facilitate satisfaction of the 'regularity conditions' that guarantee the existence of a unique Cournot equilibrium (see Appendix A.1).[17] They also facilitate the comparison between our results and De Borger *et al*. (2008)'s.

2.3 Effects of road capacity and tolls

2.3.1 Road capacity expansion

Denoting the equilibrium quantities as $X_1^*(K_1, K_2, t_1, t_2, T_1, T_2)$ and $X_2^*(K_1, K_2, t_1, t_2, T_1, T_2)$ for port 1 and port 2 respectively, we conduct comparative static analysis with respect to road capacities and tolls. Without loss of generality, we consider port 1 as the port in concern and port 2 as the rival port.

This subsection examines the effects of road capacity expansion. Notice first that the effects of chain 1's road capacity on X_1^* and X_2^* are determined by $\partial^2\pi^1/\partial X_1\partial K_1$. Specifically, the linear demand functions ensure that the reaction functions, derived from (2.9) for each port, are downward sloping. When $\partial^2\pi^1/\partial X_1\partial K_1 > 0$ (<0), an increase in K_1 will increase (decrease) the marginal profit of port 1, shifting port 1's reaction function outward (inward). However, an increase in K_1 does not affect port 2's marginal profit ($\partial^2\pi^2/\partial X_2\partial K_1 = 0$), keeping its reaction function unchanged. Consequently, the equilibrium will move along port 2's reaction function, depending on the sign of $\partial^2\pi^1/\partial X_1\partial K_1$. In particular, the equilibrium output of port 1 will rise and that of port 2 will fall if and only if $\partial^2\pi^1/\partial X_1\partial K_1 > 0$.

The effect of K_1 on port 1's marginal profit can be decomposed into four components:

$$\frac{\partial^2\pi^1}{\partial X_1\partial K_1} = -\frac{\partial D_1}{\partial K_1} - \frac{\partial D_1}{\partial V_1}\frac{\partial Y_1}{\partial K_1} - X_1\frac{\partial V_1}{\partial X_1}\frac{\partial^2 D_1}{\partial V_1\partial K_1} - X_1\frac{\partial D_1}{\partial V_1}\frac{\partial^2 V_1}{\partial X_1\partial K_1}. \tag{2.11}$$

[17] The regularity conditions – i.e. quantities are "strategic substitutes" and the equilibrium is stable – guarantee the existence of a unique equilibrium (e.g. Bulow *et al.*, 1985; Tirole, 1988). Stability of the equilibrium over the entire range of interest further renders comparative statics meaningful.

The first component is by (2.3) positive and indicates a congestion reduction directly from road capacity expansion. The second component is by (2.3) and (2.5) negative, indicating an increase in congestion as capacity expansion induces more local commuters to travel. Using (2.3) and (2.5) we can prove that the combined effect of the first two components is positive (see Appendix A.2). The last two terms in (2.11) capture effects on the marginal change of congestion delay, i.e. the second-order effects. The third term is positive, indicating that an increase in road capacity moderates the negative impact of one more output unit on the congestion delay and therefore leads to a positive marginal change of port charge. The last term indicates that road capacity expansion affects the change of total traffic volume resulted from marginal increase in cargo output and such impact on the total traffic volume then affects the congestion delay. This term is negative under linear specifications (2.10). The combined effect of the last two terms is nonetheless positive (shown in Appendix A.2). Thus, $\partial^2 \pi^1 / \partial X_1 \partial K_1 > 0$; consequently, the equilibrium output of port 1 will rise and that of port 2 will fall as region 1 expands its road capacity.[18]

The results that $dX_1^* / dK_1 > 0$ and $dX_2^* / dK_1 < 0$ are also useful in deriving other comparative-static results. For example, the impacts of road capacity expansion on port profits are given by:

$$\frac{d\pi^{1*}}{dK_1} = X_1^* \frac{\partial \rho_1}{\partial X_2} \frac{dX_2^*}{dK_1} - X_1^* \frac{\partial D_1}{\partial K_1} \frac{\partial V_1}{\partial X_1} > 0; \qquad \frac{d\pi^{2*}}{dK_1} = X_2^* \frac{\partial \rho_2}{\partial X_1} \frac{dX_1^*}{dK_1} < 0. \qquad (2.12)$$

As can be seen, the impact on port 1's profit depends on two effects: first, an increase in region 1's road capacity reduces the rival port's output which in turn raises the own port's demand (the ports provide substitute services). Second, the congestion reduction directly from road capacity expansion results in more cargo traffic going through port 1. As both effects are positive, road capacity expansion by the region will improve port 1's profit. As for

[18] In the present paper, the terms "chain" and "region" are interchangeable unless when welfare is in concern. The "region" is defined as an area encompassing the port, urban roads and their administrators, and local commuters. Its welfare, in this context, refers to the collective well-being of these parties, which is different from the total surplus of an intermodal chain.

the rival port, the profit impact depends solely on the indirect effect on its demand: an increase in region 1's road capacity increases port 1's output which in turn reduces port 2's demand and hence profit. The discussion leads to:

Proposition 2.1: *An increase in road capacity by an intermodal chain will, at equilibrium, (a) increase its port's output, port charge and profit, (b) reduce the rival port's output, port charge and profit, and (c) reduce road congestion in both chains.*

Proof: See Appendix A.2.[19]

Proposition 2.1's results are listed in Table 2.1. Regarding (equilibrium) port charges, the effect of increased port traffic in chain 1 outweighs the effect of decreased port traffic in chain 2 and, as a consequence, there is a downward pressure on port charges. For port 1, however, road capacity expansion in its inland directly alleviates chain 1's road congestion which allows it to overcome the downward price effect. Furthermore, note that road capacity expansion by a chain will alleviate not only its own road congestion, but road congestion of its rival chain as well, thus creating a positive externality across the regions.

[19] Note that propositions presented in this chapter are all based on the linear specification (2.10).

Table 2.1 Effects of road capacity

	Quantity competition		Price competition
	Ports max. profit	Ports max. local welfare	Ports max. profit
dX_1^* / dK_1	+	+	+ iff X_1^* is large enough and X_2^* is small enough
dp_1^* / dK_1	+	+	+ iff X_1^* is small enough (plausible) [a]
$d\pi_1^* / dK_1$	+	+ if truck toll low [b]	-
dD_1^* / dK_1	-	-	- if $dp_1^* / dK_1 > 0$
dX_2^* / dK_1	-	-	- iff X_1^* is large enough and X_2^* is small enough
dp_2^* / dK_1	-	-	- [a]
$d\pi_2^* / dK_1$	-	- if truck toll low	+ iff both X_1^* and X_2^* are small enough
dD_2^* / dK_1	-	-	?

Notes: a) These results are from De Borger *et al.* (2008), the rest of results are calculated by the authors based on De Borger *et al.* (2008)'s model. b) Truck toll low means $T_i < MEC_i(\partial V_i / \partial X_i) - X_i(\partial \rho_i / \partial X_i) - t_i(\partial Y_i / \partial X_i)$.

2.3.2 Road tolls

This subsection starts with the case where trucks and commuters are tolled separately (discriminative tolls), followed by an examination of fixed-ratio tolls. Applying the same procedure as in the previous subsection, we obtain three propositions (Propositions 2.2, 2.3 and 2.4 below).

Proposition 2.2: *An increase in commuter toll by an intermodal chain will, at equilibrium, (a) increase its port's output, port charge and profit, (b) reduce the rival port's output, port charge and profit, and (c) reduce road congestion in* <u>*both*</u> *chains.*

Proof: See Appendix A.3.

Comparing Propositions 2.2 and 2.1, we see that commuter road toll (t_i) has the same impacts on the ports' outputs, charges and profits, and road congestion as road capacity. Like the road-capacity case, the following marginal-profit effect plays a key role:

$$\frac{\partial^2 \pi^1}{\partial X_1 \partial t_1} = -\frac{\partial D_1}{\partial V_1}\frac{\partial Y_1}{\partial t_1} > 0\,, \tag{2.13}$$

where the inequality follows from (2.3) and (2.5). Essentially, an increase in commuter toll by chain 1, while having zero effect on the marginal change of port charges (i.e. no second-order effects),[20] reduces local commuter traffic which in turn lowers congestion cost of the freight shipment and improves the marginal profit of port 1. As a consequence, port 1's equilibrium output will rise while port 2's equilibrium output falls. Further, the improvement on road congestion outweighs the reduction in shippers' willingness-to-pay due to the output expansion, raising port 1's charge. Higher output and higher port charge lead to an increase in port 1's profit at equilibrium. The rival port's profit falls as the higher output in port 1 reduces the marginal shipper's willingness-to-pay for its service and hence its charge.

Proposition 2.3: *A <u>decrease</u> in truck toll by an intermodal chain will, at equilibrium, (a) increase its port's output, port charge and profit, (b) reduce the rival port's output, port charge and profit, and (c) increase its road congestion while reducing the rival's road congestion.*

Proof: See Appendix A.4.

The rationale of Proposition 2.3 is straightforward: a fall in truck toll by chain 1 reduces the generalized cost paid by shippers in chain 1 and allows its port to charge a higher fee, thus increasing port 1's marginal profit. The resulting outward shift of port 1's reaction function raises the port's equilibrium output, while reducing the equilibrium output of port 2. Furthermore, although the rising port traffic in chain 1 suppresses local commuter traffic, the former dominates the falling local traffic. As a result, unlike the above two policies (road capacity expansion and adding commuter toll), a lower truck toll benefits the port but negatively impacts on the local commuters.

Now consider that trucks and commuters are charged based on a fixed-ratio toll schedule.

[20] This may not be true if linear specifications (2.10) do not hold.

Following Yuen *et al.* (2008), there may be some *a priori* rule which says that trucks and cars are charged according to some fixed ratio (for example, a truck pays three times what a car pays). Without loss of generality, we set the ratio to one, i.e. $T_i = t_i = \tau_i$.

Proposition 2.4: *An increase in fixed-ratio toll by an intermodal chain will, at equilibrium, reduce its road congestion. Furthermore, if and only if the commuter-shipper value-of-time ratio is sufficiently small, it will: (a) increase its port's output, port charge and profit; (b) reduce the rival port's output, port charge and profit; and (c) reduce the rival's road congestion.*

Proof: See Appendix A.5.

When commuters and trucks are charged by a fixed-ratio toll, an increase in the toll by chain 1 combines the effect of raising commuter toll with that of raising truck toll. Since the two tolls have opposite impacts on port 1's profit by Propositions 2.2 and 2.3 (see also the summaries in Table 2.2), increasing fixed-ratio toll is less likely to favor port 1 than raising commuter toll alone. As a result, the value-of-time ratio α plays an important role in determining its net effect. Particularly, only when the shippers' value-of-time is large enough relative to the commuters' (i.e. α is small enough), will the impact from commuters' suppressing their traffic (due to the increased toll) dominates, resulting in port 1's gain and port 2's loss in profit. As for road congestion of its own region, it is clear that increasing commuter and truck road tolls at a fixed ratio will reduce congestion farther than just raising commuter toll regardless of α values, since shippers must pay higher tolls as well.

Table 2.2 Effects of road tolls

Impacts on:	Ports maximize profit: Increase t_1	Ports maximize profit: Increase T_1	Ports maximize profit: Increase τ_1	Ports maximize local welfare: Increase t_1	Ports maximize local welfare: Increase T_1	Ports maximize local welfare: Increase τ_1
X_1^*	+	-	+ if α low - if α high	+	0	+
p_1^*	+	-	+ if α low - if α high	+	-	? if α low - otherwise
π_1^*	+	-	+ if α low - if α high	+ if truck toll low ? otherwise	-	?
D_1^*	-	-	-	-	0	-
S_{Y1}^*	-	+	-	-	0	-
X_2^*	-	+	- if α low + if α high	-	0	-
p_2^*	-	+	- if α low + if α high	-	0	-
π_2^*	-	+	- if α low + if α high	- if truck toll low ? otherwise	0	- if toll low ? otherwise
D_2^*	-	+	- if α low + if α high	-	0	-
S_{Y2}^*	+	-	+ if α low - if α high	+	0	+

Notes: S_{Yi}^* denotes the local commuter surplus of chain *i* at the equilibrium. 'α low' means $\alpha < 1 - m_1K_1/c$ and 'α high' means $\alpha > 1 - m_1K_1/c$. The cut-off is valid iff $0 < 1 - m_1K_1/c < 1$. 'Truck toll low' means $T_i < MEC_i(\partial V_i/\partial X_i) - X_i(\partial \rho_i/\partial X_i) - t_i(\partial Y_i/\partial X_i)$. 'Toll low' means $\tau_i < MEC_i - X_i((\partial \rho_i/\partial X_i)/(\partial V_I/\partial X_i))$.

2.4 Further discussion

2.4.1 Welfare-maximizing ports

The analysis has so far dealt with profit-maximizing ports. In practice, port ownership and administration structures are diverse and it is difficult to determine a clear objective function, but the private/public classification is quite conventional and common (Bichou and Gray, 2005) and can be considered as two benchmark cases, and as a consequence, welfare optimization may be a relevant objective for a port.[21] Following De Borger *et al.* (2008), welfare for the region containing port *i* is taken as the sum of local commuter surplus (F^i), port profit and toll revenue from trucks, minus investment cost of road capacity ($r_i K_i$):

$$W^i = F^i + \pi^i + T_i X_i - r_i K_i = \int_0^{Y_i} \rho_{Li}(y)dy - \alpha D_i Y_i + \pi^i + T_i X_i - r_i K_i, \quad i = 1,2, \quad (2.14)$$

noting that the toll revenue from local commuters represents an internal transfer. The first-order conditions of (2.14) with respect to port quantities can be expressed as,

$$p_i = MEC_i \frac{\partial V_i}{\partial X_i} - \frac{\partial \rho_i}{\partial X_i} X_i - t_i \frac{\partial Y_i}{\partial X_i} - T_i, \qquad i = 1,2, \quad (2.15)$$

where $MEC_i = MEC_{Xi} + MEC_{Yi} = (\partial D_i / \partial V_i) X_i + \alpha (\partial D_i / \partial V_i) Y_i$ is the marginal external congestion cost encountered by both commuters and trucks in region *i*. A comparison between equations (2.15) and (2.9) reveals that relative to a profit-maximizing port, the port charge of a welfare-maximizing port has an additional upward adjustment so as to internalize the marginal external congestion cost encountered by local commuters (which is not a concern for a profit-maximizing port). In addition, it has an upward adjustment equivalent to the marginal revenue of road toll levied on local commuters, and a downward adjustment equivalent to the marginal revenue of road toll levied on trucks.

Proposition 2.5: *For welfare-maximizing ports, an increase in road capacity by an intermodal chain will, at equilibrium, (a) increase its port's output and port charge, (b) reduce the rival port's output, port charge and the rival region's welfare, (c) reduce road congestion in both chains and increase both regions' commuter surplus, and (d) if tolls paid by trucks are low, increase its port's profit and reduce the rival port's profit.*

Proof: See Appendices A.6 and A.7.

---

[21] This can, for example, be considered as having local public control over the port's decision variables. De Borger *et al.* (2008) also look at ports' welfare optimal behavior in their appendices and numerical analysis.

A comparison between Propositions 2.5 and 2.1 suggests that the comparative-static effects of road capacity on intermodal chains with welfare-maximizing ports are similar to those with profit-maximizing ports, except for port profits. In the case of welfare-maximizing ports, although the capacity expansion improves the value of the port's objective function (excluding investment cost), its profit may fall if tolls paid by trucks are so high that the port ends up subsidizing its shippers (i.e. setting a negative charge).[22]

As for the effects of road tolls, one interesting observation in the case of welfare-maximizing ports is that $\partial^2 W^1 / \partial X_1 \partial T_1 = 0$, i.e. changes in truck toll alone have no impact on the marginal welfare. This is because at the equilibrium the port charge equals the marginal shipper's full price subtracting congestion cost and truck toll and, consequently, the toll revenue from trucks is just an internal transfer from the port's profit. Further, $\partial^2 W^1 / \partial X_1 \partial t_1 = \partial^2 W^1 / \partial X_1 \partial \tau_1$, implying commuter toll and fixed-ratio toll have the same effects on the Cournot equilibrium outputs.

Proposition 2.6: *For welfare-maximizing ports, an increase in fixed-ratio or commuter tolls by an intermodal chain will, at equilibrium, (a) increase its own port's output, (b) reduce the rival port's output and port charge and the rival region's welfare, (c) reduce road congestion in both chains and its own region's commuter surplus while improving the rival region's commuter surplus, and (d) if tolls paid by trucks are low, reduce the rival port's profit.*

Proof: See Appendix A.8.

Proposition 2.6 is also summarized in Table 2.2. Again, as in the case of profit-maximizing ports, the impacts of commuter toll are similar to those of road capacity.[23] However, an

[22] More specifically, when the ports subsidize shippers, an increase in road capacity by a chain will increase its own port's profit and reduce the rival port's profit if and only if the percentage change in the subsidy is greater than the percentage change in throughput.

[23] Under a discriminative toll system, similar to road capacity expansion, raising commuter toll increases its own port's (equilibrium) charge and, if the truck toll is low, increases its profit as well. The proof for Proposition 2.6 is available in Appendix A.

increase in fixed-ratio toll has ambiguous impacts on its own port's charge and profit.

2.4.2 Comparison with existing literature

As indicated earlier, one of the main differences between the present paper and most of the earlier literature is that we consider quantity rather than price competition between ports. In general, which model of competition is applicable to a particular industry depends in large part on its production technology. In Cournot (quantity) competition, firms commit to quantities, and prices then adjust to clear the market implying the industry is flexible in price adjustments, even in the short run. On the other hand, in Bertrand (price) competition, capacity is unlimited or easily adjusted in the short run. Quantity competition may be more realistic than price competition in the case of ports as port capacity is difficult to increase (Quinet and Vickerman, 2004). The main reason for why port capacity is difficult to change (relative to the ease and rapidity with which prices can be adjusted) is that port investment is lumpy, time-consuming and irreversible. Indeed, with capacity constraints Van Reeven (2010) assumes quantity competition between port terminal operators based on the Kreps and Scheinkman (1983)'s argument of capacity-constrained price competition yielding quantity competition. Furthermore, Menezes *et al.* (2007) empirically estimated the market 'conduct parameters' with respect to port charges of the three largest, competing Australian seaports. Our calculation based on their results indicates that at the 0.1 level of statistical significance, the hypothesis of price competition among the ports is rejected.[24] In reality both terminal operators and port authorities can make port capacity (quantity) decisions, depending on who own which part of the facility, the type of terminal operators as well as the contractual relationship between the port authority and terminal operators. For example, the port authority of a landlord port may have the power to determine the size of the land and wharves or the capacity of common facility shared among terminals while individual terminal operators may determine the capacity of their exclusive facility such as cranes and labors within each terminal. As terminals can be owned or operated by shipping lines, local

[24] For studies using conduct parameters to assess the empirical relevance of certain oligopoly models to a particular market, see, e.g., Bresnahan (1989) and Brander and Zhang (1990).

port authorities, specialized terminal operating firms and infrastructure investment firms, their objectives could vary as well. There are also various forms of private-public partnerships under which the decision on the capacity level may stem from the negotiation process between terminal operators and port authorities. As the issue is so complicated that a separate study is necessary, the present study limits analysis to the situation that port profit or regional welfare is maximized by an entity (e.g. a dominant or even monopoly terminal operator or a port authority leasing facility to a number of terminal operators who fiercely compete with each other) which can extract almost all the profits from other port users and does not have interest in the rival port. In fact, port authority may have a dominant power over private terminal operators in determining port outputs as in many concession contracts the port authority requires throughput guarantees which specify minimum throughput the terminal operator should achieve otherwise a penalty will be imposed on the terminal operator (see Notteboom, 2006a, for detailed discussion on throughput guarantees).

It is therefore worthwhile to study the interaction between inland road congestion and port competition by assuming ports' competing in quantities, and compare the results with those of price competition. Table 2.1 lists the effects of road capacity on the ports' outputs, charges, profits and road congestion for both competition modes. As can be seen, clear-cut predictions are obtained under quantity competition, which is seldom the case under price competition. De Borger *et al.* (2008) indicate that under price competition, the (equilibrium) port charges plausibly increases in road capacity. If so, capacity expansion has the same effects on port charges and road congestion for the two competition modes. Perhaps the most eye-catching difference is in the effects on port profits: Under quantity competition, capacity expansion benefits its own port and hurts the rival port; under price competition we can show, based partly on De Borger *et al.*'s analysis, that capacity expansion reduces its own port's profit even if the port charge may rise, but may benefit the rival port under certain conditions. This implies that a region's motive for investing in road capacity so as to improve its port's profit is strengthened if port competition is in quantities rather than in prices.

This effect would remain even if a region's objective is to maximize its overall welfare rather than just its port's profit. To see this, following De Borger *et al.* (2008)'s two-stage game formulation in which road capacity investment is decided ahead of the competition of profit-maximizing ports given a fixed-ratio toll schedule, the optimal road investment policy is given by the first-order condition of (2.14):

$$\begin{aligned} & X_1^* \frac{\partial \rho_1}{\partial X_2} \frac{dX_2^*}{dK_1} + \left(\tau_1 - \alpha Y_1 \frac{\partial D_1}{\partial V_1} \right) \frac{\partial V_1}{\partial X_1} \frac{dX_1^*}{dK_1} \\ & \qquad + \left(X_1^* \frac{\partial p_1}{\partial K_1} - \alpha Y_1 \frac{\partial D_1}{\partial K_1} + \left(\tau_1 - \alpha Y_1 \frac{\partial D_1}{\partial V_1} \right) \frac{\partial Y_1}{\partial K_1} \right) = r_1 \end{aligned} . \tag{2.16}$$

Equation (2.16) is analogous to the policy derived in De Borger *et al.* under price competition. In particular, the first term is the induced port profit via changes in the rival port's output and is, as indicated above, positive. The corresponding term under price competition is the induced port profit via changes in the rival port's charge, which is, by contrast, plausibly negative. The incentive to elicit a favorable strategic quantity response by the rival port will thus create a tendency for the region to invest more on road capacity as compared to the case of price competition. Furthermore, if road capacity expansion is desired, the regional government is less likely to receive objection from its port if the mode of competition is in quantities; rather, it may be even possible to raise fund from the port for the expansion project. Under price competition, by contrast, the port may require lump-sum transfers from the government so as to compensate for its loss due to such expansion. Note that even if road toll equals the marginal external cost of commuters, the optimal capacity level tend to be higher under quantity competition, since the effect of road expansion on port profit is positive under quantity competition while negative under price competition.

Another difference between the present paper and De Borger *et al.* (2008) is that this paper studies in more detail the role of tolls. The discriminative/fixed-ratio tolls have been examined in, e.g., De Borger *et al.* (2005, 2007) and Ubbels and Verhoef (2008) where each region imposes road tolls on local commuters and 'transit' truck traffic.[25] Port competition is not considered in these papers and so transit truck traffic is not port-related; rather, such traffic is generated from jurisdictions outside of the regions under consideration. Applying

[25] The discriminative/fixed-ratio tolls have also been considered in Yuen *et al.* (2008) and De Borger-De Bruyne (2011) under different contexts as indicated in the introduction.

the same approach we derive the optimal tolls for commuters and trucks respectively. Same as De Borger *et al.* (2005, 2007), the optimal commuter toll is equal to MEC_1 (the marginal external congestion cost encountered by both commuters and trucks in region *i*). The optimal truck toll is given below:

$$T_1^* = MEC_1 - MEC_{X1}\frac{\partial V_1}{\partial X_1} - X_1^* \frac{\partial \rho_1}{\partial X_2}\frac{\partial X_2^* / \partial T_1}{\partial X_1^* / \partial T_1}. \tag{2.17}$$

The second term on the RHS of (2.17) is the deduction of marginal externality internalized by the ports and the last term indicates competition intensity. As both terms are negative, contrary to the earlier papers, the optimal truck toll is lower than MEC_1. Hence, the local government taxes more on the local commuters than on the shippers (i.e. transit traffic under De Borger *et al.*'s setting). Equation (2.17) can be rewritten into:

$$T_1^* = MEC_{Y1} - MEC_{X1}\frac{\partial Y_1}{\partial X_1} - X_1^* \frac{\partial \rho_1}{\partial X_2}\frac{\partial X_2^* / \partial T_1}{\partial X_1^* / \partial T_1}. \tag{2.18}$$

Since the second term on the RHS of (2.18) is positive, when ports are local monopolists (i.e. the last term of (2.18) disappears), the optimal truck toll is above MEC_{Y1} (the marginal external congestion cost of commuters) which is consistent to the literature. This is no longer the case when ports compete however: To attract shippers from the rival and improve its port's profit, the local government will lower the truck toll, sometimes even below MEC_{Y1}. The situation won't happen in the earlier papers as port competition is absent.

Under the fixed-ratio toll, the optimal pricing rule for region 1 is given by:

$$\begin{aligned}\tau_1^* &= MEC_1 - MEC_{X1}\frac{\partial V_1}{\partial X_1}\left(\frac{dX_1^*}{d\tau_1}\Big/\frac{dV_1^*}{d\tau_1}\right) - X_1^* \frac{\partial \rho_1}{\partial X_2}\left(-\frac{dX_2^*}{d\tau_1}\Big/\frac{dV_1^*}{d\tau_1}\right)\\ &= MEC_{Y1} + MEC_{X1}\left(\frac{\partial Y_1}{\partial \tau_1}\Big/\frac{dV_1^*}{d\tau_1}\right) - X_1^* \frac{\partial \rho_1}{\partial X_2}\left(-\frac{dX_2^*}{d\tau_1}\Big/\frac{dV_1^*}{d\tau_1}\right)\end{aligned}. \tag{2.19}$$

Based on Proposition 2.4, it is easy to show that if the value-of-time ratio, α, is small (large, respectively), the optimal road toll must be greater (smaller, respectively) than MEC_1.[26] A comparison between (2.9) and (2.19) suggests how much congestion externality imposed by one additional shipper is internalized by the fixed-ratio toll together with port charge. By summing up the first term of (2.9) and the first two terms of (2.19), we obtain:

$$MEC_1 + MEC_{X1}\frac{\partial V_1}{\partial X_1}\left(1-\frac{dX_1^*}{d\tau_1}\bigg/\frac{dV_1^*}{d\tau_1}\right) = MEC_1 + MEC_{X1}\frac{\partial V_1}{\partial X_1}\frac{dY_1^*}{d\tau_1}\bigg/\frac{dV_1^*}{d\tau_1}. \qquad (2.20)$$

If congestion externality is fully internalized, (2.20) will equal to MEC_1. However, it is easy to show that (2.20) is larger than MEC_1 regardless the value of α. Therefore, in terms of the congestion externality alone, shippers will always pay more than the amount of externality they impose on other road users. Abstracting away port competition, not only the above will continue to hold, but also the local government will always set a toll higher than MEC_{Y1}. The former result differs from the result in Ubbels and Verhoef (2008) who do not include port profit into the objective function and consequently, find the local government always charging the toll higher than MEC_1. The latter result is in line with the findings in De Borger *et al.* (2005, 2007) and Ubbels and Verhoef (2008), but the nature of such over-charge differs. In the literature, it is driven by the local government's incentive to capture the toll revenues from the transit traffic, while in our case, holding commuter traffic constant, an increase in road toll requires the same amount of reduction in port charge so as to maintain the throughput. The revenue extracted on the road is thus exactly offset by the port profit loss, and so the over-charge is from the fact that a higher toll reduces congestion by contracting commuter traffic such that the port can charge higher to extract the congestion cost savings from shippers. However, when port competition enters into the picture and α is high, the aforementioned over-charge may turn into an under-charge, because when the value-of-time of shippers is relatively small, road congestion cost savings will not compensate for the high tolls paid, and as a result a high road toll is more likely to turn away shippers to the rival port

[26] This is because the last two terms in the first line of (2.19) are both positive if $\alpha < 1 - m_1K_1/c$ and negative if $\alpha > 1 - m_1K_1/c$.

and hence the local government may have to charge a toll below MEC_{Y1}.

2.5 Concluding remarks

This paper has analytically examined the interaction between urban road congestion and port competition, treating port competition as part of rivalry between alternative intermodal transportation chains. Contrary to the case of ports' competing in prices analyzed in the earlier literature, our quantity-competition model predicted that an increase in road capacity by an intermodal chain will likely benefit its port while hurting the rival port (especially when tolls on truck traffic are low). This implies that a region's motive for investing in road capacity would be strengthened if port competition is in quantities rather than in prices. As the non-cooperative capacity rivalry between regions (chains) would lead to over-investment in port-related road (and other) capacities, our analysis provided an explanation for the prevailing observation that such capacities are 'on the high side' in port ranges where ports compete vigorously (e.g. the Le Havre-Hamburg range). The high capacities may nevertheless be socially desirable globally as we showed that road congestion of *both* regions would fall as a result of one region's capacity expansion.

In addition to the above policy implication regarding capacity investment, our analysis offered several implications for pricing policies. For one, road tolls may be used by a region to gain a strategic advantage by influencing port competition. Owing to the strategic considerations, the marginal social cost pricing for roads advocated by, e.g., the EC may not be implemented by a regional government. Second, the nature of deviation from the marginal social cost pricing can depend on the toll systems adopted (discriminative vs. fixed-ratio tolls) and the value-of-time of road users (shippers vs. commuters). For instance, when fixed-ratio tolls are used, a local government would adopt the above marginal social cost pricing by setting congestion toll above the marginal external congestion costs, provided that the value-of-time of shippers is sufficiently large relative to that of commuters. When a discriminative toll system is implemented, however, commuters are tolled at the marginal external congestion costs while truck tolls are much lower. Third, our analysis showed that while a high toll by a region relieves its road congestion, it may increase road congestion in the rival region. This and other results of the paper suggested the importance for major and competing seaport regions to coordinate on their pricing and investment decisions regarding port-related roads and facilitates.

3 Effects of Urban Road Congestion on Port Rivalry and Efficiency: Empirical Analysis of U.S. Container Ports[27]

3.1 Introduction

The performance of seaports is strongly intertwined with the development and performance of their respective inland networks (Notteboom and Rodrigue, 2008). It is found that the shippers' ocean carrier and/or port selection process has been heavily affected by the landside operation attributes, such as land carrier service quality and proximity to the hinterland (Lirn *et al.*, 2004; Slack, 1985; Ugbonma *et al.*, 2006; Yuen *et al.*, 2012). Research on major container ports in China and the Asia-Pacific region has found hinterland connection as one of the key factors in determining port competitiveness and productivity (Yuen *et al.*, 2012, 2013). It has also been pointed out by Fan *et al.* (2012) that seaports which possess more hinterland transport infrastructure are more likely to survive in newly established trade flow markets. As depicted by De Langen (2008), hinterland accessibility or connectivity is a broad concept. In this chapter, we focus on road congestion around the ports and, to a lesser extent, rail services, because these are the most commonly used modes of transport on the inland portion of intermodal transport chains.

In the United States, for example, the limited capacity of the highway system resulted in an inability to withstand demand shocks, which caused congestion on key freight transport segments (TTI, 2010). Road congestion delays raise traveling times and fuel costs, lower the reliability of commercial truck operations and increase the chance of missing schedules (FHWA, 2004). All of these factors could translate into costs endured by shippers who select the intermodal chain to ship their cargoes.[28] As a large proportion of the containers are moved into and out of seaports by trucking, the congestion at urban roads surrounding the

[27] Chapter 3 has been split into two papers. One has been published: Wan, Y., Zhang, A. and Yuen, A. (2013) Urban road congestion, capacity expansion and port competition: empirical analysis of U.S. container ports, *Maritime Policy and Management*, 40(5), 417-438. The other one has been accepted for publication: Wan, Y., Yuen, A. and Zhang, A. (2013) Effects of hinterland accessibility on U.S. container port efficiency, *International Journal of Shipping and Transport Logistics*.

[28] TTI (2010) estimated that in 2009, the average (per urban area) congestion costs suffered by trucks in the U.S. amounted to 1,273 million US dollars and 215 million US dollars for very large (population over 3 million) and large (population over 1 million but less than 3 million) urban areas, respectively.

seaports has become one of the essential factors that influence a port's ability to sustain its competitiveness. Capacities of local road, rail and trucking have been identified as top capacity drivers of container ports (Maloni and Jackson, 2005). A survey on trucking firms based in California – where roads around the ports of Los Angeles and Long Beach were heavily congested – revealed that road congestion has seriously affected trucking firms' business, including access to port and rail terminals (Golob and Regan, 2000).

In chapter 2, we analytically examined how capacity expansion and congestion pricing of urban roads affect port competition. Although road congestion pricing has not been widely implemented, road capacity expansion is one of the most common and important congestion mitigation strategies (FHWA, 2004).[29] The survey conducted by Golob and Regan (2000) found that adding more freeway lanes is strongly supported by U.S.-based commercial trucking firms, especially short haulers connecting maritime ports and rail terminals. Therefore, to accommodate more port-related traffic and hence enhance competitiveness of the port in their jurisdictions, policy makers are likely to take road capacity expansion into serious consideration when searching for measures to alleviate road congestion. For example, a number of major UK road improvement schemes have been carried out in order to benefit ports around this area (Pettit and Beresford, 2008). However, as mentioned in chapter 2, the impacts of road capacity expansion are in general ambiguous, depending on assumptions about how ports compete and other factors.[30] Thus, it is of interest to examine empirically the impacts of road congestion, as well as the effect of road capacity expansion on the port competition.

Based on a sample of major U.S. container ports, this chapter empirically examine urban road congestion, road capacity expansion and port performance. More specifically, the paper investigates the following three questions: (i) whether an increase in road congestion reduces

[29]Other congestion mitigation strategies include: improving operational efficiency, providing truck dedicated facilities, improving traffic management, and managing demand (e.g. through congestion pricing). For more detailed discussion, see FHWA (2004) and Golob and Regan (2000).

[30] Other theoretical studies in this strand of the literature include Yuen *et al.* (2008) and Zhang (2008). The former investigates the effects of congestion pricing implemented at a gateway port on its hinterland's optimal road pricing, road congestion and social welfare, whereas the latter focuses on the corridor congestion and capacity investment rather than urban road congestion *per se*.

the container throughput of the port affiliated to the urban area while benefiting the rival ports nearby; (ii) how road capacity changes in both the port's and its rival's respective urban areas will affect the port's container throughput; and (iii) how rail services and road congestion will affect port efficiency. We find that port throughput is related to urban road congestion and capacity investment. Specifically, a 1% increase in road congestion is associated with 0.90-2.48% reduction in the container throughput of the seaport affiliated with that urban area, while implying an increase in the throughput of its rival port by 0.62-1.69%. Further, the impact of road capacity expansion on container throughput differs among the ports in our sample, depending in particular on road capacities around a port and its rival ports. Despite the differential impacts of road capacity expansion, adding local road capacity tends to benefit the port in the urban area in question and harm its rivals when such road expansion solely affects ports' throughput via road congestion. The container port efficiency is measured by data envelopment analysis (DEA). Provision of on-dock rail facility at container terminals is negatively correlated with container port efficiency. The impacts of Class I rail services, on the other hand, are ambiguous. In general, there is a negative association between road congestion around the port and port efficiency. However, road congestion may have a negligible, or even positive, relationship with DEA scores of ports with larger container throughput volume.

Empirical work which investigated both port competition and hinterland characteristics is relatively rare. Most of the studies predicted the distribution of container traffic volumes among competing ports with numerical simulations and/or multinomial logit model based either on the total delivery costs or on the distance to access the seaport from hinterland (e.g. Garcia-Alonso and Sanchez-Soriano, 2009; Luo and Grigalunas, 2003; Nir *et al.*, 2003; Veldman and Bückmann, 2003; Zongdag *et al.*, 2010). None of those studies looked into the impact of road capacity or road congestion delays on port competition.[31] The impact of road capacity provision on the road traffic volume (combining both commercial trucks and commuter cars) has been extensively studied (e.g. see Noland and Lem, 2002; Kitamura,

[31] Pope et al. (1995) assessed the impact of adding a new section of highway on road congestion around the port of Hampton Roads by a simulation model, but port competition was excluded from their analysis.

2009 for useful reviews), but this literature focuses on local residents and commuter cars and does not pay any particular attention to port cities and port-related activities. There is an extensive literature on seaport productivity with DEA (see Panayides *et al.*, 2009, for a recent review), but container ports in North America are rarely covered by previous studies. Empirical evidence on the impacts of hinterland connection on seaport productivity is also scarce. Turner *et al.* (2004) examined 26 North American ports from 1984 to 1997 and measured the impacts of rail services on these ports' productivity, but road congestion was excluded from their work.

The paper is organized as follows. In Section 3.2, we briefly go through the theory regarding to the connection between road congestion, capacity and port competition as well as port productivity. Then, we introduce the research methods and our econometric models in Section 3.3, followed by the description of the dataset in Section 3.4. Section 3.5 presents the major empirical results and Section 3.6 provides the concluding remarks.

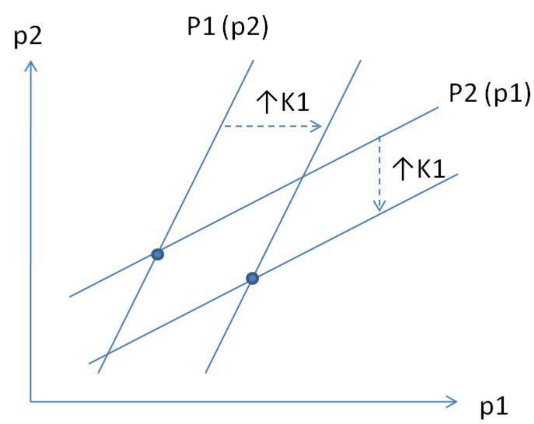
3.2 Road congestion, capacity and port performance: theory

3.2.1 Road congestion, capacity and port throughput

Recall the analytical model in chapter 2. Road congestion and port output, such as throughput, affect each other simultaneously. On the one hand, inequality (2.5) suggests when port throughput X_i increases, the overall traffic V_i will increase, resulting in more congestion delays on the road. On the other hand, when the road congestion around port *i* increases, the generalized cost of using port *i* increases as indicated by equation (2.2), and as a result, the container throughput of port *i* tends to decrease, while that of port *j* tends to increase. The cost of delay due to congestion on roads connecting gateway ports with intermodal transport facilities (such as rail terminals) or inland destinations is commonly identified as a part of the generalized cost. This portion of generalized cost may be significant. For example, trucks travelling in the Los Angeles-Long Beach area incurred $2.3 billion road congestion cost, accounting for 0.55% of the value of commodity carried in 2009 (TTI, 2010). Note that for consumer goods shipped on the Asia-U.S. or Asia-Europe routes, ocean shipping costs alone typically account for 1.5% of shelf values (Rodrigue and Comtois, 2009). Faced with such cost increase, shippers may divert to other ports and the container port's output level will reduce. Thus, empirically, we want to test whether road congestion in the urban area near a port will reduce the port's container throughput, while increasing its rival's throughput.

The impacts of road capacity on container throughputs, however, depend on the mode of competition between ports. According to Proposition 2.1, when demand and congestion delay functions are all linear, under quantity competition, a port's throughput tends to increase when the road capacity nearby increases (i.e. $dX_i^* / dK_i > 0$) but decrease when the road capacity around the rival port increases (i.e. $dX_j^* / dK_i < 0$). Under price competition, however, the impacts of road capacity on ports' throughputs are in general ambiguous. Note that in such a case, port charges are strategic complements and thus the reaction functions are upward sloping (see Figure 3.1).

Figure 3.1 An illustration of the price competition case



As both K_i and K_j enter the reaction functions, both ports' reaction functions will shift if K_i changes. If demand functions and congestion delay functions are all linear, an increase in K_i will lead to the reaction function of port *i* shifting outward and that of port *j* shifting inward. De Borger *et al.* (2008) proved that at the equilibrium an increase in K_i will reduce port *j*'s charge while raising port *i*'s charge if port *i*'s output level is low.[32] The impact on

[32] De Borger *et al.* (2008) argued that it is quite plausible to have port *i*'s charge increase in road capacity K_i.

port *i*'s equilibrium throughput is given by

$$\frac{dX_i^*}{dK_i} = \frac{\partial X_i}{\partial p_i}\frac{dp_i^*}{dK_i} + \frac{\partial X_i}{\partial p_j}\frac{dp_j^*}{\partial K_i} + \frac{\partial X_i}{\partial K_i}. \tag{3.1}$$

It can be shown that the second term on the RHS of equation (3.1) is negative, while the third term, the direct impact of capacity expansion on throughput, is positive. If the capacity increase leads to an increase in port charge, then the first term will be negative as well.[33] Accordingly, the overall impact of an increase in road capacity on the port's throughput is undetermined. The impact on the rival's equilibrium throughput is given by

$$\frac{dX_j^*}{dK_i} = \frac{\partial X_j}{\partial p_i}\frac{dp_i^*}{dK_i} + \frac{\partial X_j}{\partial p_j}\frac{dp_j^*}{\partial K_i} + \frac{\partial X_j}{\partial K_i}. \tag{3.2}$$

All the three terms on the RHS of equation (3.2) have signs opposite to those in equation (3.1). Again, an increase in port *i*'s road capacity has ambiguous impact on its rival port's container throughput.

Although the impact of road capacity change on throughput is analytically ambiguous under a price competition model, the quantity competition setting suggests that road capacity expansion is likely to benefit the port nearby and harm its rival. On the other hand, previous surveys (Golob and Regan, 2000; Maloni and Jackson, 2005) found that both commercial trucking firms and port managers believe that increasing road capacity is an effective solution to road congestion, which, in turn, facilitates port development. Given these survey findings and the ambiguous results in analytical studies, we will empirically investigate whether an increase in road capacity in an urban area will have a positive impact on the container throughput of the port nearby and a negative impact on that of the rival port.

---

[33] De Borger *et al.* (2008) showed that $\partial X_i / \partial p_i < 0$, $\partial X_i / \partial p_j > 0$, $\partial X_i / \partial K_i > 0$, and $\partial X_j / \partial K_i < 0$.

3.2.2 Road congestion and port productivity

The theory about the relationship between port efficiency and road congestion has yet been formally established. Nonetheless, since productivity is usually defined as the ratio of outputs to inputs, it is natural to decompose the impact of road congestion into two aspects: the influences on container port outputs and inputs, respectively.

The direct impact of road congestion on port output levels has been partly discussed in Section 3.2.1, based on the argument that shippers consider the generalized cost along the intermodal transport chain as the basis of making port choices. Demand faced by a container port is also related to, or in many cases is derived from, economic activities in manufacturing, trading and logistics industries located in the broadened economic region surrounding the port. Those industries require not only frequent access to marine transportation services but also efficient connections among landside facilities, such as factories, warehouses, distribution centres, outlets, rail terminals and other nodes in a supply chain. Even if access to container ports is not a concern, the latter will be adversely affected by urban road congestion.[34] Basically, road congestion hinders effective adoption of just-in-time logistics, a trending practice of today's manufacturing and distribution business. Delays and uncertainty due to road congestion drive up inventory level, employee travel time, delivery cost, etc. All provide incentives to relocate distribution and production facilities, substantially reducing the customer base of the container port nearby and leading to a detrimental impact on the port's outputs. Given that investment in inputs of container ports, such as infrastructures and land, are lumpy and irreversible, any form of output reduction caused by road congestion might bring about a decrease in port productivity.

Road congestion affects the input side of port productivity mainly from the lengthened container dwell time in the port. Time waiting in marine terminal stacks for truck pickup and vessel loading accounts for, respectively, 19.6% and 24.7% of the total time spent on moving a 40-foot container between the North American East Coast and Western Europe (Rodrigue,

[34] The adverse impact of road congestion on the productivity of supply chain and truck-dependent businesses has been widely identified (for a brief literature review, see Weisbrod and Fitzroy, 2011).

2006) and can be a significant source of inefficiency in container port operation. According to Rodrigue and Notteboom (2009), there is a tendency for container ports to reduce the free dwell time and raise charges for staying longer in deep-sea terminals so as to maximize the throughput. High level of road congestion, recurring daily or not, increases the amount of time that trucks have to spend before accessing the port and raises container dwell time in marine terminals. When road congestion is expected or recurring, behaviourally, some truck drivers may depart much earlier than necessary and arrive at the port gate before their designated pickup time. This phenomenon is predicted by the bottleneck model which has been used to describe the commuter behaviour during the morning rush hours (e.g. Arnott et al., 1990) and can be applied here as well. Similar phenomenon has been observed in the port industry as well. In January 2012, among the 4,551 trucks showing at the gate of Port Botany in Sydney, Australia, 3,578 of them were early arrivals (Sydney Ports, 2012). The Port of Melbourne discovered that trucks which arrived at the port earlier than the booked timeslots caused long queues along roads in and around the port area (Port of Melbourne, 2005). In such cases, those early arrivals might block the trucks arriving on time and amplify container dwell times in the port. Road congestion delays affect the case of dropping off containers for export as well. To prevent missing the vessel schedule, trucks may depart for port earlier, lengthening the stay of containers in terminal stack before being loaded onto the vessels. If a container misses the vessel, it will be deferred and assigned to another vessel, leading to extra dwell time as well as additional handling and paper works, i.e. more resources have to be consumed to deal with a late container.

In summary, road congestion tends to increase container dwell time, extend occupation of storage areas and consume extra resources, which in turn reduces the terminal capacity and impairs the utilization of port equipments, labours as well as land, a highly constrained resource for most of the seaports. Lack of storage space in terminals may also keep the containerships idle and adversely affect the loading/unloading rate. Alternatively, to keep up with the desired throughput level, container terminals may have to increase their sizes and make extra investment in other relevant storage resources to accommodate more containers at the same time, which may further reduce port infrastructure productivity. As road transportation is one stage of the intermodal shipping process, road congestion exacerbates the uncertainty of container pickup or drop-off times and hence amplifies the operational variability at the port. According to the tradeoffs heuristic between capacity utilization, inventory and variability or the so-called OM triangle (Lovejoy, 1998; Schmidt, 2005), facing with increased level of uncertainty, to maintain the same level of throughput, container ports may either invest more in capacity or increase storage space. Both require an increase in resource inputs at the container port, causing lower port productivity.

Based on the above arguments from both the output and input sides, our hypothesis is that an increase in road congestion will reduce a container port's infrastructure productivity.

3.3 Methodology and variables

We use ordinary least square (OLS) and two-stage least square (2SLS) regression analysis to address the first two research questions introduced in Section 3.1 and examine the relationship between road congestion, road capacity and container port throughputs. To answer the third question, i.e. how road congestion affects port efficiency, we take a standard two-stage approach. In the first stage, the container port productivity is measured by DEA. Then, Tobit regression analysis is undertaken to explore the relationship between the DEA efficiency scores and the hinterland connectedness, such as provision of rail services and road congestion around the ports. Our estimation is based on a panel dataset that includes 13 major container ports in the U.S. from 1982 to 2009.

3.3.1 Measure port efficiency with DEA

We measure the infrastructure efficiency among the container ports sampled by calculating DEA scores. In the DEA model, we use container throughput (including containers moved by all intermodal modes, such as barge, truck and rail) as the output and consider container terminal size, total length of berths and total number of cranes and gantries[35] at container

[35] As lifting capabilitiy varies across cranes and gantries, the number of cranes and gantries is not a perfect measure of inputs but this is commonly used and accepted in the literature due to the lack of detailed information on each crane or gantry as well as the lack of a comprehensive and precise measure for the overall input of a group of cranes and gantries.

terminals as the inputs. As investment in port infrastructure is lumpy and port expansion projects usually take several years to complete, it is common that the amounts of these inputs may be constant over many years before or after a sudden addition of port capacity. Despite the low variation in inputs, container throughput changes rapidly over the years. Thus, output-orientated DEA is more appropriate. DEA scores are calculated by assuming constant return to scale (CCR model) as well as variable return to scale (BCC model).[36]

Two approaches are widely used to measure productivity with DEA for panel data. The cross-sectional approach constructs one production frontier for each period and technical efficiency scores are calculated based on the corresponding frontier of each time period. This approach enables comparison of relative technical efficiency among ports in the same period while removing the impact of technology progress. The pooled-time-period approach treats a port sampled in different years as different decision making units (DMUs) and fits a single efficiency frontier by pooling all the time periods and ports together. The second approach is applied here for a few reasons. First, it is consistent with Turner *et al.* (2004) so that our results can be compared with the earlier study. Second, the frontier estimated by DEA is very sensitive to outliers especially when the number of DMUs is small. This is exactly the case in this study. Finally, the cross-sectional approach fails to account for efficiency change over time and random effects. To deal with this problem DEA window analysis has been proposed (Charnes *et al.*, 1985). The cross-sectional approach and the pooled-time-period approach are two extremes of window analysis. Cullinane *et al.* (2004, 2005) implemented window analysis in the container port context and argued for the importance to reflect ports' ability to catch up with both the latest technology and the best-practice techniques over time in the DEA scores. Thus, the pooled-time-period approach is a better choice if the long-run efficiency is in concern. In addition to the above rationales, to check the robustness of our results, we replicate our analysis with the cross-sectional approach and obtain qualitatively

[36] The CCR model is based on the work of Charnes *et al.* (1978, 1981), whereas the BCC model is based on Banker *et al.* (1984). Ports with high DEA scores can have high congestion level due to huge throughput volume. Thus, one may argue that quantity is only one dimension of a port's output while quality, such as congestion delay, should be considered as well. Although this paper does not incoperate congestion into efficiency calculation, it can be done by employing quality-adjusted DEA (see Sherman, *et al.*, 2006, and Lozano and Gutlerrez, 2011, for details).

similar results from the second stage regression analysis. Therefore, it is adequate to focus on the pooled-time-period approach and present only this part of the results.

3.3.2 Econometric specifications

To examine the relationship between urban road congestion and container throughput, we estimate the following model:

$$\ln X_{it} = \alpha_0 + \alpha_1 \ln T_t + \alpha_2 \ln P_{it} + \alpha_3 \ln D_{it} + \alpha_4 \ln DR_{it} + u_i + \varepsilon_{it}\,, \qquad (3.3)$$

where X_{it} is the truck-related container throughput for port *i* in year *t*, T_t is the value of U.S. international trade in year *t*, P_{it} is the population in the catchment area of port *i* in year *t*, D_{it} is the road congestion delay in the urban area around port *i* in year *t*, DR_{it} is the road congestion delay in the urban area of port *i*'s rival port in year *t*, u_i is the fixed effect for port *i*, and ε_{it} is the error term for port *i* in year *t*.

Note that we apply a log-log regression model in order to reduce the magnitude of heteroskedasticity which is likely to be an issue in cross-sectional data (Verbeek, 2005). Besides, the coefficients in (3.3) can subsequently be directly interpreted as elasticities. Container traffic which does not use trucks to move into or out of the seaport is not counted in the dependent variable. U.S. container ports use a number of inland transportation modes. Trucking is common for shipment delivery within 350 miles of the port of entry, while intermodal rail (on-dock and near/ off-dock) dominates the inland shipping market with a distance longer than 950 miles. Cargos moved through ports on the east coast use intermodal rail less frequently than those shipped through ports on the west coast, since the main destinations of consumption are relatively closer to the east coast. In addition, on-dock rail yards are widely available in west-coast terminals. The share of on-dock rail has been increasing since early 2000. In 2008, among all modes of transport, on-dock rail accounts for 23.7% of containers shipped through ports in the San Pedro Bay Area (The Port of Los Angeles, 2009). On the east coast, however, on-dock facilities are less commonly used: for example, even in NYNJ among the 5,529,908 TEUs handled in 2012, only 433,481 containers are moved by on-dock rail facilities (Port of New York and New Jersey, 2013). On the west coast, in order to exploit the cost savings from using the larger 53-foot domestic containers, a significant portion of imported containerized cargos need to be transloaded at distribution centers outside of the port before being shipped further inland by rail. Thus, except for on-dock rail and barge, all the other modes require truck drayage on urban roads surrounding the port and will be affected by urban road congestion.[37] Container movements of such modes are hence truck-related and should be counted into the dependent variable. Throughout this paper, except when measuring port productivity by calculating efficiency scores, container throughput refers to the truck-related part of the total traffic volume via the port, i.e. removing barge[38] and on-dock rail. Moreover, our models capture the fixed effect of each port, because seaports vary in their operations, geographical locations, availability of port-related services, capability of handling large vessels, etc. Consequently, some ports naturally have advantages over the other ports in terms of attracting container traffic.

Equation (3.3) includes two control variables, T_t and P_{it}. It is expected that α_1 is positive because maritime traffic is mainly derived from international trade, and, therefore, container throughput should be positively correlated to the international trade volume. In fact, the container throughputs for the selected U.S. ports have an increasing trend, as do other independent variables, since the economies and the shipping demand have a growing trend in general. Therefore, by including the value of U.S. international trade as the independent variable, we control for the increase in container throughput owing to the economic growth and the shipping demand increase. For a particular container port, part of the truck-related throughput is destined to, or originated from, the port's captive catchment area,[39] where the

[37] A consulting report for the port of Los Angeles and the port of Long Beach finds that an increase in truck transportation costs will affect containers shipped by off-dock rail, transloaded to rail or shipped by truck alone with distance more than 150 miles the most. Near-dock rail and trucking within the range of 50-150 miles will be mildly affected. The impact on short distance trucking (less than 50 miles) will be the least (Moffatt & Nichol and BST Associates, 2007).

[38] Containers may be barged from one port to the other and then trucked to the final destination. However, the scale of barge service is very limited due to longer transit time. In 2012 only 6,227 containers accessed Port of Hampton Roads via inland waterway barge and the frequency of such service is also quite low. Intracoastal short-sea shipping is also limited in the U.S. Thus, ignoring this mode will not significantly affect our results.

[39] Here, catchment area is different from urban area around the port. Urban area around the port can be considered as the metropolitan area the port belongs to. Usually, a port's catchment area is much larger than the

port has a locational advantage over any other port and hence inter-port competition over this part of the container traffic is unlikely to occur. We control for the impact of captive catchment area by including catchment area population as one of the independent variables, because population affects the consumption of imported goods and the capacity of producing export goods and, as a result, ports with larger populations in their own catchment areas are likely to have higher captive container throughputs. Thus, α_2 is expected to be positive, as well. Once the population in the captive catchment area is controlled for, the coefficients of α_3 and α_4 will reflect the relationship between the variation in urban road congestion and the variation in the part of throughput which is unaffected by the consumption or production in the captive catchment area.

As mentioned in Section 3.2.1, an increase in container throughput will increase the severity of road congestion around the port, while congested roads may redirect shippers to the rival port. Thus, we may have an endogeneity problem: D_{it} is likely to be correlated with the error term and will lead to inconsistent estimation. We deal with this problem through instrumental variables (IVs): the urban population (pop_{it}) and area ($area_{it}$), or the urban population density ($dens_{it}$) which is a ratio of pop_{it} and $area_{it}$. Roads in urban areas which accommodate larger populations or population densities are likely to be more congested. Besides, people who live in larger urban areas may need to travel longer distances, thereby creating more traffic on the roads. Thus, road congestion should be correlated with those instrumental variables. Conversely, because contemporary container ports serve destinations deep into the hinterland instead of being confined to the immediate urban area around the ports, such factors as the population density of the associated urban area is less likely to affect container throughput directly. Therefore, the abovementioned instrumental variables are legitimate. As long as the endogeneity problem is properly resolved, we expect α_3 to be negative and α_4 to be positive.

metropolitan area and may include metropolitan areas in vicinity states. For example, the urban area of the port of Oakland consists of the populated area within the San Francisco-Oakland metropolitan area while the catchment area extends to San Jose and even a few counties in Nevada. The exact definitions of urban area and catchment area for ports in our sample are given in Section 3.4.

One may argue that the rival's road congestion might be endogenous, but we think the impact of port output on the rival's road congestion should be insignificant. Container throughput might affect rival's road congestion in two ways. First, a growth in port throughtput may raise the chance of having containers diverted to the nearest rival port if the port is overloaded. Such diverted containers will increase the road congestion on the roads near the rival port. Second, when the port's output level increases, economies of scale will play a role by reducing handling cost and hence port charge; on the other hand, more logistics service providers and forwarders will be attracted to the port as well. Consequently, more shippers will be attracted from the rival port, reducing road congestion around the rival port. As an increase in throughput can have both negative and positive effect on the rival port's road congestion and these effects are all indirect, the endogeneity problem if any should be relatively mild. Thus, given a small sample size as we have in this study, we want to avoid further complicating the model and hence consider rival's road congestion as exogenous.

To examine the relationship between urban road capacity and container throughput, we estimate the following model:

$$\begin{aligned}\ln X_{it} &= \beta_0 + \beta_1 \ln T_t + \beta_2 \ln P_{it} + \beta_3 \ln LM_{it} + \beta_4 \ln LMR_{it} \\ &\quad + \beta_5 (\ln LM_{it})^2 + \beta_6 (\ln LMR_{it})^2 + \beta_7 (\ln LM_{it})(\ln LMR_{it}) + v_i + \eta_{it}\end{aligned}, \quad (3.4)$$

where LM_{it} is the road capacity in the urban area near port *i* in year *t*, LMR_{it} is the road capacity in the urban area near the port *i*'s rival in year *t*, v_i is the fixed effect for port *i*, and η_{it} is the error term for port *i* in year *t*.

According to chapter 2, the port's equilibrium throughput is a function of the port's own road capacity and its rival's road capacity, but the functional form depends on the structures of the demand functions and the congestion cost function. In general, the impacts of road capacity on throughput are non-linear and the road capacity elasticities are also not constant but rather are functions of capacities, as well. Without an explicit functional form, to capture the feature that the road capacity elasticities may depend on LM_{it} and LMR_{it}, we apply the second-order Taylor expansion around the point at which both $\ln LM_{it}$ and $\ln LMR_{it}$ are zero.

Consequently, the coefficients β_3 and β_4 only reflect the elasticities at the (0, 0) point, which does not have a practical meaning. Instead, we need to estimate the elasticities within the range of the observed points. The own-road capacity elasticity and rival-road capacity elasticity can be estimated by taking partial derivatives of $\ln LM_{it}$ and $\ln LMR_{it}$ respectively. That is,

$$\frac{\partial \ln X_{it}}{\partial \ln LM_{it}} = \beta_3 + 2\beta_5 \ln LM_{it} + \beta_7 \ln LMR_{it} \text{, and}$$
$$\frac{\partial \ln X_{it}}{\partial \ln LMR_{it}} = \beta_4 + 2\beta_6 \ln LMR_{it} + \beta_7 \ln LM_{it} \,. \tag{3.5}$$

In the theoretical models, it is assumed that road capacity affects ports' throughputs only through the change of road congestion. However, in reality, road capacity expansion may affect the ports' competition through other channels. For example, holding the road congestion unchanged, an increase in urban road capacity may imply an improvement in the accessibility to the hinterland by, for instance, having shorter routes and thus attracting more shippers. On the other hand, an increase in urban road supply may suggest an improvement in the connectivity to the rival port nearby, consequently intensifying the port competition. Furthermore, an investment in urban road capacity may be accompanied by an investment in other transport infrastructure in and outside of the urban area, reducing the cost of using other transportation modes, such as air, rail and long-distance trucking. As a result, local firms may change their sourcing destinations, as well as transportation modes, and these factors may eventually affect the demands of maritime shipping. Therefore, it is of interest to distinguish the effect of road capacity investment through road congestion from other channels. In view of this, we consider the model that controls for the road congestion while estimating the impact of road capacity investment through other factors:

$$\begin{aligned}\ln X_{it} = {} & \delta_0 + \delta_1 \ln T_t + \delta_2 \ln P_{it} + \delta_3 \ln D_{it} + \delta_4 \ln DR_{it} + \delta_5 \ln LM_{it} + \delta_6 \ln LMR_{it} \\ & + \delta_7 (\ln LM_{it})^2 + \delta_8 (\ln LMR_{it})^2 + \delta_9 (\ln LM_{it})(\ln LMR_{it}) + \omega_i + \phi_{it}\end{aligned} \,, \tag{3.6}$$

where ω_i is the fixed effect for port *i*, and ϕ_{it} is the error term for port *i* in year *t*.

To examine the correlation of road congestion with port efficiency, we apply Tobit regression (Tobin, 1958) and use DEA scores as the dependent variable. Tobit regression is widely used to investigate the relationship between DEA scores and exogenous explanatory variables.[40] The reason of using Tobit regression is that DEA scores is confined in the (0, 1] interval and thus OLS regression is no longer appropriate. Tobit regression is appropriate when the dependent variable of some observations is censored or is a corner solution of an optimization problem. DEA scores belong to the second category, as they are generated by sovling a series of linear programming problems. Each linear programming problem looks for the maximum efficiency of a decision making unit (DMU) subject to the constraint that the efficiency of every sampled DMU is between zero and unity. According to Hoff (2007), when DEA estimates a DMU's efficiency score to be unity, this is a cornor solution and occurs by construction of the DEA score generating process. That is, if we can observe the true efficiencies, we should not observe a group of DMUs which are scored exactly at unity. Those DMUs with unity scores are considered to be more efficient than the rest because they are located on the production frontier estimated by DEA, but some of these "efficient" DMUs may be more efficient than the other DMUs on the frontier in spite of having the same unity DEA scores. Thus, for DMUs with unity DEA scores, we are not able to observe the latent (true) efficiency scores. Tobit regression instead assumes truncated normal distribution for the dependent variable to match the fact that if the true efficiecy score is larger or equal to unity we will only observe unity from DEA. The regression model is specified as follows:

$$\begin{aligned} E_{it}^{*} &= \gamma_0 + \gamma_1 \ln P_{it} + \gamma_2 \ln D_{it} \\ &\quad + \gamma_3 OP_{it} + \gamma_4 NCLASS1_{it} + \gamma_5 ONDOCK_{it} + \gamma_6 LARGE_i + \xi_{it} \end{aligned}, \tag{3.7}$$

where

$$E_{it} = \begin{cases} E_{it}^{*} & if \;\; E_{it}^{*} < 1 \\ 1 & if \;\; E_{it}^{*} \geq 1 \end{cases}.$$

[40] The appropriateness of Tobit regression for DEA scores has been challenged by some recent papers (e.g. McDonald, 2009; Simar and Wilson, 2011), but the issue is still in debate. As only five observations in our sample are censored, we find that the results from OLS are similar to those from Tobit regression.

In equation (3.7), E_{it} stands for the observed DEA efficiency score for port *i* in period *t*, while E^{*}_{it} is the unobservable (i.e. latent) efficiency for port *i* in year *t*. ξ_{it} is the error term for port *i* in year *t*. According to the discussion in Section 3.2.2, we expect γ_2 to be negative if road congestion adversely affects port efficiency. OP_{it} measures intra-port competition. $NCLASS1_{it}$ measures the level of rail services. $ONDOCK_{it}$ is a dummy variable indicating the provision of on-dock rail facility. $LARGE_i$ is a dummy variable indicating the large container ports.

3.4 Data and variable construction

The data set consists of 13 U.S. ports, and the observations are taken from 1982 to 2009 annually.[41] Six of the 13 ports are on the west coast, namely, Seattle, Tacoma, Portland, Oakland, Los Angeles and Long Beach. The other seven ports are on the east coast: Boston, New York-New Jersey (NYNJ), Baltimore, Hampton Roads, Charleston, Jacksonville and Miami.[42] In answering the first two questions, we conceptually aggregate ports located in neighboring cities as one port. For example, the Port of Los Angeles and the Port of Long Beach are considered to be one port, LALB, despite their separate managements, because these two ports are located close to each other such that the traffic from/to the two ports probably uses the same road system. Thus, it is inappropriate to treat the two ports as rivals instead of a single port, and as in the present study, we look at the impacts of road system on their throughputs. On the other hand, cities in which the two ports are located are conventionally pooled into the same urban area in the datasets we used for urban road congestion and road capacity. For the same reasons, we also combine the Port of Seattle and the Port of Tacoma as a single port, SeaTac. As a result, we use 11 *de facto* ports for the first

[41] By looking at the U.S. ports, we can focus on the interaction between the port throughput and the road system. As the Jones Act limits feeder services between U.S. ports, U.S. ports have little transshipment and have a 'strong inland orientation' (Notteboom and Rodrigue, 2005), and thus their throughputs are more likely to be affected by port access conditions.
[42] West-coast U.S. ports might compete with Canadian west-coast ports, namely, Vancouver and Prince Rupert. East-coast U.S. ports might also compete with their counterparts in Canada, e.g. Montreal.

two research questions (Table 3.1).[43]

Table 3.1 Ports studied and their respective rival

Port	Container throughput in 2009 (in 1,000 TEU's)	Average market share[a] 1982-2009 (%)	Rival port	Distance to the rival port (miles)
Boston	187	1.64	NYNJ	190
NYNJ	4,561	29.86	Baltimore	170
Baltimore	525	6.38	NYNJ, Hampton Roads	170
Hampton Roads	1,745	10.02	Baltimore	170
Charleston	1,181	10.43	Jacksonville	197
Jacksonville	754	4.38	Charleston	197
Miami	807	5.78	Jacksonville	326
SeaTac	3,130	21.60	Portland	145
Portland	96	1.99	SeaTac	145
Oakland	2,050	13.45	LALB	344
LALB	11,816	52.77	Oakland	344

Notes: a. West-coast port market share: $S_i = X_i \big/ \sum_j X_j \quad j \in \{U.S.\ Pacific\ coast\ ports\}$; East-coast port market share: $S_i = X_i \big/ \sum_j X_j \quad j \in \{U.S.\ Atlantic\ coast\ ports\}$

For each port, we consider its rival port being the one in our sample with the shortest distance to that particular port. Competition between ports through the rivalry between alternate intermodal chains fully or partially served by commercial trucks is most likely to be affected by road congestion and relevant to the context of this study. Ports on the same coast are likely to compete in the common inner-hinterland, by connecting to intermodal rail services. For example, ports on the U.S. west coast compete intensively for inbound containers destined to the Midwest states. However, an overlap of hinterlands served by trucks alone is most likely to occur between ports located closest to each other, because trucks are often used to transport short-to-medium-haul cargos. Thus, we consider the closest port as the most influential rival and its road condition as one of the independent variables. On the other hand, Luo and Grigalunas (2003) argued that competition between the west-coast ports and the east-coast ports for container traffic might be present, due to U.S. intercontinental railways. In effect, the west coast and the east coast are probably competing for the America-bound Asian cargos destined for coastal and inland regions in the east of the continent and lower

[43] When calculating DEA scores and addressing the third research question, we keep Los Angeles, Long Beach, Seattle and Tacoma as independent ports, since their DEA scores significantly differ from their nearby neighbors'. Meanwhile, due to data availability, the port of Miami is excluded from calculating DEA scores and the observation period is shortened to 2000-2009.

Ohio Valley (Fan *et al.*, 2011). These cargos arrive at the east coast either directly through the all-water route via the Panama Canal or by being transited by double-stack trains from the west coast. The all-water route takes about eight more days to reach the Midwest states than the west coast intermodal option. Urban road conditions around the west coast ports are unlikely to change the eight-day time difference too much. Thus, even if urban road conditions around the west coast ports are improved, cargos which choose the east coast ports are less likely to trade off the reliability and cost savings of the all-water route against the limited time savings. Moreover, west coast ports have advantage over time-sensitive shipments and containers imported from North-east Asia, while east coast ports have overwhelming cost advantage over containers from south-west Asia, such as India. Thus, competition between ports of the two coasts, if any, is less relevant in our discussion.

The annual container throughput volumes (in 1,000 TEUs), aggregated across all inland shipping modes, are taken from the American Association of Port Authorities. As the present study investigates the impact of road systems on container throughput, the throughput data should be adjusted by subtracting the amount transported into and out of the interior regions directly by trains, i.e. on-dock rail, and/or inland waterway. However, as the modal split ratios for inland transportation at each seaport are not available, we approximate the truck-related throughput using the data provided by the *U.S. Commodity Flow Surveys*. These surveys report the tons of cargoes shipped into and out of a certain urban area by transportation modes. Based on this information, we calculate the fraction of cargoes shipped by trucks.[44] We multiply this fraction with the container throughput to estimate the number of TEU's moved by trucks to enter or exit the port, which is the dependent variable (X_{it}) in the econometric models.

The three input variables for DEA score calculation are obtained from various issues of *Containerization International Yearbook* (CIY, 2001~2010). We sum up the sizes of

[44] Unfortunately, this series of surveys is only conducted once every 4-5 years, and we have data for 1993, 1997, 2002 and 2007, but our dataset has observations from 1982 to 2009. Therefore, we assign the fraction calculated from the 1993 survey to all periods in 1982-1993, the 1997 fraction to all periods in 1994-1997, the 2002 fraction to all periods in 1998-2002 and the 2007 fraction to the remaining years.

container terminals within a port as the measure of terminal size in DEA. The numbers of ship-shore gantries, yard gantries and quay cranes are summed together to form another input variable, total number of gantries and cranes. The total length of berths in container terminals within a port is considered as the third input variable, total berth length.

The data for the value of U.S. annual international goods trade (T_t) are from the U.S. Census Bureau. To decide the population in the catchment area (P_{it}), we first need to decide for each port the relevant catchment area. As the shortest distance between the seaports in the sample is 145 miles, each port has overwhelming advantage over containers of which inland shipping distance is shorter than 50 miles. Thus, we consider the area within the 50-mile radius of the port to be the port's captive catchment area. Populations in counties within the respective radius of the port are summed up as an estimate of the population in the catchment area. The county level population data are obtained from the U.S. Census Bureau. To check the robustness of our models, we also consider areas within the 250-mile radius as catchment areas. On the one hand, the 250-mile catchment areas do overlap to some extent, but they have large portions which do not overlapped with rivals' catchment areas. On the other hand, based on the data provided by the *U.S. Commodity Flow Surveys*, the majority of trucking traffic is within 250 miles, accounting for 95% and 76% of total trucking traffic in terms of weight and value respectively. This suggests that the maximum economic trucking distance is roughly 250 miles and this distance defines the farthest reach of captive hinterland.

The *Annual Urban Mobility Report 2010* prepared by Texas Transportation Institute (TTI) provides urban road congestion delay measurements for major urban areas. An urban area is defined as the populated area with population density over 1,000 persons per square mile within a metropolitan region. This report records road congestion measures aggregated across all the freeways and arterial streets in the urban area. Although data that only include roads heavily used by container trucks fit to our research question better, these data are, unfortunately, not available.[45]

[45] A number of congestion measures are available in the report. We choose delay per peak traveler (DPPT) and

Although road capacity can be theoretically defined as the maximum traffic volume that can pass through per unit of time, such exact measurements are not available. Rather, this paper uses lane-miles to measure the amount of road available in an urban area and, therefore, approximate road capacity. This is a common approach in the literature, which estimates the road traffic demand induced by adding road capacity (e.g. Downs, 1962; Goodwin, 1996; Cervero, 2002; Duranton and Turner, 2009). The *Annual Urban Mobility Report 2010* reports lane-miles of freeways and arterial streets. This report also provides urban area population and square-miles, which are the instrumental variables, pop_{it} and $area_{it}$, for the estimation of equations (3.3) and (3.6).

The more terminal operators serve the same port, the more intensive competition among these operators might be. Thus, we measure the level of intra-port competition (OP_{it}) by the number of container terminal operators in the port. However, the number of terminal operators is constrained by the size of the port and the demand. Smaller ports usually have only one or two container terminal(s) and hence the number of terminal operators at these ports tends to be small as well and may not be a good indicator of low competition intensity. Larger ports instead usually have a number of container terminals operated by a few terminal operators, but some terminals may be operated by the same terminal operator, leading to a lower level of competition. Therefore, we also use the ratio between the number of operators and the number of container terminals as another measure of intra-port competition for robust check. The names of terminal operators are provided in CIY.

As detailed data of rail services are not available, we use the number of Class I railroads serving the port, denoted as $NCLASS1_{it}$, as a proxy. In the United States, large railroads with operating revenues exceeding USD $250 million are classified as Class I railroads. Larger number of Class I railroads implies fiercer competition among rail carrier and hence higher service quality, larger rail service capacity and higher service frequency. Names of rail

road congestion index (RCI) when addressing the first two research questions and the third research question, respectively. RCI is a better approximate to road congestion than DPPT, but it was not available until we completed our studies for the first two research questions.

carriers serving the container terminals are in most of the cases available in CIY. However, rail carrier information is incomplete in CIY and so we use the Port Series Reports published by the US Army Corps of Engineers as a complementary source. Information about on-dock rail facility is available in CIY. As long as at least one of the container terminals in the port provides on-dock rail facility, the value of $ONDOCK_{it}$ is one; otherwise, it is zero.

Ports with annual container throughput over three million TEUs are considered to be a large port and the value of the dummy variable, $LARGE_i$, is one for those ports. In this sample, Long Beach, Los Angeles and NYNJ belong to this category. All the other ports are considered to be small ports.

3.5 Empirical results

3.5.1 The impacts of road congestion and expansion on port throughput

Tables 3.2-3.4 present, respectively, the major results of regression models in (3.3), (3.4) and (3.6) with the panel data described in Section 3.4. The coefficients for *ln*(*T*), *ln*(*P250*) and *ln*(*P50*) are the estimated elasticities of container throughput with respect to U.S. international trade value, 250-mile catchment area population and 50-mile catchment area population respectively. As expected, all of these coefficients are positive and statistically significant in all models. The international trade elasticity of container port throughput ranges from 0.50 to 1.22, and most of them are below 1, meaning that in the U.S., the percentage growth in container throughput is smaller than the percentage growth in international trade value. On the other hand, we found that a 1% increase in the population of the 250-mile catchment area is associated with a 0.42-2.41% growth in container throughput, while the elasticity on population in the 50-mile catchment area ranges from 0.53 to 1.16. In general, the estimated coefficients of the 250-mile catchment area population are larger than the coefficients of the 50-mile catchment area population.

Table 3.2 shows that the result for the relationship between urban road congestion and port throughput is consistent with our expectation. In models D1 and D2 (with large and small catchment areas, respectively), the OLS coefficients of own road congestion, *ln*(*D*), are negative and statistically significant, while the coefficients of the rival's road congestion, *ln*(*DR*), are positive but not statistically significant. However, after using the instrumental variables for *ln*(*D*) (i.e. the natural logarithms of urban population and urban area for models DIV1-2 and the natural logarithm of urban population density for models DIV3-4) and applying the two-stage least square method (2SLS), the signs of all the coefficients remain unchanged, which are all statistically significant, and their magnitudes increased. The Durbin-Wu-Hausman tests suggest that the endogeneity issue exists, and thus the instrumental variable approach should be adopted. In models DIV1-4, we observe that the impact of rival's road congestion is smaller than the impact of own road congestion. In particular, a 1% increase in own road congestion implies a reduction in container throughput by 0.90-2.48%, while a 1% increase in rival's road congestion implies an increase in container throughput by 0.62-1.69%. This finding is consistent with the theoretical assumptions for demand functions: when the road congestion around a certain port (and thus the full price of using this port) increases, only a portion of the shippers who choose not to use the port will switch to the rival port indicated in the sample because the rest of them will either divert to other rival ports farther away which are not indicated in this study or even choose not to ship the goods at all.

Table 3.2 Fixed effect OLS and 2SLS on congestion delays: equation (3.3)

	OLS		2SLS			
	D1	D2	DIV1	DIV2	DIV3	DIV4
ln(*T*)	0.5027** (0.0769)	0.5575** (0.0595)	0.5987** (0.0982)	0.7431** (0.0910)	0.8214** (0.1744)	0.7818** (0.1349)
ln(*P250*)	1.2616** (0.2123)		1.4377** (0.2507)		1.8463** (0.3999)	
ln(*P50*)		0.9116** (0.0925)		1.0430** (0.1030)		1.0703** (0.1277)
ln(*D*)	-0.2257* (0.1171)	-0.3236** (0.1302)	-0.9044** (0.3644)	-1.4276** (0.3642)	-2.4787** (0.9264)	-1.6575** (0.6815)
ln(*DR*)	0.1641 (0.1375)	0.1713 (0.6028)	0.6228** (0.2655)	0.9035** (0.2598)	1.6868** (0.6408)	1.0560** (0.4631)
Constant	-9.2987** (1.7150)	-4.3839** (0.6028)	-10.9512** (2.1445)	-5.4857** (0.7159)	-14.7839** (3.5085)	-5.7151** (0.9415)
N	306	306	306	306	306	306
F	159.38	171.23	261.41	281.51	138.24	252.04
Adjusted R^2	0.9292	0.9448	0.9222	0.9265	0.8526	0.9181
Instruments			*ln*(*area*) and *ln*(*pop*)		*ln*(*density*)	
			Durbin-Wu-Hausman chi-square test H0: *ln*(*D*) is exogenous			
Chi^2 (1)			4.6231	14.7312	13.4677	6.2290
p-value			0.0315	0.0001	0.0002	0.0126

Notes: * significant at α=0.10. ** significant at α=0.05. Values in the brackets are robust standard errors for OLS and standard errors for 2SLS. Fixed effect coefficients are omitted to save space.

Table 3.3 Fixed effect OLS on road capacities (lane-miles): equation (3.4)

	SO1	SO2
ln(*T*)	0.9434** (0.1021)	0.8605** (0.0889)
ln(*P250*)	0.4150* (0.2414)	
ln(*P50*)		0.5340** (0.1181)
ln(*LM*)	-5.7785** (2.1502)	-5.3648** (2.1587)
ln(*LMR*)	15.6863** (2.3505)	10.5065** (2.1940)
$[ln(LM)]^2$	0.6388** (0.1737)	0.6828** (0.1749)
$[ln(LMR)]^2$	-0.6043** (0.1968)	-0.2393 (0.1829)
ln(*LM*)×*ln*(*LMR*)	-0.7210** (0.2546)	-0.8421** (0.2200)
Constant	-38.5351** (9.3778)	-18.8594* (9.5978)
N	306	306
F	101.50	101.66
Adjusted R^2	0.9475	0.9522

Notes: * significant at α=0.10. ** significant at α=0.05. Values in the brackets are robust standard errors. Fixed effect coefficients are omitted to save space.

The impacts of urban road capacity expansion are estimated by the fixed effect OLS on lane-miles (models SO1 and SO2 with large and small catchment areas respectively in Table 3.3). Note that because equation (3.4) is a Taylor expansion around the point where both *ln*(*LM*) and *ln*(*LMR*) are zeros, which is out of our sample range, the coefficients of these two variables do not have practical meanings.[46] However, by estimating equation (3.4), we can achieve a better understanding of the higher-order impacts of road capacity changes. In fact, almost all of the second-order terms are statistically significant, suggesting that higher order effects should not be ignored. Recall equation (3.5), i.e. our definitions of road capacity elasticities of container throughput. The estimation of β_5 is positive, although that of β_7 is negative, implying that the own-road capacity elasticity increases in own road capacity while decreasing in its rival's road capacity. As a result, ports with much larger road capacity relative to their rivals are more likely to benefit from road capacity expansion. Regarding the impact of the rival's road capacity expansion, because the estimation of β_6 is negative, the rival-road capacity elasticity decreases in both own road capacity and rival's road capacity. Therefore, ports are likely to be harmed by their rivals' road investments, unless the road capacities of both competing ports are low.

[46] We also fit the pure first-order models by removing all the second-order terms in equation (3.4). Assuming there is no higher order effects, the coefficients of *ln*(*LM*) and *ln*(*LMR*) can be interpreted as, respectively, the own-road capacity and the rival-road capacity elasticities. We find that the coefficient of *ln*(*LM*) is negative and statistically significant, while the coefficient of *ln*(*LMR*) is negative but not statistically significant.

Table 3.4 Fixed effect OLS and 2SLS on congestion delays and road capacities (lane-miles): equation (3.6)

	SOD1a	SOD2a	SOD1b	SOD2b	SOD1c	SOD2c
ln(*T*)	0.9383** (0.0985)	0.8735** (0.0819)	0.9503** (0.0961)	0.9442** (0.0917)	1.0971** (0.1392)	0.9914** (0.1152)
ln(*P250*)	0.5168** (0.2476)		0.5514** (0.2393)		0.9751** (0.3615)	
ln(*P50*)		0.5847** (0.1203)		0.6663** (0.1076)		0.7209** (0.1346)
ln(*D*)	-0.2571** (0.1158)	-0.3311** (0.1103)	-0.3712 (0.3283)	-0.9027** (0.3311)	-1.7681** (0.7837)	-1.2845** (0.6023)
ln(*DR*)	0.2987** (0.1450)	0.3466** (0.1409)	0.3793 (0.2493)	0.7464** (0.2485)	1.3667** (0.5656)	1.0136** (0.4322)
ln(*LM*)	-6.3443** (2.1013)	-6.1414** (2.0873)	-6.5114** (1.9259)	-7.1025** (1.9003)	-8.5590** (2.6011)	-7.7445** (2.1947)
ln(*LMR*)	15.5709** (2.4403)	10.2277** (2.2758)	15.6803** (1.9312)	10.3127** (2.1166)	17.0200** (2.5001)	10.3696** (2.2694)
$[ln(LM)]^2$	0.6852** (0.1761)	0.7487** (0.1704)	0.6759** (0.1854)	0.7376** (0.1728)	0.5620** (0.2383)	0.7302** (0.1854)
$[ln(LMR)]^2$	-0.5615** (0.2174)	-0.1745 (0.2022)	-0.5782** (0.1969)	-0.2036 (0.2007)	-0.7831** (0.2651)	-0.2231 (0.2165)
$ln(LM) \times ln(LMR)$	-0.7802** (0.2818)	-0.9193** (0.2449)	-0.7499** (0.3026)	-0.8220** (0.2832)	-0.3794 (0.4181)	-0.7571** (0.3146)
Constant	-36.0600** (9.9426)	-14.3861 (9.9384)	-36.4781** (8.2794)	-11.6297 (9.0437)	-36.0890**(10.3630)	-9.3794 (10.1070)
N	306	306	306	306	306	306
F	87.58	88.54	294.33	303.34	188.16	264.04
Adjusted R^2	0.9483	0.9536	0.9481	0.9494	0.9188	0.9419
Instruments			*ln*(*area*) and *ln*(*pop*)		*ln*(*density*)	
			Durbin-Wu-Hausman chi-square test H0: *ln*(*D*) is exogenous			
Chi^2 (1)			0.1490	3.9748	6.4815	3.5113
p-value			0.6995	0.0462	0.0109	0.0610

Notes: * significant at α=0.10. ** significant at α=0.05. Values in the brackets are robust standard errors for OLS and standard errors for 2SLS. Fixed effect coefficients are omitted to save space.

In addition to the general observations discussed above, based on the results stated in Table 3.3, we further estimate the road capacity elasticities of throughput for each port and each year covered in our sample and analyze the behavior of individual ports. Except NYNJ, Miami and LALB, each port in our sample has negative own-road capacity (LM) elasticity on average (Table 3.5). However, in general, LM elasticity has a slightly increasing trend over time (Figures 3.2 and 3.3). The rival-road capacity (LMR) elasticity of throughput is likely to be negative, as well, except those of Charleston and Jacksonville. We observe a clear downward sloping pattern for each port's LMR elasticity over time (Figures 3.4 and 3.5). Therefore, over the past three decades, own road capacity expansion tended to be more beneficial to or less harmful, in terms of throughput levels, for the affiliated port, while rival's road capacity expansion tended to become more harmful.

We further estimated the regression model which includes road congestions as explanatory variables in equation (3.6). Table 3.4 shows the results for models SOD1-2a without IVs and models SOD1-2b and SOD1-2c with IVs as in Table 3.2. We find that a port's container throughput is negatively correlated with its own road congestion while positively correlated with its rival's road congestion. This finding is consistent with our findings from estimating equation (3.5). Moreover, after controlling for road congestion, the relationship between container throughput and road capacities persists in two respects. First, the signs of the coefficients of road capacity-related variables remain the same (Table 3.4). Second, more importantly, the previously mentioned patterns of estimated road capacity elasticities over time do not change, either (Table 3.5).

Table 3.5 The average elasticities estimated by equations (3.4) and (3.6)

Ports	Boston		NYNJ		Baltimore		Hampton Roads		Charleston		Jacksonville		Miami	
Models	LM elasticity	LMR elasticity	LM elasticity	LMR elasticity	LM elasticity	LMR elasticity	LM elasticity	LMR elasticity	LM elasticity	LMR elasticity	LM elasticity	LMR elasticity	LM elasticity	LMR elasticity
Eq.(5)														
SO1	-1.67	-2.93	1.06	-1.72	-1.94	-1.87	-1.43	-0.31	-2.46	1.52	-1.02	1.87	0.19	0.03
SO2	-1.69	-1.82	1.35	-2.00	-1.94	-1.11	-1.32	-0.36	-2.36	1.06	-0.78	0.80	0.47	-0.68
Eq.(7)														
SOD1a	-2.01	-2.71	0.94	-1.71	-2.29	-1.67	-1.74	-0.19	-2.84	1.65	-1.29	1.90	0.00	0.03
SOD2a	-2.07	-1.48	1.26	-1.97	-2.34	-0.80	-1.67	-0.18	-2.82	1.24	-1.08	0.83	0.29	-0.67
SOD1b	-2.03	-2.67	0.84	-1.58	-2.32	-1.62	-1.80	-0.11	-2.90	1.71	-1.39	2.01	-0.10	0.16
SOD2b	-2.25	-1.12	0.89	-1.39	-2.57	-0.46	-1.99	0.21	-3.19	1.55	-1.54	1.25	-0.13	-0.16
SOD1c	-2.37	-2.16	-0.40	0.06	-2.75	-1.08	-2.60	0.80	-3.70	2.49	-2.62	3.34	-1.37	1.70
SOD2c	-2.37	-0.87	0.64	-1.01	-2.71	-0.23	-2.21	0.47	-3.44	1.76	-1.85	1.54	-0.41	0.19

Ports	SeaTac		Portland		Oakland		LALB	
Models	LM elasticity	LMR elasticity	LM elasticity	LMR elasticity	LM elasticity	LMR elasticity	LM elasticity	LMR elasticity
Eq.(5)								
SO1	-0.57	-0.08	-1.81	-0.39	-1.73	-2.88	0.71	-2.29
SO2	-0.36	-0.50	-1.73	-0.28	-1.76	-1.78	0.94	-2.22
Eq.(7)								
SOD1a	-0.81	-0.02	-2.14	-0.24	-2.07	-2.66	0.56	-2.24
SOD2a	-0.62	-0.41	-2.12	-0.06	-2.14	-1.43	0.81	-2.13
SOD1b	-0.89	0.08	-2.20	-0.17	-2.10	-2.62	0.47	-2.12
SOD2b	-1.00	0.04	-2.43	0.30	-2.32	-1.07	0.48	-1.58
SOD1c	-1.95	1.34	-2.88	0.60	-2.43	-2.13	-0.59	-0.68
SOD2c	-1.25	0.35	-2.63	0.54	-2.44	-0.83	0.27	-1.21

Figure 3.2 Estimated LM elasticity of throughput (250-mile catchment area)

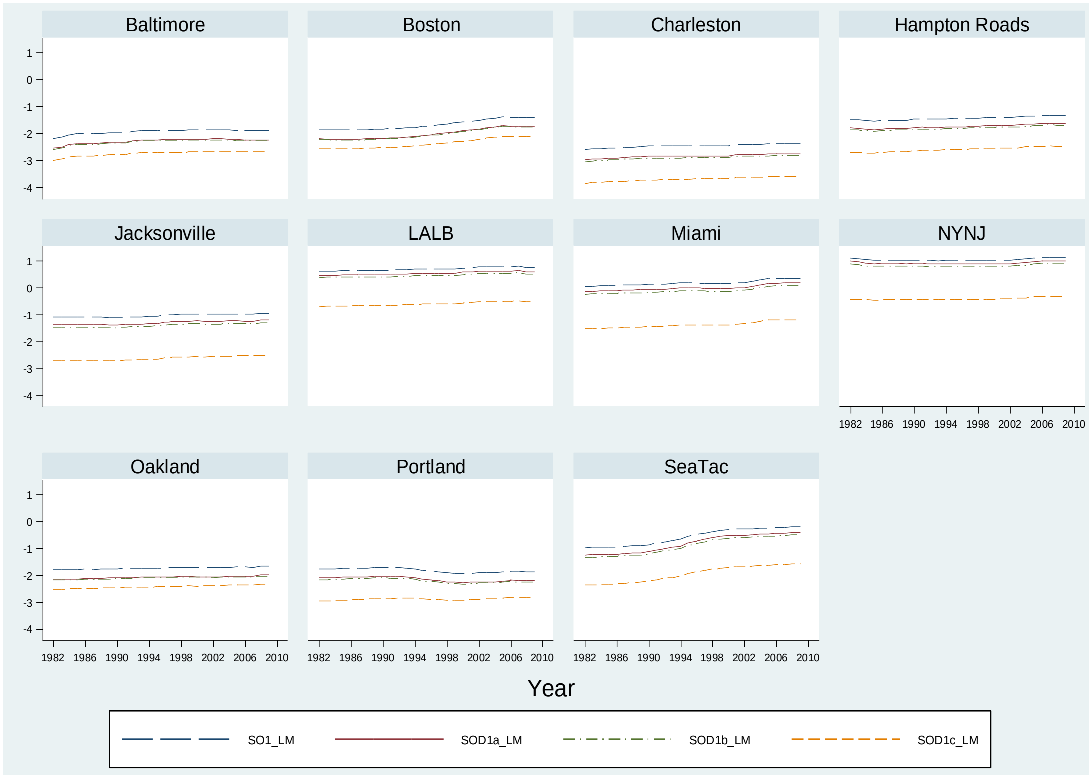


Figure 3.3 Estimated LM elasticity of throughput (50-mile catchment area)

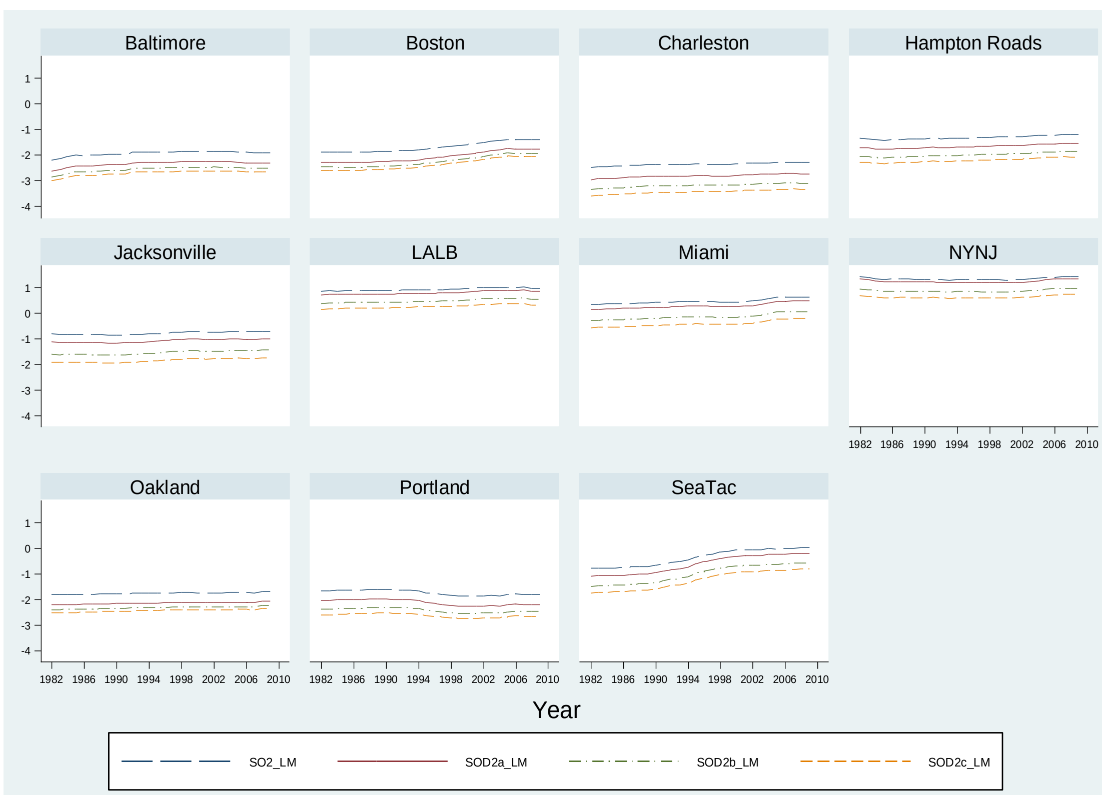


Figure 3.4 Estimated LMR elasticity of throughput (250-mile catchment area)

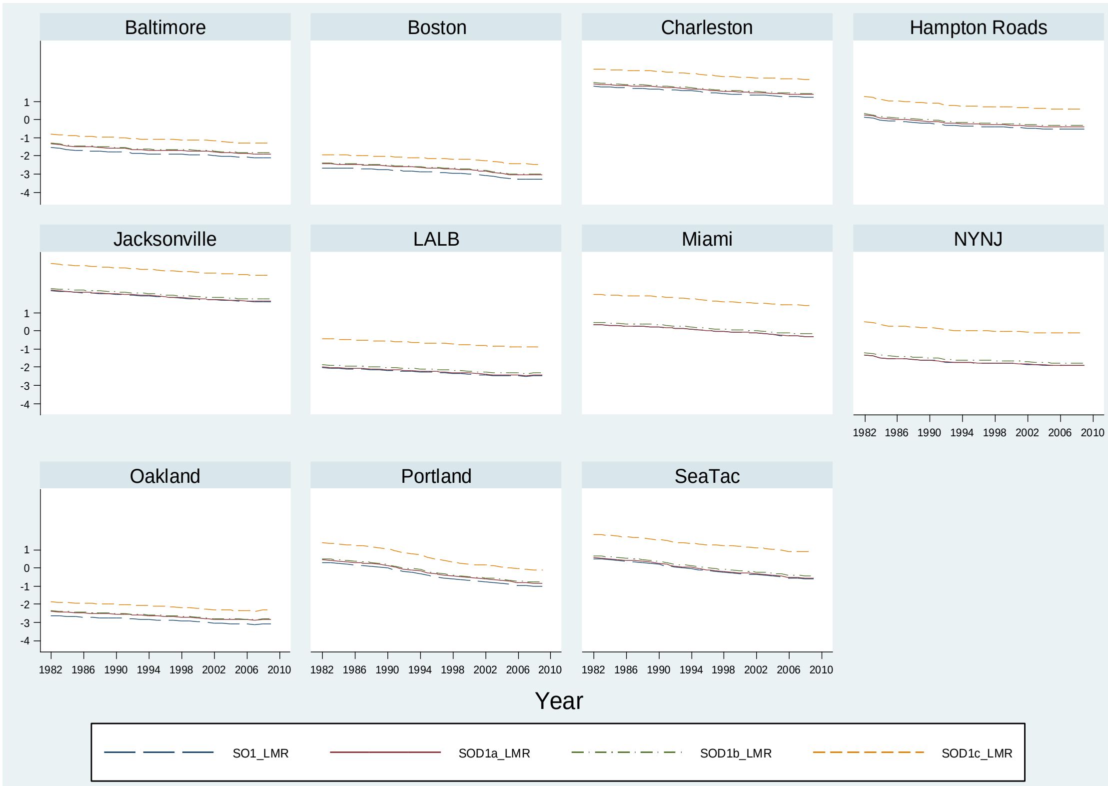
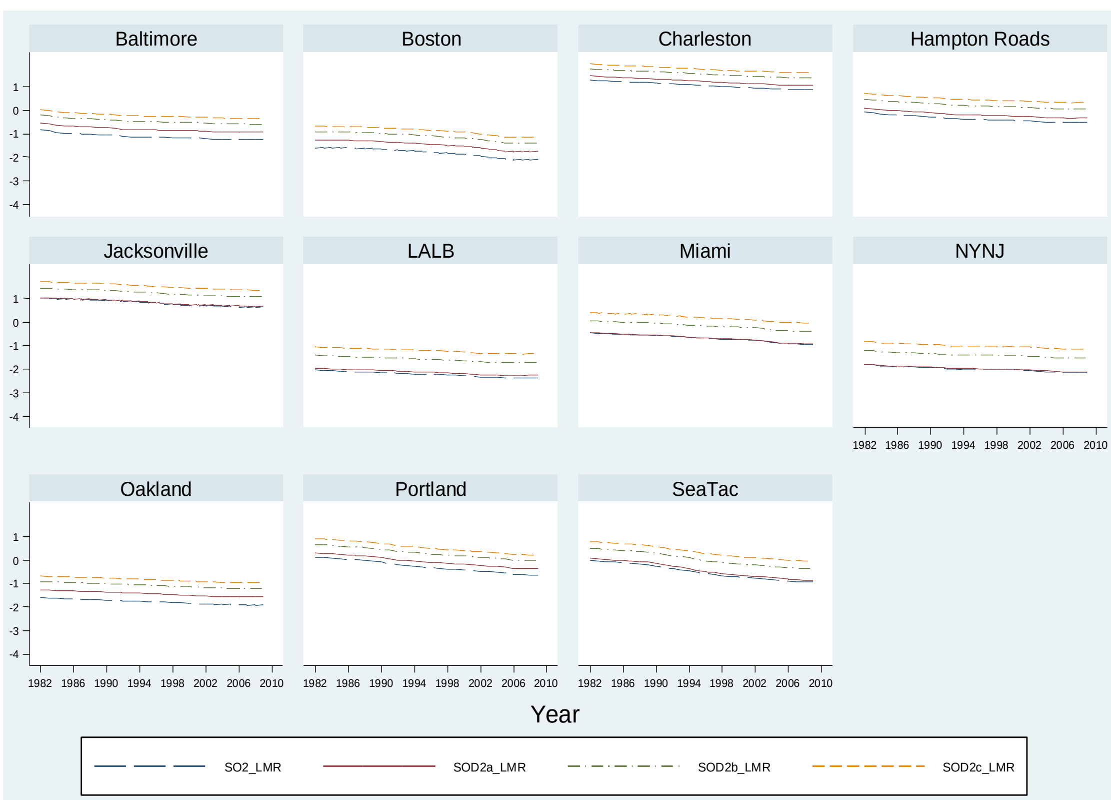


Figure 3.5 Estimated LMR elasticity of throughput (50-mile catchment area)



However, it is worthwhile to have a closer comparison between the results of regression models in (3.4) and (3.6). The LM elasticities decrease (become more negative or less positive) after controlling for road congestions (Table 3.5). This result can also be observed from Figures 3.2 and 3.3, as all of the curves of SOD models (i.e. after controlling for road congestion delays) are below the curves of SO models (without controlling for road congestion delays). This finding suggests that an increase in own road capacity expansion is more (less) strongly associated with container throughput reduction (increase) when road congestion is controlled for. In other words, this finding may imply that road capacity expansion is positively correlated with container throughput via the changes in road congestion, while the impact of road capacity expansion on throughput through other channels (as mentioned in Section 3.3) is negative. Regarding the LMR elasticities, we find that the curves representing LMR elasticities of SOD models are probably above the curves of SO models in Figures 3.4 and 3.5, implying that the rival's road capacity expansion is less (more) strongly associated with container throughput reduction (increase) given the road congestions being controlled for. One possible interpretation for this result is that rival's road capacity expansion may have a negative correlation with a port's container throughput via the adjustment in road congestions but a positive correlation with a port's container throughput through channels other than road congestion delays. Therefore, given these results, we might indeed distinguish the impact of road capacity expansion on throughput through road congestion and other channels.

A comparison between the theoretical predictions of the impacts of road capacity expansion and the above empirical findings provides two key points. First, via the endogenous adjustment of road congestion, port output tends to move in the same direction as own-road capacity but in the opposite direction as rival-road capacity, conforming to the prediction of the quantity competition model. Although we cannot conclude at this stage that quantity competition prevails in port competition, our finding suggests that the quantity competition model may provide a plausible prediction and it could be the focus of further investigation. Second, we also observe that urban road capacity may affect port output through other mechanisms which are not established in existing theoretical models. Such effects are not negligible and likely to cause the net relation between port output and road capacity to be positive for some ports while negative for the others. To provide a more complete and precise picture of this issue, future studies should include and explicitly model the impact of roads connecting to air, marine and rail facilities competing with the port in concern and treat roads directly connecting to the ports and roads in the other parts of the urban areas separately.

3.5.2 The impacts of hinterland accessibility on port efficiency

Tables 3.6 and 3.7 present the DEA scores for CCR and BCC models respectively. *CRS_P* stands for DEA scores from CCR model while *VRS_P* stands for DEA scores from BCC model. We do not find any outlier in our sample. Efficiency varies over time, but not necessarily increasing. On average, the efficiency scores initially have a slightly increasing trend but drop in 2008 and 2009, consistent with international trade reduction after the financial crisis (Figure 3.6). Smaller ports, such as Boston, Jacksonville and Portland, are quite sensitive to the assumption of scale efficiency. The three largest ports, Long Beach, Los Angeles and NYNJ, are relatively more efficient than the other ports.

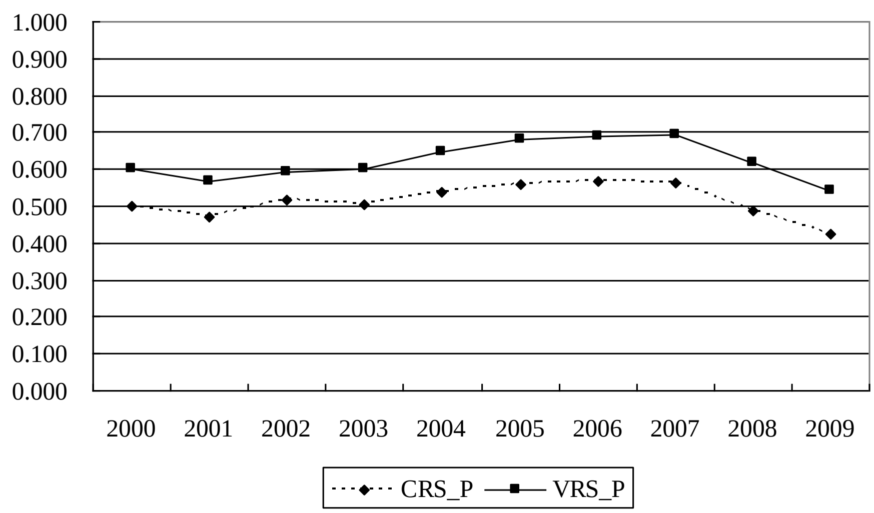
Table 3.6 DEA scores assuming constant return to scale (CCR model)

CRS_P	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Baltimore	0.204	0.198	0.204	0.212	0.224	0.242	0.252	0.245	0.171	0.146
Boston	0.139	0.133	0.142	0.158	0.440	0.474	0.502	0.552	0.523	0.469
Charleston	0.606	0.568	0.592	0.565	0.603	0.643	0.617	0.550	0.431	0.311
Hampton Roads	0.375	0.363	0.401	0.459	0.504	0.552	0.570	0.593	0.422	0.353
Jacksonville	0.546	0.539	0.528	0.449	0.472	0.504	0.499	0.461	0.453	0.490
Long Beach	0.689	0.630	0.732	0.500	0.621	0.852	0.925	0.824	0.685	0.547
Los Angeles	0.848	0.722	0.851	1.000	0.914	0.679	0.768	0.735	0.691	0.588
NYNJ	0.746	0.811	0.917	0.716	0.651	0.552	0.580	0.604	0.565	0.489
Oakland	0.356	0.336	0.408	0.378	0.366	0.407	0.428	0.444	0.415	0.457
Portland	0.238	0.228	0.209	0.278	0.178	0.104	0.139	0.169	0.159	0.063
Seattle	0.574	0.507	0.577	0.596	0.712	0.838	0.797	0.761	0.551	0.513
Tacoma	0.657	0.631	0.641	0.726	0.751	0.864	0.715	0.804	0.778	0.646
Average	0.498	0.472	0.517	0.503	0.536	0.559	0.566	0.562	0.487	0.423

Table 3.7 DEA scores assuming variable return to scale (BCC model)

VRS_P	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Baltimore	0.215	0.208	0.214	0.223	0.235	0.254	0.265	0.258	0.175	0.150
Boston	0.215	0.205	0.220	0.244	0.798	0.858	0.909	1.000	1.000	0.897
Charleston	0.634	0.594	0.619	0.587	0.625	0.666	0.638	0.569	0.454	0.328
Hampton Roads	0.386	0.373	0.411	0.471	0.518	0.567	0.586	0.609	0.433	0.363
Jacksonville	0.869	0.857	0.839	0.849	0.893	0.954	0.943	0.871	0.856	0.925
Long Beach	0.690	0.631	0.735	0.594	0.737	0.907	0.985	0.948	0.810	0.646
Los Angeles	0.876	0.722	0.851	1.000	1.000	0.884	1.000	0.986	0.927	0.797
NYNJ	0.761	0.828	0.936	0.721	0.652	0.625	0.660	0.687	0.666	0.577
Oakland	0.361	0.341	0.448	0.382	0.369	0.410	0.431	0.447	0.419	0.464
Portland	0.639	0.612	0.562	0.746	0.337	0.197	0.263	0.319	0.301	0.119
Seattle	0.602	0.532	0.607	0.627	0.750	0.881	0.839	0.798	0.571	0.531
Tacoma	0.941	0.902	0.684	0.767	0.834	0.958	0.744	0.849	0.821	0.682
Average	0.599	0.567	0.594	0.601	0.646	0.680	0.689	0.695	0.619	0.540

Figure 3.6 Average DEA scores



Tobit regression results for equation (3.7) are provided in Table 3.8. The four models presented here produce consistent coefficient estimations. Container port efficiency is negatively correlated with catchment area population. Intra-port competition has a positive relationship with port DEA scores: an increase in either operator-terminal ratio (*OP_TER*) or the number of container terminal operators (*NOPERATOR*) is associated with an increase in DEA scores. On average, large ports have higher DEA scores than small ports and the difference is huge. We also find that ports providing on-dock facility tend to have lower DEA scores, consistent with Turner *et al.*'s finding. The magnitude of this negative effect is much larger for BCC DEA scores. In agreement with our expectation, road congestion is negatively associated with DEA scores.

Table 3.8 Tobit regression results for the base case model

	DV = *CRS_P*		DV = *VRS_P*	
	CP1-1	CP1-2	VP1-1	VP1-2
ln(*P250*)	-0.1618** (0.0275)	-0.1619** (0.0244)	-0.1927** (0.0292)	-0.2303** (0.0294)
ln(*OP_TER*)	0.0985** (0.0409)		0.1829** (0.0524)	
NOPERATOR		0.0409** (0.0094)		0.0075 (0.0129)
NCLASS1	-0.0710** (0.0207)	-0.1667** (0.0505)	-0.1370** (0.0478)	-0.1405* (0.0722)
ONDOCK	-0.0681** (0.0298)	-0.0393 (0.0365)	-0.1563** (0.0522)	-0.1551** (0.0614)
ln(*D*)	-0.6712** (0.1225)	-1.2184** (0.2504)	-0.9565** (0.2001)	-0.9108** (0.3329)
LARGE	0.5417** (0.0400)	0.5982** (0.0610)	0.6176** (0.0722)	0.6626** (0.0802)
CONSTANT	2.3542** (0.2712)	2.4302** (0.2399)	3.0539** (0.2748)	3.3177** (0.2675)
N	120	120	120	120
F	51.36	45.16	55.76	44.13
Log pseudo-likelihood	56.0619	64.9689	31.3147	25.4917

Notes: * significant at α=0.10; ** significant at α=0.05; Values in the brackets are robust standard errors.

In addition to estimate equation (3.7), we are also interested in how the impact of own road congestion on port efficiency differs between large and small ports. Therefore, we drop $LARGE_i$ dummy from (3.7) and add one interactive term, $LARGE_i \times \ln(D_{it})$. Table 3.9 presents the estimation of the modified model. Note that this time the coefficients of NCLASS1 tend to be positive or statistically insignificant, suggesting that a positive association between port efficiency and rail service levels may exist. More importantly, we find that road congestion affects large ports and small ports differently. Road congestion clearly has a negative correlation with container port productivity, but for large ports there is also a positive association between those two variables. If we estimate the impact of road congestion for small and large ports with two separate interaction terms, $(1-LARGE_i)\times\ln(D_{it})$ and $LARGE_i \times \ln(D_{it})$, respectively, road congestion has a strong negative association with efficiency of small ports. However, the relationship between road congestion and large ports' efficiency scores is either positive or statistically insignificant. One explanation is that large ports defined in our sample are in fact primary ports of entry to the Continental U.S. These ports usually have superior logistics service providers, such as forwarders, trucking firms and insurance companies, gather around, offering high quality services with competitive prices. Consequently, the benefits from those services and convenience at those primary ports may outweigh the cost of delays on the road and shippers may choose those ports even if roads surrounding those ports are highly congested. Smaller ports, on the other hand, do not possess the same advantage as those primary ports and hence when shippers are considering using small ports, traffic condition on roads connecting the smaller ports and the hinterland becomes an important criterion. Therefore, road congestion does not substantially affect large ports' throughput and efficiency scores but has a significant negative impact on small ports.

Table 3.9 Regression results for the port difference model

	DV = *CRS_P*		DV = *VRS_P*	
	CP2-1	CP2-2	VP2-1	VP2-2
ln(*P250*)	-0.1138** (0.0255)	-0.1225** (0.0217)	-0.1378** (0.0287)	-0.1762** (0.0314)
ln(*OP_TER*)	0.1956** (0.0392)		0.2929** (0.0544)	
NOPERATOR		0.0649** (0.0084)		0.0283** (0.0129)
NCLASS1	0.0464** (0.0204)	-0.0807** (0.0277)	-0.0027 (0.0247)	-0.0254 (0.0437)
ONDOCK	-0.0607* (0.0333)	-0.0104 (0.0356)	-0.1484** (0.0494)	-0.1434** (0.0659)
ln(*D*)	-1.0595** (0.1597)	-2.0379** (0.2450)	-1.3879** (0.1889)	-1.4442** (0.3346)
LARGE×*ln*(*D*)	1.4480** (0.1180)	1.8038** (0.1374)	1.6448** (0.1238)	1.7169** (0.1743)
CONSTANT	1.7504** (0.2495)	1.8969** (0.2051)	2.3618** (0.2504)	2.5750** (0.2621)
N	120	120	120	120
F	40.54	38.04	49.78	28.01
Log pseudo-likelihood	56.3942	72.1990	31.2596	19.3657

Notes: * significant at α=0.10. ** significant at α=0.05. Values in the brackets are robust standard errors.

In Table 3.8, the coefficients for *NCLASS1* are negative and statistically significant, while in Table 3.9 their signs vary across specifications. Turner *et al.* (2004), however, find that the number of Class I railroads is positively associated with DEA scores with data taken in early 1990s. The difference between our findings and Turner *et al.*'s may stem from the fact that the deregulation started in the 1970s triggered a series of mergers and acquisitions in the U.S. rail industry since 1980s and consequently the number of Class I North American railroads has reduced from 56 in 1975 to seven in 2005 (Slack, 2009). Extensive mergers and acquisitions have resulted in a dominant duopolistic structure in port-oriented rail transport markets: except Boston and NYNJ, all the other ports in our sample are served by two Class I railroads, substantially reducing the variation in *NCLASS1* in our study. As shown by the pair-wise correlation test, no statistically significant correlation between *NCLASS1* and DEA scores is found. Thus, for samples taken in recent years, the number of Class I railroads might not be a perfect indicator of rail service status. When this variable is dropped from the regression models, our findings on the other factors, such as provision of on-dock facility and road congestion, persist, suggesting that inclusion of this variable does not significantly affect the robustness of our results.[47]

3.6 Concluding remarks

In this paper, we empirically study the impacts of urban road congestion and road capacity expansion on seaport container throughput, based on data derived from a sample of major container ports in the U.S. We find that a port's container throughput is statistically significantly associated with the congestion delays on its own urban roads, as well as delays on its rival's roads. Specifically, a 1% increase in own road congestion implies a reduction in container throughput by 0.90-2.48%, while a 1% increase in rival's road congestion implies an increase in container throughput by 0.62-1.69%. These associations between throughput and road congestion are mild without remedies for endogeneity but they become much stronger when endogeneity is taken into account. Thus, by mitigating road congestion in the urban area nearby, the port management would be able to effectively compete with its rivals. Our regression results are consistent to the observation that over the past decades, container ports in the U.S., especially those on the west coast, have invested substantially to expand on-dock rail capacity. By adding on-dock capacity, the ports can alleviate the pressure of growing container traffic on the roads nearby and at the same time accommodate more cargos shifted from other ground transportation modes due to road congestion.

Urban road capacity expansion, on the other hand, has different implications on container throughput. We find that own urban road capacity is positively correlated with the container throughput of the Port of NYNJ, the Port of LALB and the Port of Miami while negatively

[47] We observe a firstly increasing and then decreasing trend in the DEA scores, attributed to the impact of the 2008 financial crisis. To check whether our major findings are influenced by such trends, we estimate a number of variations of our models, such as including year dummies and U.S. international trade values. In general, these models produce results consistent with those presented in this paper.

correlated with the throughputs of the other ports in our sample. Furthermore, except the Port of Charleston and the Port of Jacksonville, the container throughput of a port tends to be negatively correlated with its rival's road capacity. The relationship between road capacity and container throughput via the changes in road congestion delays is largely consistent with the predictions obtained from the quantity-competition analytical model. That is, via the change of road congestion, an increase in road capacity implies an increase in output by the port nearby but a decrease in output by its rival port. Therefore, adding more roads might be an effective strategy to improve a port's competitiveness, provided that road capacity expansion solely affects road congestion while having little negative impact on the port through other channels. Local governments and port management should be cautious when deciding to provide more roads so as to reduce hinterland congestion and increase throughput, as adding road capacity might be harmful to the port throughput overall, although beneficial in terms of mitigating road congestion. Another caution that one needs to bear in mind when interpreting our result is the potential endogeneity problem between container throughput and road supply. This problem is much milder than the endogeneity between throughput and road congestion delay, because there could be a huge time lag between throughput growth and road capacity build-up due to the long planning, proofing and construction period for expanding roads, while the effect of traffic growth on road congestion takes place immediately.

We then examine the relationship between hinterland access conditions and U.S. container port efficiency. In various models, road congestion around a port tends to be negatively correlated with the port's DEA scores. Furthermore, we find that the efficiency of large ports is less sensitive to road congestion around them than that of small ones. As on-dock rail facility it requires land space – a port's input in the operation – which could be used to load/unload and store containers, and hence provision of such facility may lead to lower infrastructure efficiency scores. It might be true that on-dock rail facility helps to not only reduce road congestion but also speed up the ship-rail transferring process and hence less storage space at port is needed. Yet port managers should be cautious when deciding whether to make such investment as the benefits might be offset by the port efficiency loss if the terminal land space is limited.

Our study offers a couple of avenues for further investigation when better and more detailed data are available. First off, due to the lack of publicly accessible data, it is impossible to specify the individual competing markets based on origin-and-destination pairs at the moment and hence, by pooling all the markets together, we can only obtain the impact of road congestion averaged across markets. It is important to improve dataset and identify competing markets accurately in future studies. Second, we do not have lane-mile and road congestion data for roads around the ports or roads heavily used by port-related traffic. It is likely that extra capacity in the overall urban areas that we considered is added at the areas less used by commercial trucks and if so, it may amplify the effects of road capacity expansion via factors other than road congestion. Third, we approximate the modal split shares with data from the *U.S. Commodity Flow Surveys*, which include all domestic freight movements but are not limited to port-related container traffic. There might be a large share of truck-rail intermodal transport for port-related container movements, but we are not able to distinguish this possibility from truck-only container movements. Finally, this study does not find an unambiguous association between port efficiency and the number of Class I railroads. This may be due to the limitation of using the number of Class I railroads as the proxy of rail connection conditions. A better measurement of the accessibility to the rail network, like the frequency of rails from and to a port, is needed for future investigations.

4 Seaport Competition and Strategic Investment in Accessibility

4.1 Introduction

Over the past few decades, the port industry has undergone a number of major changes, including privatization, growth of container throughput, and globalization. Such changes have intensified seaport competition. As a node in the global supply 'chain' (Heaver, 2002), a port connects its hinterland – both the local and interior (inland) regions – to the rest of the world by an *intermodal transport* network. Talley and Ng (2013) deduce that determinants of port choice are also determinants of maritime transport chain choice. Among these determinants, hinterland accessibility is of major concern. It is argued that hinterland accessibility in particular has been one of the most influential factors of seaport competition (e.g. Notteboom, 1997; Kreukels and Wever, 1998; Fleming and Baird, 1999; Heaver, 2006). Empirical studies on major container ports in China and the Asia-Pacific region have found port-hinterland connection as a key factor in determining port competitiveness and productivity (Yuen *et al.*, 2012). In chapter 3, we found negative correlation between local road congestion and throughput and productivity of sampled container ports in the U.S.

As it is the intermodal chains rather than individual ports that compete (Suykens and Van De Voorde, 1998), seaport competition has been largely affected by the transportation infrastructure around the port as well as the transportation system in the inland. Consequently, plans on local transport infrastructure improvements, such as investment in road capacity, rail system and dedicated cargo corridors, are critical for local governments of major seaport cities as well as inland regions where shippers and consignees locate. Jula and Leachman (2011) study the allocation of import volume between San Pedro Bay Ports (i.e. Los Angelus and Long Beach ports) and other major ports in the U.S. and find that adequate port and landside infrastructure plays a significant role for San Pedro Bay Ports to maintain competitiveness.

Theoretical works discussing the interplay between ports and their landside accessibility are emerging (see De Borger and Proost, 2012, for a comprehensive literature review). One stream of the literature studies a single intermodal chain. Yuen *et al.* (2008) models a gateway port and a local road connecting the port to the hinterland and investigates the effects of congestion pricing implemented at the port on the hinterland's optimal road pricing, road congestion and social welfare. De Borger and De Bruyne (2011) examine the impact of *vertical integration* between terminal operators and trucking firms on optimal road toll and port charge, allowing trucking firms to possess market power. The other stream focuses on transport facility investment in the context of seaport or airport competition. De Borger et al. (2008), Zhang (2008), and chapter 2 study the impact of urban road or cargo corridor expansion on the performance of competing seaports. De Borger and Van Dender (2006) and Basso and Zhang (2007) study the investment decisions of two congestible but competing port facilities. The major difference between these two papers is that the former assumes ports face demand from final users (e.g. shippers and passengers) directly, while the later incorporates the vertical structure between the upstream ports and downstream carriers which in turn face demands from final users. One issue which has been overlooked by those papers is that transport infrastructure investment decisions made by individual local governments can affect the well-being of other port regions as well as the inland region through the mechanism of port competition. In the literature of seaport competition, to our knowledge, there is little work investigating the strategic behaviors of and interactions among seaport regions and inland region when making infrastructure investment decisions.

Thus, the focus of the present paper is the strategic investment decisions of local governments on local as well as inland transportation infrastructure in the context of seaport competition. In particular, we consider two seaports with their respective captive catchment areas and a common hinterland for which the seaports compete in prices. The two seaports and the common hinterland belong to three independent local governments, each determining the level of investment for its own regional transportation system. Based on this model, we answer the following questions: (1) how do infrastructure investment decisions affect port competitiveness? (2) How does transport infrastructure improvement affect each region's welfare? (3) How do optimal investment decisions look like under various forms of coordination (coalitions) among local governments? (4) Do port ownerships play a role in answering the above three questions? Although some of the aforementioned analytical papers also consider duopoly ports competing for a common hinterland, they focus on the competition and welfare effects of road or corridor expansions on the port regions while abstracting away the infrastructure decision of the common hinterland. Our setting is closest to Takahashi (2004) and Czerny *et al.* (2013), but there are a few major differences: (1) Takahashi does not care about investment decision of the inland region and assume local governments make both price and investment decisions; (2) Czerny *et al.* focus on port privatization games and ignore facility investment decisions; and (3) the present paper is the first one to examine the infrastructure investment rules under different ownership types and various forms of coordination among local governments of the seaport regions and the inland region.

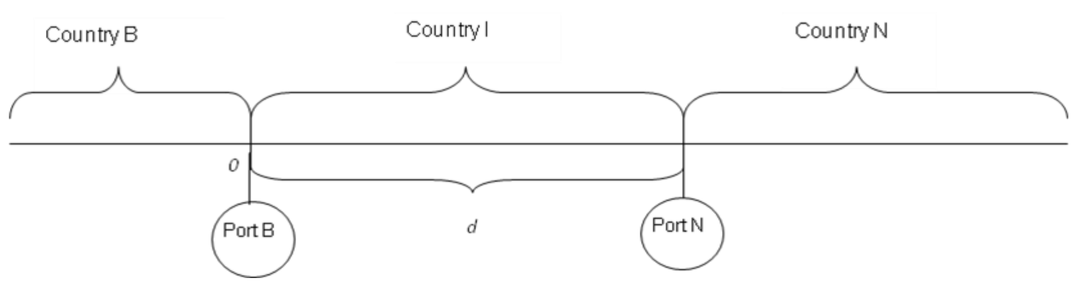
Our main findings are as follows. Increasing investment in the common hinterland lowers charges of both competing ports. Port ownership plays crucial roles in regional governments' strategic investment decisions. For public ports, an increment in investment in the captive catchment area of a certain port will cause severer reduction in its port charge than that of the rival port. However, for private ports, under certain conditions, improving a port region's accessibility may raise the charge of the port by a larger amount than that of the rival port. As a result, an increase in investment in the port region will reduce the welfare of the rival port region but improve the welfare of the common inland region if ports are public. The opposite may occur if ports are private. We also examine the equilibrium investment rules under various coalitions of local governments. In general, for regional governments of public ports, their incentive of infrastructure investment is the lowest when two port regions coordinate. They will invest more once at least one of them coordinates with the inland region. The inland region, on the other hand, always has higher incentive to invest at lower level of coordination. If the ports are private, the port regions' incentive of investment may be the highest when they coordinate while investment may be at the low end if the port region is coordinated with the inland.

The rest of the paper is organized as below. We present the basic model in Section 4.2. In Section 4.3, we derive the pricing decision of public seaports and private seaports respectively. The non-cooperative investment decisions of local governments are derived in Section 4.4. Section 4.5 compares the infrastructure decision in non-cooperative scenario with three forms of coalitions among local governments. Section 4.6 contains concluding remarks.

4.2 Basic model and shippers demand

We consider a linear continent, with three countries, *B*, *I* and *N*. Countries *B* and *N* have ports, but country *I* does not (Figure 4.1). The ports are non-congested regarding ship traffic and cargo handling and they deliver the cargoes right in the frontier between their countries and country *I*. We put the origin of coordinates at the boundary between port *B* and country *I*, and country *I* has a length of *d*.

Figure 4.1 Basic model



For simplicity, we assume that countries *B* and *N* start from the boundary points of country *I* and extent infinitely on the line. In all three countries, shippers, i.e. people or firms that want something shipped in from abroad, are distributed uniformly with a density of one shipper per unit of length. We assume that all shippers desire the same product and each has a demand to ship one unit of containerized cargoes.

Liners and forwarders bring the containers from abroad into the two ports for a fee, but the shippers are the ones that have to decide through which port the containers enter the continent and pay the port fee. Shippers have to pay then for an inland transportation service to bring the container to their address. We assume that the inland transportation costs are t_B, t_I and t_N per unit of distance in each country's non-congestible transportation network respectively.

Assume that liners and forwarders behave competitively, and hence bringing the containers into one or the other port costs the same. Thus, we will collapse their action to charge a given fee per container, which is set to zero without further loss of generality. The relevant players in this game then are: the two public ports, governments *B*, *N* and *I* and the shippers.

As for objective functions, private ports will maximize profit; while governments or public ports will maximize regional welfare which should include infrastructure expenditure, port profits and national shipper surplus. Shippers are considered because they contribute to a port's traffic and therefore to their profits. Liners and forwarders will not be considered.

The timing of the game is as follows. In the first stage, governments decide investment in accessibility, that is *t*'s. In the second stage, ports decide on prices to maximize their respective objectives. Finally, shippers decide whether they will demand the product or not, and which port to use. This defines the catchment areas of each port (and the market size for the forwarders). The game is solved by backward induction and we start with shippers' decisions.

Shippers have unit demands (per unit of time) and derive a gross-benefit of *V* if they get a container; otherwise their benefit is zero. Shippers care for the full price. Consider a shipper located in country *I* (i.e. at $0 < z < d$). If the shipper decides to use port *B* to bring in the container, she derives a full price of $\rho_B = p_B + t_I z$, and net utility of $U_B = V - \rho_B = V - p_B - t_I z$. Similarly, if she uses port *N*, she derives a net-utility: $U_N = V - \rho_N = V - p_N - t_I (d - z)$. Note that ρ_h is the full price, p_h is the port fee (per container), and t_I is the inland transportation cost that shippers from country *I* have to pay.

We assume that every shipper in country *I* gets a container and that both ports bring in containers for country *I*, then the shipper who's indifferent between using either port is given by $\rho_B = \rho_N$, that is $\tilde{z} = d/2 + (p_N - p_B)/2t_I$. These assumptions will hold as long as $0 < \tilde{z} < d$ and $U_B(\tilde{z}) = U_N(\tilde{z}) \geq 0$. That is, $| p_N - p_B | < dt_I \leq 2V - (p_B + p_N)$. This condition also implies that part of country *B* shippers will demand containers as well and those containers will be brought in through the national port. The same goes for *N*. We define

z^l as the last shipper on the left side of port *B* who gets a container. Similarly, we define z^r as the last shipper on the right side of port *N* who gets a container. Hence, taking into account the distribution of shippers along the line, the direct demands that each port faces is given by

$$Q_B = \tilde{z} + \left|z^l\right| = \tilde{z} + \frac{V - p_B}{t_B} \quad \text{and}$$

$$Q_N = (d - \tilde{z}) + (z^r - d) = (d - \tilde{z}) + \frac{V - p_N}{t_N} .$$

Replacing $\tilde{z}$, we obtain the following demands

$$Q_B = \frac{d\,t_B + 2V}{2t_B} + \frac{p_N}{2t_I} - \left(\frac{2t_I + t_B}{2t_I t_B}\right) p_B \text{ and } Q_N = \frac{d\,t_N + 2V}{2t_N} + \frac{p_B}{2t_I} - \left(\frac{2t_I + t_N}{2t_I t_N}\right) p_N \quad (4.1)$$

Let $k_B = 1/t_B$, $k_N = 1/t_N$ and $k_I = 1/2t_I$, and then the demand functions in (2.1) reduce to:

$$Q_B = (d/2) + k_B V - (k_B + k_I)\, p_B + k_I\, p_N \quad \text{and}$$

$$Q_N = (d/2) + k_N V + k_I\, p_B - (k_N + k_I)\, p_N \qquad (4.2)$$

This is a linear demand system with the standard dominance of own-effects over cross-effects, i.e., $\left|-(k_h + k_I)\right| > \left|k_I\right|$ for $h = B, N$, since $k_B, k_N, k_I > 0$. Furthermore, (4.2) shows that two ports produce substitutes. The substitutability arises due to the presence of country *I*'s shippers who may use either port for their shipment. To see this, recall that a port obtains its business from two markets: the captured national shippers and the overlapping shippers in country *I*. For port *h* ($h = B, N$) the quantity of the captured market may be denoted as Q_{hh}, and that of the overlapping market Q_{hI}. These quantities can be calculated as,

$$\begin{aligned} Q_{BB} &= k_B (V - p_B), \qquad Q_{BI} = (d/2) + k_I (p_N - p_B) \\ Q_{NN} &= k_N (V - p_N), \qquad Q_{NI} = (d/2) + k_I (p_B - p_N) \end{aligned} \qquad (4.3)$$

Clearly, we have $Q_{hh} + Q_{hI} = Q_h$. As can be seen from (4.3), the port demand of a captured market depends only on the price of its own. On the other hand, the port demand of the overlapping market depends on the prices of both ports: here, the two ports offer substitutable services. In particular, with $Q_{BI} + Q_{NI} = d$ – a fixed number – the gain in demand by one port is the loss in demand of the other port, and vice verse. Note that total demand in captive markets varies in prices and transportation costs, but as we assume that the inland market is always fully covered by the two ports and each port has positive demand, total demand from the inland is fixed. If the above mentioned inland market coverage assumption is violated, total inland demand will vary as well, but the two ports will no longer compete but become two monopolies as inland shippers who locate near to the ports will ship but those who are in the middle of the inland will not ship at all. Another merit of imposing this assumption is to avoid the situation that one port lowers its price to the extent that shippers inside the other port's captive area find shipping via the rival port located far away is cheaper than with the local port. Then, the rival port will obtain all the business of the local port, leading to discontinuity problem of the demand function. The present study confines analysis to cases that inland market and transportaion costs are so large that demand discontinuity will not occur. All the other cases can be considered as an extension in the future. We shall further assume all the four quantities in (4.3) are positive, implying that $p_B < V$, $p_N < V$, and p_B and p_N are not too different from each other, i.e. $| p_B - p_N | < d/2k_I$.[48]

4.3 Equilibrium prices for ports

4.3.1 Public ports

Consider first that each port decides on its price to maximize regional welfare. This is the case in which the port is publicly operated: the port authority chooses the region's social

[48] For public ports, at equilibrium, Q_{BI} and Q_{NI} are both positive for any k_I, k_B and k_N > 0 (see Appendix B).

surplus as its objective. More specifically, region *B*'s welfare is the sum of region *B*'s consumer surplus and the port's profit, minus the infrastructure cost $c_B(k_B)$. Here, we care about improvement in infrastructure within a region rather than inside a port. Such investment may involve lots of direct investment from local governments but not terminal operators. Therefore, in the present study, we assume infrastructure investment costs are born by local governments rather than by the ports.

$$\begin{aligned} W^B(p_B, p_N; k_B, k_I) &= CS^B + \pi^B - c_B(k_B) \\ &= (k_B/2)(V - p_B)^2 + p_B Q_B - c_B(k_B) \end{aligned} \tag{4.4}$$

In (4.4) region *B*'s consumer surplus is calculated as $CS^B = \int_0^{k_B(V-p_B)} [V - p_B - (z/k_B)]dz$, and the port has zero operating cost and so its profit is just equal to revenue $p_B Q_B$. Also note that k_I enters the $W^B(\cdot)$ function via $Q_B(\cdot)$. Similarly, region N's welfare can be expressed as,

$$\begin{aligned} W^N(p_B, p_N; k_N, k_I) &= CS^N + \pi^N - c_N(k_N) \\ &= (k_N/2)(V - p_N)^2 + p_N Q_N - c_N(k_N) \end{aligned} \tag{4.5}$$

The equilibrium port prices are determined by the following first-order conditions:

$$W_H^H \equiv \frac{\partial W^H}{\partial p_H} = -Q_{HH} + Q_H + p_H \frac{\partial Q_H}{\partial p_H} = Q_{HI} - p_H(k_H + k_I) = 0,\ H \in \{B, N\}. \tag{4.6}$$

The ports' second-order conditions are satisfied, because $W_{BB}^B = -k_B - 2k_I < 0$ and $W_{NN}^N = -k_N - 2k_I < 0$ (subscripts again denoting partial derivatives). Further, the equilibrium is unique and stable, as $\Delta_W \equiv W_{BB}^B W_{NN}^N - W_{BN}^B W_{NB}^N = k_B k_N + 2k_B k_I + 2k_N k_I + 3k_I^2 > 0$. Equation (4.6) can be rewritten into:

$$Q_{HI} + p_H \frac{\partial Q_{HI}}{\partial p_H} + p_H \frac{\partial Q_{HH}}{\partial p_H} = \frac{\partial \pi_{HI}}{\partial p_H} + p_H \frac{\partial Q_{HH}}{\partial p_H} = 0,\ H \in \{B, N\}.$$

That is, at equilibrium the marginal profit from the inland market is positive while the net impact of price increase on the captive region equals to the impact on the profit loss due to reduced captive demand which is negative.

We use $p^{WB}(k_B, k_N, k_I)$ and $p^{WN}(k_B, k_N, k_I)$ to denote the equilibrium port charges for public ports where the superscript *W* denotes for public ports. Then, we obtain, by equation (4.6), the identities $W_B^B(p^{WB}, p^{WN}; k_B, k_I) \equiv 0$ and $W_N^N(p^{WB}, p^{WN}; k_N, k_I) \equiv 0$. Totally differentiating these identities with respect to k_B yields

$$p_B^{WB} \equiv \partial p^{WB}(k_B, k_N, k_I)/\partial k_B = -p^{WB}(k_N + 2k_I)/\Delta_W < 0; \tag{4.7}$$

$$p_B^{WN} \equiv \partial p^{WN}(k_B, k_N, k_I)/\partial k_B = -p^{WB}k_I/\Delta_W < 0. \tag{4.8}$$

Thus, an increase in k_B will reduce the equilibrium charges of both ports. The intuition behind this result is as follows. First, it can be easily seen that the first-order conditions (4.6) generate two upward-sloping reaction functions – noting that $W_{BN}^B = W_{NB}^N = k_I > 0$ and so strategy variables p_B and p_N are strategic complements in the port game. Second, an increase in k_B reduces W_B^B, the marginal welfare increment with respect to p_B, thereby shifting port *B*'s reaction function downward. Given that port *N*'s reaction function remains un-shifted, the price equilibrium moves down along *B*'s reaction function, leading to a fall in both p^{WB} and p^{WN}. Moreover, we have

$$p_B^{WB} - p_B^{WN} = -p^{WB}(k_N + k_I)/\Delta_W < 0. \tag{4.9}$$

Consequently, the reduction in p^{WB} – following an increase in k_B – is greater than the reduction in p^{WN}, reflecting the fact that port *B*'s reaction function is steeper than port *N*'s.

As for the effects of k_I on port charges p^{WB} and p^{WN}, it can be calculated,

$$\begin{aligned} p_I^{WB} &\equiv \partial p^{WB}(k_B, k_N, k_I)/\partial k_I = -(d/2\Delta_W^2)[(k_N + 3k_I)^2 + k_N(k_N - k_B)] \\ p_I^{WN} &\equiv \partial p^{WN}(k_B, k_N, k_I)/\partial k_I = -(d/2\Delta_W^2)[(k_B + 3k_I)^2 + k_B(k_B - k_N)] \end{aligned}. \tag{4.10}$$

Summing up the two equations in (4.10), we get:

$$p_I^{WB} + p_I^{WN} = -(d / 2\Delta_W^2)[(k_N + 3k_I)^2 + (k_B + 3k_I)^2 + (k_N - k_B)^2] < 0 \tag{4.11}$$

Inequality (4.11) shows that an increase in k_I will reduce the equilibrium charges for at least one port. Further, by (4.10), an increase in k_I will reduce the equilibrium charges of both ports if and only if $(k_N + 3k_I)^2 + k_N(k_N - k_B) > 0$ and $(k_B + 3k_I)^2 + k_B(k_B - k_N) > 0$, which hold if the two port regions are not too asymmetric. We shall assume this is the case for the remainder of the paper. The above comparative static results are summarized as follows:

Lemma 4.1: *Assuming public ports, then (i) an increase in k_B reduces the equilibrium charges of both ports – and here, the reduction in p^{WB} is greater than the reduction in p^{WN}. (ii) The effects of an increase in k_N can be similarly given. (iii) An increase in k_I reduces the equilibrium charges of both ports.*

The intuition behind the positive effect of k_B, k_N, and k_I on port charges may be seen as follows. With the present demand and other specifications, the equilibrium port prices can be calculated as,

$$\begin{aligned} p^{WB}(k_B, k_N, k_I) &= \frac{(k_N + 3k_I)d}{2(k_B k_N + 2k_B k_I + 2k_N k_I + 3k_I^2)} \\ p^{WN}(k_B, k_N, k_I) &= \frac{(k_B + 3k_I)d}{2(k_B k_N + 2k_B k_I + 2k_N k_I + 3k_I^2)} \end{aligned} \tag{4.12}$$

Assuming symmetric equilibrium, (4.12) reduces to

$$p^{WB} = p^{WN} = \frac{d}{2(k_H + k_I)}, \quad \text{where} \quad k_H = k_B = k_N \quad .$$

Therefore, essentially an increase in k_B, k_N, or k_I will make the demands more elastic, and thereby reduce the prices that the ports can charge. When the size of inland region, *d*, increases, the equilibrium charges will increase and the port with worse local accessibility will raise port charge more than the other port.

4.3.2 Private ports

Now consider two private ports competing simultaneously. Taking the land-side infrastructure decisions as given, each private port maximizes its profit:

$$\pi^H = p_H(Q_{HH} + Q_{HI}) \text{, where } H \in \{B, N\}.$$

Taking first-order conditions with respect to p_H leads to the following:

$$Q_{HH} + Q_{HI} = p_H(k_H + k_I) \tag{4.13}$$

Equation (4.13) can be rewritten as

$$\frac{\partial \pi_{HI}}{\partial p_H} = -\frac{\partial \pi_{HH}}{\partial p_H}, \; H \in \{B, N\}.$$

That is, at equilibrium, except for the special case where the marginal profits for both captive and inland markets are zero, the maginal profits in the two markets have different signs and one is offset by the other. When the equilibrium p_H is much lower than the shipping utility V (i.e. $p_H < V/2$), an increase in price leads to a gain in the captive market but a loss in the inland market; otherwise, the opposite will hold.

Again, the second-order conditions are satisfied as $\pi^H_{HH} = -2(k_H + k_I) < 0$. Solving for (4.13), we obtain the equilibrium port changes:

$$p^{\pi B}(k_B, k_N, k_I) = \frac{2V(k_N k_I + 2k_B(k_N + k_I)) + (3k_I + 2k_N)d}{2\Delta_\pi} \text{, and}$$

$$p^{\pi N}(k_B, k_N, k_I) = \frac{2V(k_B k_I + 2k_N(k_B + k_I)) + (3k_I + 2k_B)d}{2\Delta_\pi}, \tag{4.14}$$

where the superscript π denotes the equilibrium of private ports and $\Delta_\pi \equiv \pi_{BB}^{B}\pi_{NN}^{N} - \pi_{BN}^{B}\pi_{NB}^{N}$ $= 4(k_B k_N + k_B k_I + k_N k_I) + 3k_I^2 > 0$. Consequently, the difference between the equilibrium port charges is:

$$p^{\pi B} - p^{\pi N} = \frac{(k_B - k_N)(Vk_I - d)}{\Delta_\pi}. \tag{4.15}$$

Based on (4.15), it is straightforward to reach Lemma 4.2.

Lemma 4.2: *For private ports, at equilibrium, if $k_I V - d > 0$ holds, the sign of $p^{\pi B} - p^{\pi N}$ depends on the sign of $k_B - k_N$; and if $k_I V - d < 0$ holds, the sign of $p^{\pi B} - p^{\pi N}$ depends on the sign of $k_N - k_B$.*

Contrary to the case of public ports, when $k_I V - d > 0$, port B will be able to charge higher than port N if and only if the transportation infrastructure in country B is superior to that in country N. Similar to Section 4.3.1, we derive comparative statics for equilibrium port charges by differentiating both sides of (4.13) with respect to k_B and using the Cramer's rule. That is,

$$p_B^{\pi B} = \frac{\partial p^{\pi B}(k_B, k_N, k_I)}{\partial k_B} = \frac{2(k_N + k_I)(V - 2p^{\pi B})}{\Delta_\pi} = \frac{2(k_N + k_I)(3k_I + 2k_N)(k_I V - d)}{\Delta_\pi^2},$$

$$p_B^{\pi N} = \frac{\partial p^{\pi N}(k_B, k_N, k_I)}{\partial k_B} = \frac{k_I(V - 2p^{\pi B})}{\Delta_\pi} = \frac{k_I(3k_I + 2k_N)(k_I V - d)}{\Delta_\pi^2}, \text{ and}$$

$$p_B^{\pi B} - p_B^{\pi N} = \frac{(2k_N + k_I)(3k_I + 2k_N)(Vk_I - d)}{\Delta_\pi^2}.$$

Based on Lemma 4.2, we know that if $k_I V - d > 0$ holds, $p_B^{\pi B} > 0$ and $p_B^{\pi N} > 0$. This is again opposite to the case of public ports where an improvement in the transportation infrastructure in any port country will cause a decrease in port charges. This difference between public and private ports will eventually lead to differentiated results for the investment decisions made by individual local governments. In addition, we have $p_B^B - p_B^N > 0$. However, if $k_I V - d < 0$ holds, we obtain similar outcomes as in the case of public ports. That is, $p_B^{\pi B} < 0$, $p_B^{\pi N} < 0$ and $p_B^B - p_B^N < 0$

Differentiate both sides of (4.13) with respect to k_I and use Cramer's rule:

$$
\begin{aligned}
p_I^{\pi B} &= \frac{\partial p^{\pi B}(k_B, k_N, k_I)}{\partial k_I} = \frac{2k_N p^{\pi N} - (3k_I + 4k_N)p^{\pi B}}{\Delta_\pi} \\
&= -\frac{1}{2\Delta_\pi^2}\left((8k_N^2 k_B + 24k_B k_N k_I + 6k_I^2 k_N + 12k_I^2 k_B)V + ((3k_I + 2k_N)^2 + 4k_N(k_N - k_B))d\right).
\end{aligned}
$$

Assumptions for the shippers' demand equilibrium require that $d < 4Vk_I$, implying that $24k_B k_N k_I V - 4k_N k_B d = 4k_N k_B(6Vk_I - d) > 0$. Therefore, $p_I^{\pi B}$ must be negative. Similarly, we have:

$$
p_I^{\pi N} = \frac{\partial p^{\pi N}(k_B, k_N, k_I)}{\partial k_I} = \frac{2k_B p^{\pi B} - (3k_I + 4k_B)p^{\pi N}}{\Delta_\pi} < 0.
$$

Therefore, the above analysis leads to Lemma 4.3.

Lemma 4.3: *Assuming private ports, then we have (i) if $k_I V - d > 0$, an increase in k_B <u>increases</u> the equilibrium charges of both ports – and here, the <u>increase</u> in $p^{\pi B}$ is greater than the <u>increase</u> in $p^{\pi N}$; (ii) if $k_I V - d < 0$, an increase in k_B <u>reduces</u> the equilibrium charges of both ports – and here, the <u>reduction</u> in $p^{\pi B}$ is greater than the <u>reduction</u> in $p^{\pi N}$; (iii) the effects of an increase in k_N can be similarly given; and (iv) an increase in k_I reduces the equilibrium charges of both ports.*

4.4 Non-cooperative infrastructure equilibrium

This section derives the equilibrium infrastructure investments rules when the social planers for the three countries simultaneously choose the level of infrastructure accessibility which in turn affects regional welfare through subsequent port competition. Taking the ports' price decisions into account, a port region's welfare is given by:

$$\phi^H(k_B, k_N, k_I) \equiv W^H(p^B(k_B, k_N, k_I), p^N(k_B, k_N, k_I); k_H, k_I), \ H = B, N, \quad (4.16)$$

where we suppress the notation for private ports and public ports and use $p^H(k_B, k_N, k_I)$ to denote the equilibrium charge of port *H*. Social surplus of region *I*, the inland country, is just equal to its consumer surplus, CS^I, minus the infrastructure cost $c_I(k_I)$:

$$\phi^I(k_B, k_N, k_I) \equiv CS^I(p^B(k_B, k_N, k_I), p^N(k_B, k_N, k_I); k_I) - c_I(k_I) \quad (4.17)$$

In (4.17),

$$CS^I = \int_0^{\tilde{z}} [V - p_B - (z/2k_I)]dz + \int_0^{d-\tilde{z}} [V - p_N - (z/2k_I)]dz \quad (4.18)$$

where $\tilde{z}$ is the shipper of region *I* who is indifferent between using port *B* and using port *N*, and $\tilde{z} = (d/2) + k_I(p_N - p_B)$.

Governments decide on investment in accessibility, that is, the *k*'s. In particular, the non-cooperative infrastructure equilibrium arises when each government chooses its welfare-maximizing infrastructure investment, taking the investment of the other governments as given at the equilibrium value. Specifically, it is characterized by the following first-order conditions,

$$\phi_B^B \equiv \partial\phi^B / \partial k_B = 0, \qquad \phi_N^N \equiv \partial\phi^N / \partial k_N = 0, \qquad \phi_I^I \equiv \partial\phi^I / \partial k_I = 0 \quad (4.19)$$

We now take a closer look at each of the marginal effects in (4.19), starting with port country *B*. The effects of k_N on country *N*'s welfare can be similarly analyzed. As indicated earlier, the impacts of k_B on the regional welfare of country *B* is:

$$\phi_B^B = W_B^B p_B^B + W_N^B p_B^N + \frac{\partial W^B}{\partial k_B}$$
$$= \left[Q_{BI} - (k_B + k_I) p^B\right] p_B^B + p^B k_I p_B^N + \frac{(V - p^B)(V + p^B)}{2} - c_B^{'}(k_B) \quad . \tag{4.20}$$

If ports maximize regional welfares when choosing their charges, the first term of (4.20) becomes zero and (4.20) reduces to

$$\phi_B^{WB} = p^{WB} k_I p_B^{WN} + \frac{(V - p^{WB})(V + p^{WB})}{2} - c_B^{'}(k_B), \tag{4.21}$$

where the first term is negative by Lemma 4.1. It represents the reduction of market share and hence revenue in the hinterland market as the rival port reduces its port charge when country *B*'s infrastructure improves. The second term is positive, as it is the direct increase in the (gross) benefit of *B*'s shippers due to less transport friction (cost) in country *B*.

In the case of private ports, (4.13) implies the first term of (4.20) becomes $-Q_{BB}$ and hence we have

$$\phi_B^{\pi B} = -Q_{BB} p_B^{\pi B} + p^{\pi B} k_I p_B^{\pi N} + \frac{(V - p^{\pi B})(V + p^{\pi B})}{2} - c_B^{'}(k_B). \tag{4.22}$$

When $k_I V - d > 0$, the first term is a negative indirect effect as higher investment in *B*'s infrastructure leads to higher port charges, less country *B*'s shipping demand and hence the (gross) benefit of country *B*'s shippers reduces. The indirect effect due to price adjustment of the rival (the second term of (4.22)) and the direct effect on the (gross) benefit of country *B*'s shippers (the third term of (4.22)) are both positive. When $k_I V - d < 0$, the effect of region *B*'s accessibility on the first term becomes positive while that on the second term becomes negative, as port charges decrease in port region's accessibility.

We next consider the effect of k_I on region *I*'s welfare. From (4.17)-(4.18) we obtain,

$$\phi_I^I = [\frac{\partial CS^I}{\partial p_B} p_I^B + \frac{\partial CS^I}{\partial p_N} p_I^N] + \frac{\partial CS^I}{\partial k_I} - c_I^{'}(k_I)$$
$$= [(-Q_{BI}) p_I^B + (-Q_{NI}) p_I^N] + \frac{Q_{BI}^2 + Q_{NI}^2}{4k_I^2} - c_I^{'}(k_I) . \quad (4.23)$$

For both public and private ports, equation (4.23) holds. Moreover, both the first and second terms on the right-hand side (RHS) of (4.23) are, by Lemma 4.1 and Lemma 4.3, positive. While the second term reflects the direct effect of an infrastructure improvement, the first term represents the indirect effect of an infrastructure improvement (via its impact on the port charges, which in turn benefits region *I*'s shippers). The two positive terms are balanced against the cost of infrastructure improvement, $c_I^{'}(k_I)$.

The impact of infrastructure investment on other regions can also be derived. In particular, the effect of k_B on region *N*'s welfare can be written as:

$$\phi_B^N = W_N^N p_B^N + W_B^N p_B^B = [Q_{NI} - (k_N + k_I) p^N] p_B^N + p^N k_I p_B^B . \quad (4.24)$$

As mentioned in Section 4.3.1, since ϕ_B^N is evaluated at equilibrium, when the competing ports are both public, W_N^N is zero and (4.24) reduces to $\phi_B^{WN} = p^{WN} k_I p_B^{WB} < 0$. Intuitively, an increase in k_B will lower port *N*'s profit from the inland market due to substantial price-cut by port *B*. Port *N* will lower its price as well, which leads to a gain from the captive market as captive demand increases and a loss from the inland market as lower price substantially lowers inland profit margin while the number of shippers attracted from the rival port is very limited. At equilibrium, these two trade-offs due to a decrease in port *N*'s price have to be balanced out, leaving the negative impact of the reduction in port *B*'s price as the only effective influence on region *N*'s equilibrium welfare.

When both ports are private, (4.24) becomes $\phi_B^{\pi N} = -Q_{NN} p_B^{\pi N} + p^{\pi N} k_I p_B^{\pi B}$. Thus, $\phi_B^{\pi N}$ is decomposed into two components with opposite signs. For example, when $k_I V - d > 0$, the first component is negative as an increase in k_B raises country *N*'s port charge and hence lowers the consumer surplus of *N*'s shippers while the marginal change of port *N*'s profit with respect to its own price increase is zero at equilibrium. However, the price charged by port *B* increases as well, making port *N* more attractive to hinterland shippers and hence raises port *N*'s profit. We can show that the net effect is positive by using the first-order conditions (4.13) and equation (4.14): $\phi_B^{\pi N} = k_I (Q_{NN} + 2Q_{NI})(V - 2p^{\pi B}) / \Delta_\pi > 0$. As predicted by Lemma 4.3, the price increase from port *B* is larger than port *N*, so the revenue gain from region *I*'s market can compensate the surplus loss of shippers' in country *N*. As a result, the welfare of country *N* will increase eventually. However, when $k_I V - d < 0$, we can show that $V - 2p^{\pi B} < 0$ and hence $\phi_B^{\pi N} < 0$.

The effect of k_B on region *I*'s welfare:

$$\phi_B^I = \frac{\partial CS^I}{\partial p_B} p_B^B + \frac{\partial CS^I}{\partial p_N} p_B^N = (-Q_{BI}) p_B^B + (-Q_{NI}) p_B^N . \tag{4.25}$$

Therefore, for public ports, an increase in k_B will benefit country *I*'s shippers since the port charges of both ports will decrease (i.e. $\phi_B^{WI} > 0$), while for private ports, an increase in k_B will reduce country *I*'s welfare (i.e. $\phi_B^{\pi I} < 0$) when $k_I V - d > 0$ and increase country *I*'s welfare (i.e. $\phi_B^{\pi I} > 0$) when $k_I V - d < 0$. We can derive similar results for the effect of k_N on region *B*'s welfare as well as on region *I*'s welfare.

The effect of k_I on region *B*'s welfare:

$$\begin{aligned} \phi_I^B &= W_B^B p_I^B + W_N^B p_I^N + \partial W^B / \partial k_I \\ &= \left[Q_{BI} - (k_B + k_I) p^B\right] p_I^B + p^B k_I p_I^N + p^B (p^N - p^B) \end{aligned} . \tag{4.26}$$

As mentioned in Section 4.3.1, for public ports, at equilibrium W_B^B is zero. Thus, the first term of (4.26) is zero and equation (4.26) reduces to

$$\phi_I^{WB} = p^{WB} k_I p_I^{WN} + p^{WB} (p^{WN} - p^{WB}) . \tag{4.27}$$

The first term on the RHS of (4.27) is negative, because increasing the accessibility of the inland region leads to lower charge of port *N* so that some inland shippers will switch to port *N*. Again, although port *B* will also lower its port charge, such positive and negarive impacts from *B*'s price reduction will cancel with each other around the equiliriburm point. When the accessibility of country *B* is worse than country *N*, i.e. $k_B < k_N$, port *B* charges higher than port *N* and hence port *N* has competitive advantage over port *B* for inland shippers. Then, improved the accessibility of region *I* makes inland shippers more sensitive to this price difference between port *B* and port *N* and more willing to use port *N*; as a result, the second term on the RHS of (4.27) is negative. However, when $k_B > k_N$, we have $p^N > p^B$ and increasing k_I makes port *B* more attractive to inland shippers and hence the second term on the RHS of (4.23) will be positive.

When ports maximize profits, equation (4.26) becomes

$$\phi_I^{\pi B} = -Q_{BB} p_I^{\pi B} + p^{\pi B} k_I p_I^{\pi N} + p^{\pi B} (p^{\pi N} - p^{\pi B}) . \tag{4.28}$$

According to part (iv) of Lemma 4.3, the first term on the RHS of (4.28) is positive, equivalent to the amount of surplus increase for country *B*'s shippers as an increase in k_I causes port *B* to cut price. As port *N* cuts price as well, it attracts some inland shippers away from port *B* and thus the second term on the RHS of (4.28) is negative. Similar to the case of public ports, the sign of the last term on the RHS of (4.28) depends on the relative accessibility of country *B* and country *N*.

We can obtain similar comparative static result for the effect of k_I on country *N*'s welfare. The above discussion leads to Propositions 4.1 and 4.2.

Proposition 4.1: *Assuming public ports, then (i) an increase in* k_B *(*k_N*) reduces the welfare of region N (region B); (ii) an increase in* k_B *or* k_N *raises region I's welfare; and (iii) an increase in* k_I *reduces the welfare of the port region with less accessible infrastructure, while may or may not increase the welfare of the other port region.*

Proposition 4.2: *Assuming private ports, (i) if* $k_I V - d > 0$*, an increase in* k_B *(*k_N*)*

increases the welfare of region N (region B), while an increase in k_B or k_N reduces region I's welfare; (ii) if $k_I V - d < 0$, an increase in k_B (k_N) reduces the welfare of region N (region B), while an increase in k_B or k_N increases region I's welfare; and (iii) an increase in k_I has ambiguous effect on the other regions' welfares.

Suppose two port regions have the same level of accessibility, i.e. $k_B = k_N = k_H$, and this leads to $p^N = p^B = p^H$. Then, the last term of (4.27) disappears and $\phi_I^{WB} < 0$. Intuitively, when inland accessibility increases, both ports' prices will reduce by the same amount and hence each port still obtain half of the inland market share, but the profit from inland market reduces as the port will earn less from each shipper. In the captive market lower port charge induces more captive demand, but this gain is substantially lower than the loss from the inland market around the equilibrium point. In the case of private ports, (4.28) can be rewritten as $\phi_I^{\pi B} = (-(2k_H + 3k_I)^2 p^{\pi H} / 2\Delta_\pi^2)[(k_H + k_I)d - 2Vk_H^2]$. Thus, if ports are both private, an increase in inland accessibility will raise the port regions' welfare if and only if the port regions' accessibility is high enough and inland accessibility is low enough such that $(k_H + k_I)/k_H^2 < 2V/d$. Intuitively, when inland accessibility improves, inaddtion to the impacts mentioned above, there will be an extra consumer surplus gain from the captive market due to lower port charge. This part of the benefit is not internalized by the private port and hence is not balanced out at ports' price competition stage. If port regions' accessibility is high, demand stems from the port regions is more sensitive to the price. As a result, the price-cut due to increased inland accessibility will induce a large number of additional shippers in region *B*, leading to a substantial increase in region *B*'s consumer surplus which is large enough to compensate the revenue loss in the inland market, and hence raise welfare for the port regions. If we assume that the two port regions have the same functional forms of investment costs, i.e. $c_B(\cdot) = c_N(\cdot) = c_H(\cdot)$. By imposing symmetry, at equilibrium, regions *B* and *N* will choose the same level of accessibility. Then, the above discussion will apply and lead to Proposition 4.3.

Proposition 4.3: *Suppose $c_B(\cdot) = c_N(\cdot) = c_H(\cdot)$. At non-cooperative equilibrium for investment decisions, (i) if both ports are public, an increase in inland accessibility will* *reduce port regions' welfare; (ii) if both ports are private, an increase in inland accessibility will raise (reduce) welfare for other regions if the port regions' accessibility is high (low).*

4.5 Infrastructure equilibrium under coalitions

This section examines the equilibrium infrastructure investment decisions given that the three regions co-operate in various forms. Without loss of generality, we consider three forms of coalitions.

Coalition 1: region B and region N coordinate while region I remains independent

The social planners of regions *B* and *N* choose k_B and k_N together to maximize the joint welfare of these two regions. The joint welfare of two port regions is

$$\phi^{BN}(k_B, k_N, k_I) \equiv \phi^B(k_B, k_N, k_I) + \phi^N(k_B, k_N, k_I).$$

The optimal investment rule is characterized by:

$$\begin{aligned} \phi_B^{BN} &\equiv \partial\phi^B / \partial k_B + \partial\phi^N / \partial k_B = \phi_B^B + \phi_B^N = 0 \\ \phi_N^{BN} &\equiv \partial\phi^B / \partial k_N + \partial\phi^N / \partial k_N = \phi_N^B + \phi_N^N = 0. \\ \phi_I^I &\equiv \partial\phi^I / \partial k_I = \phi_I^I = 0 \end{aligned} \tag{4.29}$$

Assuming public ports, from Propositions 4.1 we can derive that at equilibrium $\phi_B^{WB} > 0$ and $\phi_N^{WN} > 0$. As the governments' second-order conditions must be satisfied, for given levels of k_I and k_N, $\phi_{BB}^B \equiv \partial^2\phi^B / \partial {k_B}^2 < 0$. As a result, given fixed k_I and k_N (or k_B), k_B (or k_N) will be set below the non-cooperative scenario. This is because under coalition 1, the two port regions internalize the negative externality on each other, as improving accessibility will definitely reduce the other port's profit due to price war. Under this coalition, the optimal investment rule for the inland region remains the same as in Section 4.4 by setting equation (4.23) equal zero. Assuming private ports, if $k_I V - d < 0$, the above results will still hold, however, if $k_I V - d > 0$, we can show with Proposition 4.2 that $\phi_B^{\pi B} < 0$ and $\phi_N^{\pi N} < 0$, implying that governments of port regions will investment than the non-cooperative scenario, given fixed investment levels of other players, because doing so will increase the welfare of the partner port region as well.

Coalition 2: region B and region I coordinate while region N remains independent

The social planners of regions *B* and *I* choose k_B and k_I together to maximize the joint welfare of these two regions. The joint welfare of regions *B* and *I* is

$$\phi^{BI}(k_B, k_N, k_I) \equiv \phi^B(k_B, k_N, k_I) + \phi^I(k_B, k_N, k_I) \ .$$

The optimal investment rule is characterized by:

$$\begin{aligned} \phi_B^{BI} &\equiv \partial\phi^B / \partial k_B + \partial\phi^I / \partial k_B = \phi_B^B + \phi_B^I = 0 \\ \phi_N^N &\equiv \partial\phi^N / \partial k_N = \phi_N^N = 0 \\ \phi_I^{BI} &\equiv \partial\phi^B / \partial k_I + \partial\phi^I / \partial k_I = \phi_I^B + \phi_I^I = 0 \end{aligned} \quad . \qquad (4.30)$$

From Propositions 4.1 and 4.2, we can derive that at equilibrium $\phi_B^{WB} < 0$ while $\phi_B^{\pi B} > 0$ if $k_I V - d > 0$ and $\phi_B^{\pi B} < 0$ if $k_I V - d < 0$. Therefore, given a fixed k_N and k_I, k_B will be set above the non-cooperative scenario if the ports maximize regional welfares. This is because under coalition 2, regions *B* and *I* internalize the positive impact of better infrastructure in region *B* on the surplus of shippers in region *I* due to lowered port charge. The same result holds if ports maximize profits and $k_I V - d < 0$. However, given private ports, if $k_I V - d > 0$, k_B will be set below the non-cooperative scenario, as increasing accessibility of region *B* will induce higher port charge and hence adversely affect region *I*'s welfare.

The sign of ϕ_I^I depends on the sign of $-\phi_I^B$, which is positive unless k_B is substantially larger than k_N when ports maximize regional welfares, as shown in Section 4.4. Thus, given fixed k_B and k_N, k_I will be set below the non-cooperative scenario unless region *B*'s accessibility is sufficiently better than region *N*. This is caused by taking into account the impact of increasing k_I on the profit of port *B*. The investment rule for region *N* remains the same as in the non-cooperative case. If ports maximize profits, the sign of $-\phi_I^{\pi B}$ is ambiguous and hence k_I can be higher or lower than the non-cooperative scenario.

Coalition 3: all three regions coordinate

The central planner decides k_B , k_N and k_I to maximize the total welfare across all the three regions. The total welfare of the three regions is

$$\phi^{BNI}(k_B,k_N,k_I) \equiv \phi^B(k_B,k_N,k_I) + \phi^N(k_B,k_N,k_I) + \phi^I(k_B,k_N,k_I)\,.$$

The optimal investment rule is characterized by:

$$\begin{aligned}
\phi_B^{BNI} &\equiv \partial\phi^B/\partial k_B + \partial\phi^N/\partial k_B + \partial\phi^I/\partial k_B = \phi_B^B + \phi_B^N + \phi_B^I = 0\\
\phi_N^{BNI} &\equiv \partial\phi^B/\partial k_N + \partial\phi^N/\partial k_N + \partial\phi^I/\partial k_N = \phi_N^B + \phi_N^N + \phi_N^I = 0\,,\\
\phi_I^{BNI} &\equiv \partial\phi^B/\partial k_I + \partial\phi^N/\partial k_I + \partial\phi^I/\partial k_I = \phi_I^B + \phi_I^N + \phi_I^I = 0
\end{aligned} \tag{4.31}$$

where

$$\phi_B^N + \phi_B^I = -(k_N + k_I)\,p^N p_B^N - (p^N k_I - Q_{BI})\,p_B^B\,, \tag{4.32}$$

$$\phi_N^B + \phi_N^I = -(k_B + k_I)\,p^B p_N^B - (p^B k_I - Q_{NI})\,p_N^N\,, \tag{4.33}$$

$$\begin{aligned}
\phi_I^B + \phi_I^N = {} & \left[Q_{BI} - (k_B + k_I)p^B\right]p_I^B + \left[Q_{NI} - (k_N + k_I)p^N\right]p_I^N\\
& + k_I(p^B p_I^N + p^N p_I^B) - (p^N - p^B)^2
\end{aligned}\,. \tag{4.34}$$

If ports are pubic and maximize regional welfare, we can rewrite equations (4.32), (4.33) and (4.34) as

$$\phi_B^{WN} + \phi_B^{WI} = \left(-d/2\Delta_W\right)p_B^{WB}(k_B k_N + 2k_B k_I + k_N k_I) - Q_{NI}\,p_B^{WN} > 0,$$

$$\phi_N^{WB} + \phi_N^{WI} = \left(-d/2\Delta_W\right)p_N^{WN}(k_B k_N + 2k_N k_I + k_B k_I) - Q_{BI}\,p_N^{WB} > 0,$$

$$\phi_I^{WB} + \phi_I^{WN} = k_I(p^{WB} p_I^{WN} + p^{WN} p_I^{WB}) - (p^{WN} - p^{WB})^2 < 0\,.$$

Note that though the effect of k_B on region *N*'s welfare is negative while that on region *I*'s welfare is positive, the positive impact on region *I* dominates and hence the net effect on those two regions is positive. Therefore, it is straightforward to show that given fixed k_N and k_I, the optimal k_B in coalition 3 is higher than the non-cooperative scenario. Note that 0 < $\phi_B^{WN} + \phi_B^{WI} < \phi_B^{WI}$ implies that given fixed k_N and k_I, ϕ_B^{WB} under coalition 3 is larger than ϕ_B^{WB} under coalition 2. Together with $\phi_{BB}^{B} < 0$, coalition 3 induces less infrastructure investment in region *B* than coalition 2. It is also easy to show that $\phi_I^{WI} > 0$ and hence given fixed k_N and k_B, the optimal k_I in coalition 3 is below the non-cooperative scenario. Similar analysis applies to the investment rule of region *N*.

If ports are private and maximize profits, equations (4.32), (4.33) and (4.34) reduce to:

$$\phi_B^{\pi N} + \phi_B^{\pi I} = (p_B^{\pi B}/2)(-d - k_I p^{\pi N} + 2k_I p^{\pi B}), \tag{4.35}$$

$$\phi_N^{\pi B} + \phi_N^{\pi I} = (p_N^{\pi N}/2)(-d - k_I p^{\pi B} + 2k_I p^{\pi N}), \tag{4.36}$$

$$\phi_I^{\pi B} + \phi_I^{\pi N} = (k_I p^{\pi B} - Q_{NN}) p_I^{\pi N} + (k_I p^{\pi N} - Q_{BB}) p_I^{\pi B} - (p^{\pi N} - p^{\pi B})^2. \tag{4.37}$$

The signs of above expressions depend on the sign of $k_I V - d$ as well as the magnitudes of port charges. In particular, taking (4.35) as an example, using Lemma 4.3, we can derive Table 4.1 which indicates conditions under which the optimal k_B is higher or lower than the non-cooperative scenario.

Table 4.1 The sign of $\phi_B^{\pi B}$

	$k_I V - d > 0$	$k_I V - d < 0$
$2p^{\pi B} - p^{\pi N} > d/k_I$	$\phi_B^{\pi B} < 0$	$\phi_B^{\pi B} > 0$
$2p^{\pi B} - p^{\pi N} < d/k_I$	$\phi_B^{\pi B} > 0$	$\phi_B^{\pi B} < 0$

A comparison between coalitions 2 and 3 together with Proposition 4.2 reveals that when $k_I V - d > 0$, $\phi_B^{\pi B}$ under coalition 2 is higher than that under coalition 3, suggesting that coalition 3 induces more infrastructure investment in the port region than coalition 2.

Nevertheless, applying the same logic, when $k_I V - d$ <0, coalition 3 induces less infrastructure investment in the port region than coalition 2. The sign of (4.37) is in general ambiguous. However, it is interesting to look into the situation of symmetric equilibrium where we assume that $c_B(\cdot) = c_N(\cdot) = c_H(\cdot)$. Then, equation (4.37) becomes

$$\phi_I^{\pi B} + \phi_I^{\pi N} = \left(p_I^{\pi H}(2k_H + 3k_I)/\Delta_\pi\right)\left((k_I + k_H)d - 2k_H^2 V\right) > 0 \text{ iff } (k_H + k_I)/k_H^2 < 2V/d \,.$$

That is, $\phi_I^{\pi l} < 0$ and hence the optimal k_I will be set above the non-cooperative level if and only if the accessibility of port regions is high enough.

Let *NC* denote non-cooperative case and let *C1*, *C2* and *C3* denote coalitions 1, 2 and 3, respectively. Comparing the investment rules of each region under these four cases, we reveal Propositions 4.4-4.6.

Proposition 4.4: *Assuming public ports, given fixed levels of* k_N *and* k_I, $k_B^{C1} < k_B^{NC} < k_B^{C3} < k_B^{C2}$. *That is, the infrastructure investment of a port region is the lowest if two port regions coordinate, followed by non-cooperative case, and both cases invest less than the social optimal level (coalition 3). If one port region coordinates with the inland region, this port region will overinvest in infrastructure.*

Proposition 4.5: *Assuming public ports, given fixed levels of* k_B *and* k_I, $k_N^{C1} < k_N^{NC} = k_N^{C2} < k_N^{C3}$. *That is, the infrastructure investment of a port region is the lowest if two port regions coordinate, followed by the cases that the port region does not coordinate with any other region and makes decision independently. All the three cases invest less than the social optimal level (coalition 3).*

Proposition 4.6: *Assuming public ports, given fixed levels of* k_B *and* k_N, $k_I^{C3} < k_I^{NC} = k_I^{C1} < k_I^{C2}$ *if* k_B *is substantially larger than* k_N *;* $k_I^{C3} < k_I^{C2} < k_I^{NC} = k_I^{C1}$ *otherwise. That is, the infrastructure investment of the inland region is the lowest if all the three regions coordinate, followed by the case of no coordination with inland region. If one port region coordinates with the inland region, the inland region may invest more or less than the non-cooperative* *case depending on the difference between* k_B *and* k_N .

One major implication of the above three propositions is that compared with the social optimum (coalition 3), the port regions are likely to under-invest in infrastructure accessibility while the inland region overinvest, given that full coordination among all the three regions is not achieved. The incentive of underinvestment by port regions comes from the ignorance of inland shippers' welfare improvement when port regions increase their infrastructure accessibility. The incentive of overinvestment by inland region comes from the ignorance of port regions' profit loss when inland region increases its infrastructure accessibility. This is especially the case for *NC* and *C1* where region *B* and region *N* are treated symmetrically. In coalition 2, however, where only one port region will coordinate with the inland region, the port region in collusion will overinvest while the other port region will under-invest.

Similar to the case of public ports, we obtain one proposition for each regional government's investment decision under the case of private ports.

Proposition 4.7: *Assuming private ports, given fixed levels of* k_N *and* k_I *, at equilibrium: (i)* $k_B^{C2} < k_B^{NC} < k_B^{C3} < k_B^{C1}$ *if* $k_I V - d > 0$ *and* $2p^{\pi B} - p^{\pi N} > d/k_I$ *; (ii)* $k_B^{C2} < k_B^{C3} < k_B^{NC} < k_B^{C1}$ *if* $k_I V - d > 0$ *and* $2p^{\pi B} - p^{\pi N} < d/k_I$ *; (iii)* $k_B^{C2} > k_B^{NC} > k_B^{C3} > k_B^{C1}$ *if* $k_I V - d < 0$ *and* $2p^{\pi B} - p^{\pi N} > d/k_I$ *; and (iv)* $k_B^{C2} > k_B^{C3} > k_B^{NC} > k_B^{C1}$ *if* $k_I V - d < 0$ *and* $2p^{\pi B} - p^{\pi N} < d/k_I$ *.*

Proposition 4.8: *Assuming private ports, given fixed levels of* k_B *and* k_I *, at equilibrium: (i)* $k_N^{NC} = k_N^{C2} < k_N^{C3} < k_N^{C1}$ *if* $k_I V - d > 0$ *and* $2p^{\pi N} - p^{\pi B} > d/k_I$ *; (ii)* $k_N^{C3} < k_N^{NC} = k_N^{C2} < k_N^{C1}$ *if* $k_I V - d > 0$ *and* $2p^{\pi N} - p^{\pi B} < d/k_I$ *; (iii)* $k_N^{NC} = k_N^{C2} > k_N^{C3} > k_N^{C1}$ *if* $k_I V - d < 0$ *and* $2p^{\pi N} - p^{\pi B} > d/k_I$ *; and (iv)* $k_N^{C3} > k_N^{NC} = k_N^{C2} > k_N^{C1}$ *if* $k_I V - d < 0$ *and* $2p^{\pi N} - p^{\pi B} < d/k_I$ *.*

Comparing these two propositions with those of public ports, we find that optimal investment decisions with private ports are much complicated. Considering the fully coordinated case as socially optimal, overinvestment and underinvestment will both occur based on various conditions. In general, when shippers' utility is high and the size of inland market is relatively small, coordination between two port regions (coalition 1) tends to overinvest in port regions' accessibility while coordination between one port region and the inland (coalition 2) will make the port region involved in the partnership under invest in its transport infrastructure. However, when shippers' utility is low and the size of inland market is relatively large, the opposite will hold.

Proposition 4.9: *Assuming private ports and* $c_B(\cdot) = c_N(\cdot)$, *given fixed levels of* $k_B = k_N$, *at equilibrium:* $k_I^{C3} > k_I^{NC} = k_I^{C1}$ *if port regions' accessibility is large enough;* $k_I^{C3} < k_I^{NC} = k_I^{C1}$ *otherwise.*

The implication of Proposition 4.9 is that there will be underinvestment in the inland transportation infrastructure compared to the fully coordinated case when the port regions' access condition is sufficiently good, because neither the non-cooperative case nor coalition 1 take into account the positive impact of investing in inland infrastructure on the port regions' welfares; otherwise, overinvestment in inland facility is likely to occur.

4.6 Concluding remarks

This study investigates the strategic investment decisions of local governments on inland transportation infrastructure in the context of seaport competition. In particular, we consider two seaports with their respective captive catchment areas and a common hinterland for which the seaports compete. The two seaports and the common hinterland belong to three independent local governments, each determining the level of investment for its own regional transportation system. This setting is different from any work in the literature in the sense that we consider not only two competing seaports but also the infrastructure decision of the common hinterland that the ports compete for. We study two different port ownerships, public ports which maximize regional welfare and private ports which maximize their profits. In particular, increasing investment in the common hinterland lowers charges of both competing ports. We find in most of the cases differentiated results for these two ownership types.

When ports are public, increasing investment in the captive catchment area of a certain port will cause more severe reduction in its port charge than that of the rival port. As a result, an increase in investment in the port region will reduce the welfare of the rival port region but improve the welfare of the common inland region. However, an increase in investment in the inland region will harm the port region with poorer accessibility. We also examine the non-cooperative optimal investment decisions made by local governments as well as the equilibrium investment levels under various coalitions of local governments. In general, for port regions, the incentive of infrastructure investment is the lowest when two port regions coordinate. They will invest more once at least one of them coordinates with the inland region. The inland region, on the other hand, always has high incentive to invest for low level of coordination.

When ports are private, provided that the size of the inland market is small and shippers' utility is high, additional investment in the captive catchment area of a certain port will cause more increase in its port charge than that of the rival port. As a result, at non-cooperative investment equilibrium, an increase in investment in the port region may raise the welfare of the rival port region while reduce the welfare of the common inland region. However, improved accessibility in the inland region will benefit the port regions if the port regions' accessibility is high enough. In terms of equilibrium investment levels under various coalitions, in general, when shippers' utility is high and the size of inland market is relatively small, coordination between two port regions tends to overinvest in port regions' accessibility while partnership between one port region and the inland will make the port region involved under invest in its transport infrastructure.

The present paper studies both private and public ports which can be considered as two polar cases. Port governance structure has being changing through various management reforms: the power of private sector in the port industry has been gradually increased in order to, among others, enhance operational efficiency and reduce the burden of public investment. Through the reform of port asset ownership and transfer of operational responsibility, complex forms of mixed ownership structure have emerged and evolved. Thus, a natural extension of this study is to examine mixed-ownership ports which maximize the weighted sum of regional welfare and port profit subject to a budget constraint. Furthermore, it would also be interest to investigate local governments' incentives to form various types of coalitions and predict with the theoretical model whether and in which forms coalition will occur. Issues such as schedule delay cost and congestion cost can also be incorporated into this model in the future.

A complete comparison with quantity competition can also be interesting. In chapter 2, we argue that quantity competition fits the case of port competition better because port investment is lumpy and irreversible and can be considered as a commitment on quantity. In the present chapter, however, we assume ports compete in prices, because we apply the linear city model in this study which generates concise demand functions but very complicated inverse demand functions. The complicated inverse demand functions will cause the best respons functions of both ports to shift when one region's accessibility improves. To make our analysis tractable, price competition is assumed. However, we have double checked some of our basic results with quantity competition and find that in general the main results are similar. For example, in the case of public ports, if one port region improves accessibility, similar to price competition, both ports will reduce prices and such accessibility improvement will make the welfare of the rival region worse off while the inland region better off. Thus, our inference is that quantity competition should produce qualitatively similar results.

5 Airport Pricing with Concession Revenues and Heterogeneous Passengers[49]

5.1 Introduction

Air travel delay has been growing dramatically since the end of the 1990s. The delay problem has been widely discussed in policy circles: increasing the capacity of congested airports by investing in new runways or improving air traffic control technology is one possible remedy. Another solution is the imposition of congestion pricing, according to which the landing fees paid by airlines would vary with the level of congestion at the airport. Meanwhile, non-aeronautical revenues have been growing significantly to the point that they have become the main income source for many airports (Graham, 2009; Morrison, 2009). For these reasons, the impact of non-aeronautical revenues on airport pricing is of increasing concern for airport and airline management.

With respect to the issue of airport congestion pricing, literature finds a negative relationship between the socially optimal airport charge and airlines' market concentration (Basso, 2008; Basso and Zhang, 2007; Brueckner, 2002; Brueckner and Van Dender, 2008; Pels and Verhoef, 2004; Zhang and Zhang, 2006). The socially optimal charge should include only the residual share of the marginal external congestion cost (*MEC*) that is not internalized by monopoly or oligopoly carriers[50] and this amount should be further reduced to correct for market power of airlines. On the other hand, concession revenues exert a downward pressure on the aeronautical charge (Oum *et al.*, 2004; Starkie, 2002, 2008; Yang and Zhang, 2011; Zhang and Zhang, 2003, 2010). Commercial operations tend to be more profitable than aeronautical operations (Jones *et al.*, 1993; Starkie, 2001); therefore, the aeronautical charge should be reduced so as to induce a higher volume of passengers and increase the demand for concessions.

However, in order to have a more complete picture of optimal airport pricing, two more

[49] A version of chapter 5 has been published. D'Alfonso, T., Jiang, C. and Wan, Y. (2013) Airport pricing, concession revenues and passenger types, *Journal of Transport Economics and Policy*, 47(1), 71-89.

[50] For a certain carrier, its residual share is equal to one minus the carrier's market share.

aspects of the air transport business should be incorporated into the analysis.

First, passengers may not be a homogeneous group of individuals. Literature finds that, in the case of a single passenger type, the socially optimal charge never exceeds the residual share of the marginal external congestion cost (Basso and Zhang, 2007; Brueckner, 2002; Zhang and Zhang, 2006). Czerny and Zhang (2011) find that, in the case of two types of passengers with different values of time, the socially efficient airport charge may exceed the residual share of the marginal external congestion cost. Intuitively, their result implies that it can be useful to increase airport charge so as to protect business passengers with higher time value from excessive congestion caused by leisure passengers with lower time value.

Second, there is a positive correlation between the expenditure in the concessions area and the dwell time, that is, the time available between the security check and the boarding: it is during that time that passengers will have higher chance to shop. This follows the common sense that more spare time gives more opportunity for browsing in the shops and induces the need to buy refreshment. Hence, the expenditure increases as the dwell time increases. Congestion level may have an impact on the dwell time, and therefore on the expenditure in the commercial area; but, without solid empirical studies in the literature, it is unclear whether increased congestion has a negative or positive effect. The higher the volume of passengers the longer the time needed for check-in and security check. As a result, on one hand, it would be obvious that dwell time decreases as congestion goes up, since passengers spend more time in queues. However, on the other hand, higher congestion may force travelers to arrive in advance at airport terminals because they anticipate longer waiting time in queues (Appold and Kasarda, 2006; Buendia and de Barros, 2008). This can happen when air travelers are risk averse, especially when the cost of missing a flight is relatively high: business passengers may miss important business opportunities; leisure passengers may have to cancel hotel and trip reservations whose costs cannot be fully recovered. In this context, if this amount of extra time they spend in the airport is disproportionally longer than the expected extra time they need to go through check-in and security checks, dwell time will increase: passengers will have more captive time in terminals and more time to spend money in shops. The above argument applies to originating pasengers. Congestion affects domestic connecting passengers in a different way as these passengers in general do not need to go through security screening again when transiting at the airport. Congestion may cause late arrival of the preceding flight and hence the connecting passenger may miss the succeeding flight and spend more time in the airport until the next flight is available. Some risk-averse connecting passengers may choose a longer time interval for transitting so as to avoid missing flights. Even if the arrival flight is on time, the succeeding flight can be delayed as well. As long as passenger boarding time is delayed, dwell time will increase.[51] Specifically, in this paper, we *assume* that passengers will exaggerate waiting time and therefore dwell time increases. In other words, we *assume* that as congestion increases dwell time increases and so the money spent in concession activities; equivalently, that there is a positive externality of congestion on concession activities. Hence, under this assumption, when concessions are taken into account, there can be some incentives for the airport to increase congestion in order to drive up the expenditure in the commercial area.

There is a stream of empirical literature trying to explore this issue. Geuens *et al.* (2004) find that waiting time influences consumption of concession goods. Castillo-Manzana (2010) finds that the dwell time prior to embarking is positively correlated with the decisions of consuming food/beverages and making a purchase at a significance level of 99 percent in both cases. Besides, he finds that being on vacation increases the likelihood of consuming concession goods. Moreover, the average expenditure of these passengers is greater than that of business passengers. Torres *et al.* (2005) show that the more time spent in the airport, the more consumption made by passengers. In addition, he finds that those flying on business consume more than those on vacation, if they are in the airport for less than 45 minutes. In the range of 45–170 minutes, leisure travelers consume more. When staying longer than 170 minutes, business travelers consume more. Graham (2008) finds that young leisure passengers are high spenders, while business passengers are unlikely shoppers.[52] However, to the best of our knowledge, there is no contribution in literature analyzing, from a

[51] This is true for both originating and connecting passengers.
[52] One possible reason for less spending of business passengers on concession goods is that business passengers spend surplus time in airport lounges. They could consume more if there is no lounge in the airport. However, if airports compete fiercely, lounge facility will be crucial to attract business passengers. In the present study, we take business passengers' consumption behavior (as a function of congestion delay) as exogenously given and leave behind the impact of offering lounge facility.

theoretical point of view, the effects of congestion and passenger types on consumption of concession goods.

This paper adds to literature on airport pricing as it takes into account the positive externality of congestion on concessions, through its impact on dwell time, while incorporating the effect of passenger types. Specifically, we consider a model with one congestible airport serving a number of competing airlines and two types of passengers, business and leisure, with the former having a higher time value than the latter. We consider two types of airports, namely private airports maximizing their profits and public airports maximizing social welfare. We assume that only the extra surplus generated by airport concession services not attainable elsewhere is counted into the social welfare function. In other words, we only include a proportion of the surplus from concession services. This reconciles two approaches to modeling the social welfare function in airport pricing literature: if the proportion is equal to 1, all the surplus from concession activities is counted into social welfare (Yang and Zhang, 2011; Zhang and Zhang, 2003, 2010); if the proportion is equal to 0, surplus from concession activities is excluded (Czerny, 2011; Kratzsch and Sieg, 2011).

We find that for both profit and welfare maximizing airports there is a downward correction for the congestion toll, equal to the marginal airport concession profit and passengers' concession surplus, respectively, due to the positive externality of delay. Furthermore, as the passenger volume changes when the airport charge increases, there is a correction on the optimal airport charge equal to the average concession profit and expected concession surplus – for profit and welfare maximizing airports respectively – weighted for different passenger types. For some levels of delay this correction may not be a traditional mark-down but a mark-up. Finally, the comparison between privately and socially optimal airport charges shows that: (i) when concessions generate a sufficiently high proportion of extra surplus to total concession surplus, the welfare maximizing airport can have more incentives than the profit maximizing airport to decrease the congestion toll and induce delay; and (ii) depending on the difference in the passengers' values of time and the proportion of extra surplus generated by airport concessions, the profit maximizing airport may or may not impose a higher charge than the welfare maximizing airport.

The structure of the paper is as follows. Section 5.2 sets up the model. Section 5.3 and 5.4 discuss, respectively, airlines' and airport's equilibrium behaviors. Section 5.5 contains the concluding remarks.

5.2 The model

Consider a single airport, *n* competing airlines and two types of passengers, one of which has a higher time value than the other. For sake of convenience, in our analysis we refer to them as business and leisure passengers, because Morrison (1987) and Pels *et al.* (2003), among others, provide empirical evidence that business passengers have a greater value of time than leisure passengers. We denote the business and leisure passengers' value of time as v_B and v_L, respectively, with $v_B \geq v_L > 0$. Let Q_B and Q_L be the number of business and leisure passengers at the airport. For analytical tractability, we assume linear demand functions, which give

$$\rho_h(Q_h) \equiv a_h - b_h Q_h, \tag{5.1}$$

where $a_B \geq a_L > 0$, that is, the willingness to pay of business passengers for air travel is greater than that of leisure passengers; and $b_B \geq b_L > 0$, that is, the leisure passengers are more price sensitive than business passengers. The airport is congestible: the average congestion delay, $D(\tilde{Q}, K)$, depends on the total number of flights, $\tilde{Q}$, and the airport's capacity, *K*. With these specifications, at equilibrium, the (inverse) demand function must equal to the 'full price' paid by passengers:

$$\rho_h(Q_h) = p_h + v_h D(\tilde{Q}, K), \quad h \in \{B, L\} \tag{5.2}$$

where p_h is the airline ticket price for type *h* passengers. In a word, we assume that passengers make travel decisions solely based on the airfare and the expected time cost of airport congestion and hence concession demand is induced by travel. One may argue that passengers who expect to spend time eating in the airport should take into account the costs of food and beverage when purchasing the air tickets. However, it seems more appropriate to exclude food expenditure from the travel demand function for a few reasons: first, it is difficult for a non-frequent traveler to predict the price and the type of food available at terminal when purchasing an air ticket; second, even though a frequent flyer can have some knowledge on the provision of catering service in the airport, as one has to eat no matter he travels or not, only the difference between the amounts he normally spends on food and the expense on food in the airport may play a role. This difference if any is negligible compared with the air ticket price and passengers who are extremely sensible to food price will bring their own food to the airport. It is also possible that a passenger deliberately plans to purchase certain goods in the airport when planning his trip. Such cases may occur in some airports (e.g. Hong Kong) which market themselves as shopping complexes, but in general these cases are less likely to occur as long as there is no obvious advantage in shopping in the airport compared with shopping elsewhere. We use the same linear delay function as the one in Basso and Zhang (2007) and De Borger and Van Dender (2006).[53] That is, $D(\tilde{Q}, K) = \theta(\tilde{Q}/K)$, where θ is a positive parameter. Specifically, let Q be the number of passengers of all airlines. We assume, as is common in the airport pricing literature, a fixed proportion condition. That is, all the flights use identical aircraft and have the same load factor (Basso, 2008; Basso and Zhang, 2007; Brueckner, 2002; Pels and Verhoef, 2004; Zhang and Zhang, 2006, 2010). Therefore, each flight has an equal number of passengers, denoted by *S*. Then, $Q/S = \tilde{Q}$ and we obtain

$$D(\tilde{Q}, K) = \theta \frac{Q}{KS}. \tag{5.3}$$

Furthermore, without loss of generality, we normalize $KS = 1$. Therefore, we can use, in what follows, $D(Q)$ instead of $D(\tilde{Q}, K)$. From (5.1)-(5.3), it follows that

[53] Such a linear delay function makes the analytical work more feasible, but it may lead to the problem that an interior solution may not exist, that is we may have a corner solution. Nevertheless, we assume an interior solution. The purpose of using a linear delay function is to simplify the notations. Our results will hold even if no functional form of congestion delay is imposed, i.e. *D* is a general function of *Q* and *K*, is employed.

$$p_h(Q_B, Q_L) = a_h - b_h Q_h - v_h \theta Q. \tag{5.4}$$

Carriers are ex ante symmetric and offer a homogeneous good/service, that is, the flight. Let q_h^i denote the number of type *h* passengers served by airline *i*, for $h \in \{B, L\}$ and $i = 1,2,\ldots,n$. Let q^i be airline *i*'s output, that is, the total number of passengers who fly with airline *i*. Therefore, $q^i = \sum_{h \in \{B,L\}} q_h^i$, $Q_h = \sum_{i=1}^{n} q_h^i$, for $h \in \{B, L\}$, and $Q = \sum_{h \in \{B,L\}} Q_h = \sum_{h \in \{B,L\}} \sum_{i=1}^{n} q_h^i$.

Next, we specify the passengers' demand for concessions. In particular, we assume that demand for retail services depends on travel activities. In other words, we suppose that passengers make two separate decisions sequentially. First, they book the air tickets from the airlines, based on their perceived full prices; second, after arriving at the airport, they make decisions on purchasing concession goods. Our specification of the concession demand is related to, but different from, Yang and Zhang (2011), according to whom a passenger will consume one unit of the concession goods if her valuation is greater than the concession price. We suppose that the passengers' valuation for the concession goods has a positive support on the interval $[0, \bar{u}]$, where $\bar{u}$ is the highest valuation for the concession goods. We consider two random variables, u_B and u_L, representing, respectively, the valuations for the concession goods of business passengers and leisure passengers. We assume that u_h is distributed with probability density function $g_h(u;T)$, given a specific level of dwell time, *T*. As mentioned in the introduction, we assume that as congestion increases dwell time increases as well because passengers will exaggerate waiting time. Equivalently, we assume that the dwell time, $T = T(D)$, is an increasing function of congestion. Therefore, we can use, in what follows, $g_h(u;D)$ instead of $g_h(u;T(D))$. Let $G_h(u;D)$ be the cumulative distribution function of type *h* passengers' valuation. In this scenario, the probability that a type *h* passenger buys the concession goods at the price p_c is equal to the probability that her valuation for the good is greater than p_c, that is, $\int_{p_c}^{u_h} g_h(p;D)dp = \overline{G}_h(p_c;D)$, where $\overline{G}_h(u;D) = 1 - G_h(u;D)$. With this setup we want to catch the relationship between congestion and the probability of purchasing, through the dwell time. It is possible that at some point concession revenues are adversely affected by congestion and waiting time: firstly, congestion may reduce the comfort level of shopping, affecting patronage of shops and restaurants; secondly, it may increase the stress level of passengers, that is, passengers may get unnerved by waiting. A congested airport may simply not make the passengers relaxed enough to shop (Graham, 2009). On the other hand, for some people waiting may cause annoyance just leading them to search for comfort from shopping. In this paper, we assume that the impact of people finding relaxation in shopping is enough to offset that of unnerved passengers, that is, the extra dwell time leads to more retail activity. This is equivalent to assume that the probability of purchasing increases as the delay increases. In other words, $G_h(u;D)$ satisfies the first order stochastic dominance property (FOSD) with respect to *D*, that is, $\partial G_h(u;D)/\partial D \leq 0$, with a strict inequality for some value of *u*.[54] From the FOSD property, we have that $\overline{G}_h(p_c;D) \geq \overline{G}_h(p_c;\tilde{D})$, $\forall D > \tilde{D}$, that is, the probability of purchasing a unit of concession goods increases with the delay. We further assume that the positive externality of delay decreases when the concession price increases, that is $\partial^2 \overline{G}_h(p_c;D)/\partial p_c \partial D < 0$. Therefore, the concession demand function of the type *h* passengers, x_h, is given by

$$x_h(p_c, Q_h, Q_{-h}) = Q_h \overline{G}_h\big(p_c; D(Q_h, Q_{-h})\big). \tag{5.5}$$

In other words, the demand for non-aviation activities of type *h* passengers depends on the number of type *h* travelers, Q_h, the concession price, p_c, and the delay, $D(Q_h, Q_{-h})$.

The airport charges airlines a price per passenger, denoted as τ. For simplicity of presentation, the case where the airport has zero fixed costs is considered, that is, the only cost the airport bears is the operating cost per passenger[55], c_a. Since we consider ex ante

[54] This property means that for all $\tilde{u} \in [0, \overline{u}]$ the probability that $u \leq \tilde{u}$ is weakly and sometimes strictly decreasing in delay, that is, $g_h(\cdot;D)$ shifts rightward when delay increases.

[55] The qualitative results of this analysis, however, are unchanged since we assume there are no economies of scale and economies of scope.

symmetric carriers, the cost function of carrier *i* is given by

$$C^i(q^i, q^{-i}) = \left(c + \tau + \omega D(Q)\right)q^i, \tag{5.6}$$

where c is the (constant) marginal operating cost and ω is the value of time of carriers. Suppose that the airport provides concessions to (homogeneous) retailers and that the airport itself determines the concession price p_c. Finally, we assume that the airport captures all the rents from the retailers and that the unit cost of the concession goods is constant and denoted by c_c.[56]

The airport-airline vertical structure is modeled as a two stages game. In the first stage, the airport decides both the aeronautical charge, τ, and the concession price, p_c. In the second stage, taking τ as given, airlines compete in Cournot fashion[57] and simultaneously choose their outputs, that is, the number of passengers.

5.3 Airlines' equilibrium behavior

In the second stage, each airline chooses its output to maximize its profit:

$$\pi^i = \sum_{h \in \{B,L\}} q_h^i \left[p_h(Q_B, Q_L) - c - \tau - \omega D(Q)\right]. \tag{5.7}$$

To focus on the effect of the positive externality of congestion, we abstract away the possibility of price discrimination: all passengers pay a uniform airfare, *p*. Therefore, at the

---

[56] The present study assumes that lessees are perfectly competitive so that the airport can extract all the rents. In reality the amount of rents that the airport will obtain will depend on how airport is regulated, e.g. single-till or dual-till regulation, as well as the contract between airport and airlines or concession lessees. These complicated issues are prominent future extensions.

[57] Earlier studies that model a congestible airport serving air carriers with market power assume Cournot behavior (Basso and Zhang, 2007; Brueckner, 2002; Czerny, 2006; Pels and Verhoef, 2004; Zhang and Zhang 2006; 2010). Brander and Zhang (1990, 1993) find that the Cournot model seems much more consistent with the data than either the Bertrand or the cartel model. On the other hand, Neven *et al.* (1999) provide evidence that the estimated conduct in the airline market is not consistent with Cournot, but with Bertrand. However, there is a theoretical justification for assuming Cournot behavior: if firms first make pre-commitment of quantity, and then compete in prices, the equilibrium outcome will be equivalent to that of Cournot competition (Kreps and Scheinkman, 1983).

equilibrium, the condition $p_B = p_L = p$ must be satisfied. That is,

$$p(Q_B, Q_L) = a_h - b_h Q_h - v_h \theta Q \,. \tag{5.8}$$

Then, $p(Q_B, Q_L)$ can be written as a function of Q:

$$p(Q) = \frac{a_B b_L + a_L b_B}{b_L + b_B} - \frac{b_L b_B}{b_L + b_B} Q - \frac{b_L v_B + b_B v_L}{b_L + b_B} \theta Q \,. \tag{5.9}$$

The equilibrium outputs are determined by the first-order conditions:

$$\frac{\partial \pi^i}{\partial q_h^i} = p + \left(\frac{\partial p}{\partial Q} - \omega\theta \right) q^i - \omega\theta Q - c - \tau = 0, \qquad \forall i, h \,. \tag{5.10}$$

Symmetry implies that

$$p + \frac{1}{n}\frac{p}{\varepsilon} - \left(1 + \frac{1}{n}\right)\omega\theta Q - c - \tau = 0 \,, \tag{5.11}$$

where $Q/n = q^i$ and $\varepsilon = -(\partial Q/\partial p)(p/Q)$ is the elasticity of demand for airline services with respect to the ticket price. Equation (5.11) can be rewritten as:

$$p = c + \tau + \omega\theta Q + \left(\frac{b_L v_B + b_B v_L}{b_L + b_B} + \omega \right) \frac{\theta Q}{n} + \left(\frac{b_L b_B}{b_L + b_B} \right) \frac{Q}{n} \,. \tag{5.12}$$

The last component on the right-hand side (RHS) of (5.12) is the airline's mark-up due to market power. The fourth component on the RHS of (5.12) is the part of *MEC* internalized by an individual airline, where $MEC = (v_B Q_B + v_L Q_L + \omega Q) D'(Q) = (v_B Q_B + v_L Q_L + \omega Q)\theta$. From equations (5.8) and (5.9), we can write Q_h as an expression of Q. Then, *MEC* can be rewritten as:

$$\left(\frac{(v_B - v_L)\left[(a_B - a_L) - (v_B - v_L)\theta Q\right]}{(b_L + b_B)Q} + \frac{b_L v_B + b_B v_L}{b_L + b_B} + \omega \right) \theta Q \,.$$

Therefore, the amount of MEC internalized by an airline equals to $(1/n)MEC$ only when passengers' values of time are the same, which is consistent with the literature. However, when the values of times differ between passenger types, an airline can internalized more or less than $(1/n)MEC$, depending on the difference of the time values. The larger the difference, the less MEC will be internalized.

The effect of the ticket price p on Q, Q_B and Q_L is summarized in Lemma 5.1.

Lemma 5.1: *Under the linear demand specification, we have* $dQ/dp < 0$, $dQ_L/dp < 0$, *while the sign of* dQ_B/dp *is ambiguous.*

Proof: Differentiating equation (5.9) on both sides with respect to the ticket price p, we have:

$$\frac{dQ_L}{dp} = \frac{1}{\psi}\left[-b_B - (v_B - v_L)\theta\right],$$

$$\frac{dQ_B}{dp} = \frac{1}{\psi}\left[-b_L + (v_B - v_L)\theta\right],$$

$$\frac{dQ}{dp} = \frac{dQ_B}{dp} + \frac{dQ_L}{dp} = \frac{1}{\psi}\left[-b_B - b_L\right],$$

where $\psi = [b_B b_L + (v_B b_L + v_L b_B)\theta] > 0$. Since $v_B > v_L$, we obtain $dQ_L/dp < 0$, while the sign of dQ_B/dp is undetermined. Since $b_B > 0$ and $b_L > 0$, we obtain $dQ/dp < 0$. *Q.E.D.*

Therefore, an increase in the ticket price leads to a decrease in the total number of passengers and the number of leisure passengers, but it can lead to an increase or a decrease in the number of business passengers. Let $Q^*(\tau)$ denote the equilibrium total number of passengers, $Q_B{}^*(\tau)$ the equilibrium number of business passengers, $Q_L{}^*(\tau)$ the equilibrium number of leisure passengers and $p^*(\tau)$ the equilibrium airline ticket price. The comparative static of these equilibrium outcomes with respect to the airport charge, τ, is summarized in Lemma 5.2.

Lemma 5.2: *Under the linear demand specification, we have* $dp^*/d\tau > 0$, $dQ^*/d\tau < 0$, $dQ_L^*/d\tau < 0$, *while the sign of* $dQ_B^*/d\tau$ *is ambiguous.*

Proof: Differentiating equation (5.10) on both sides with respect to τ, we have:

$$\frac{dQ^*}{d\tau} = \frac{n}{(1+n)\dfrac{dp}{dQ} + Q\dfrac{d^2 p}{dQ^2} - (1+n)\omega\theta} < 0,$$

where $d^2 p/dQ^2 = 0$, as the inverse demand for air travel is linear and dp/dQ = $1/(dQ/dp) < 0$ from Lemma 5.1. Therefore,

$$\frac{dp^*}{d\tau} = \frac{dp}{dQ}\frac{dQ^*}{d\tau} > 0 .$$

From equations (5.9) and (5.10) we derive:

$$q_L^{i*} = \frac{(a_L - c - \tau)((1+n)(b_B + \theta(\omega + v_B))) - (a_B - c - \tau)(n\theta(\omega + v_L) + \theta(\omega + v_B))}{H},$$

$$q_B^{i*} = \frac{(a_B - c - \tau)((1+n)(b_L + \theta(\omega + v_L))) - (a_L - c - \tau)(n\theta(\omega + v_B) + \theta(\omega + v_L))}{H},$$

where

$$\begin{aligned} H &= (1+n)^2 (b_B + \theta(\omega + v_B))(b_L + \theta(\omega + v_L)) \\ &- (n\theta(\omega + v_B) + \theta(\omega + v_L))(n\theta(\omega + v_L) + \theta(\omega + v_B)) \end{aligned}.$$

Therefore we obtain:

$$\frac{dQ_L^*}{d\tau} = n\frac{dq_L^{i*}}{d\tau} = \frac{n}{H}[n\theta(v_L - v_B) - b_B(1+n)],$$

$$\frac{dQ_B^*}{d\tau} = n\frac{dq_B^{i*}}{d\tau} = \frac{n}{H}\left[n\theta(v_B - v_L) - b_L(1+n)\right].$$

From the concavity condition of airlines' profit function, we derive:

$$\pi_{BB}^i \pi_{LL}^i - \pi_{LB}^i \pi_{BL}^i = 4(b_B + \theta(\omega + v_B))(b_L + \theta(\omega + v_L)) - (\theta(\omega + v_B) + \theta(\omega + v_L))^2 > 0$$

with $\pi_{BB}^i \pi_{LL}^i - \pi_{LB}^i \pi_{BL}^i \,|_{n=1} = H \,|_{n=1} = \partial H/\partial n \,|_{n=1}$. Therefore, when n = 1, H > 0. Moreover,

$$\frac{\partial^2 H}{\partial n^2} = 2(b_B + \theta(\omega + v_B))(b_L + \theta(\omega + v_L)) > 0,$$

that is, $\partial H/\partial n$ is an increasing function of n. Therefore, $H > 0 \; \forall n \geq 1$. Since $v_B > v_L$, we have that $dQ_L^*/d\tau < 0$ but the sign for $dQ_B^*/d\tau$ is undetermined. *Q.E.D.*

Therefore, an increase in the airport charge leads to a decrease in the equilibrium total number of passengers and the number of leisure passengers, an increase in the equilibrium airlines ticket price but it can lead to an increase or a decrease in the equilibrium number of business passengers.

5.4 Airport pricing

Taking the second stage airlines behavior into account, the airport chooses p_c, the concession price, and τ, the charge for airlines. We consider two types of airports, namely a private airport which maximizes its profit and a public airport which is a welfare maximizer.

5.4.1 Profit maximizing airport

Consider a private airport maximizing its profit:

$$\pi_A = (\tau - c_a)Q + (p_c - c_c)\sum_{h\in\{B,L\}} Q_h \overline{G}_h(p_c; D(Q)). \tag{5.13}$$

The optimal concession price is characterized by the first-order condition with respective to

p_c:

$$p_c^{\pi} = c_c - \frac{Q_B^* \overline{G}_B\left(p_c^{\pi}; D(Q^*)\right) + Q_L^* \overline{G}_L\left(p_c^{\pi}; D(Q^*)\right)}{Q_B^* \dfrac{\partial \overline{G}_B\left(p_c; D(Q^*)\right)}{\partial p_c} + Q_L^* \dfrac{\partial \overline{G}_L\left(p_c; D(Q^*)\right)}{\partial p_c}}, \tag{5.14}$$

where the superscript π represents the profit maximization case. Since $\partial \overline{G}_B\left(p_c; D(Q^*)\right) / p_c$ < 0 with $h \in \{B, L\}$, a profit maximizing airport sets the optimal concession price above the marginal concession cost and, in particular, equal to the monopoly price. The profit maximizing airport charge is characterized by the first-order condition with respective to τ:

$$\begin{aligned} \tau^{\pi} - c_a = & \left(\frac{b_L v_B + b_B v_L}{b_L + b_B} + \omega\right)\left(1 + \frac{1}{n}\right)\theta Q^* + \left(\frac{b_L b_B}{b_L + b_B}\right)\left(1 + \frac{1}{n}\right) Q^* \\ & - \theta\left(p_c^{\pi} - c_c\right) \sum_{h \in \{B, L\}} Q_h^* \frac{\partial \overline{G}_h\left(p_c^{\pi}; D(Q^*)\right)}{\partial D} \\ & - \left(p_c^{\pi} - c_c\right) \frac{dp}{dQ} \sum_{h \in \{B, L\}} \frac{dQ_h}{dp} \overline{G}_h\left(p_c^{\pi}; D(Q^*)\right) \end{aligned} . \tag{5.15}$$

The first line on the RHS of equation (5.15) can be reduced to the results in earlier literature where only one passenger type is considered (Zhang and Zhang, 2006). The second line consists of two terms which are the focus of this paper. The first term is a correction for the congestion toll equal to the marginal airport concession profit due to the positive externality of congestion on concession activities. Since $\partial \overline{G}_h\left(p_c; D(Q)\right) / \partial D > 0$, this term is negative. Therefore, the airport has incentives to reduce the congestion toll so as to increase the passenger volume and the passengers' waiting time. This means that, in contrast with previous literature, the congestion toll may become a 'subsidy', when the positive externality of congestion is taken into account. The above discussion leads to Proposition 5.1.

Proposition 5.1: *In the case of a profit maximizing airport, there is a downward correction for the congestion toll which is equivalent to the marginal concession profit due to the positive externality of delay. Therefore, the airport has incentives to reduce the aeronautical charge so as to increase passengers' waiting time and so their consumption of concession* *goods.*

The last term is a correction on the optimal airport charge equal to the per passenger concession profit weighted for different passenger types, where the weight is the ratio of the marginal change in the number of type *h* passengers over the marginal change in the total number of passengers. This term takes into account the change in the passenger volume and hence the pool of potential consumers of concession services when the airport charge increases. When passengers have the same value of time, this term is always negative as shown in previous literature (for example, Yang and Zhang, 2011; Zhang and Zhang, 2010), but the sign of this term is no longer clear-cut when more than one type of passengers is considered. In particular, when $dQ_B/dp > 0$, that is, $(v_B - v_L)\theta > b_L$ (see Lemma 5.1), and

$$\frac{\overline{G}_L\left(p_c^{\pi};D(Q^*)\right)}{\overline{G}_B\left(p_c^{\pi};D(Q^*)\right)} < -\frac{dQ_B/dp}{dQ_L/dp}, \tag{5.16}$$

it becomes positive, that is, a mark-up on the privately optimal airport charge. Specifically, $\overline{G}_h\left(p_c^{\pi};D(Q^*)\right)$ represents the probability of purchasing the concession good for type *h* passengers when the concession price is p_c^{π}. Therefore, when this probability is sufficiently higher for business passengers than for leisure passengers, inequality (5.16) is satisfied and the last term on the RHS of equation (5.15) is a mark-up on the airport charge. According to Torres *et al.* (2005), those flying on business can consume more than those on vacation under high levels of delay. Therefore, for these levels of delay the private airport can have incentives to induce more business passengers with higher time value - and let them buy in the commercial area – by protecting them from excessive congestion caused by leisure passengers with lower time value. This leads to

Observation 5.1: *In the case of a profit maximizing airport and two types of travelers, for some levels of delay the correction on the optimal airport charge due to the impact of the changes in passenger volume on concession profit may not be a traditional mark-down but a mark-up. Therefore, the privately optimal airport charge can be higher than what would prevail if passengers are treated as a single type.*

In summary, whenever we consider the positive externality of congestion alone, there is always a downward correction on the congestion toll to exploit the higher probability of purchasing induced by longer waiting time and a mark-down to increase the pool of potential consumers for concession goods. On the other hand, if, in addition, we consider two types of travelers, resulting from a trade-off between business and leisure passengers, the aforementioned mark-down may become a mark-up. Intuitively, such a mark-up is likely to occur when the level of delay is high.

5.4.2 Welfare maximizing airport

Consider a public airport whose mandate is to maximize social welfare (SW). It is the sum of two parts, namely, surplus from aeronautical services, S^a, and a proportion, $\delta \in [0,1]$, of the surplus from concession services, S^c, which are given by

$$S^a = \int_0^{Q_B} \rho_B(y)dy + \int_0^{Q_L} \rho_L(y)dy - \theta Q(v_B Q_B + v_L Q_L) - \theta\omega Q^2 - (c + c_a)Q$$

and

$$S^c = \sum_{h\in\{B,L\}} \int_{p_c}^{\bar{u}} Q_h \overline{G}_h(z; D(Q))dz + (p_c - c_c) \sum_{h\in\{B,L\}} Q_h \overline{G}_h(p_c; D(Q)).$$

In our formulation, if $\delta = 1$, all the surplus generated by the concession services is extra surplus (that is, surplus which is unattainable elsewhere), which is commonly assumed in the literature (Yang and Zhang, 2011; Zhang and Zhang 2003, 2010). If $0 < \delta < 1$, only part of the concession surplus is extra surplus. If $\delta = 0$, none of the concession services generate extra surplus (Czerny, 2011; Kratzsch and Sieg, 2011). The reason why only a proportion, δ, of the surplus from concession services is counted into the social welfare function is that only under certain occasions concession services generate extra surplus. In other words, a difference may exist between the types of concession services at the airport. For example, the overall demand for food and beverages may not depend much on whether individuals fly or not fly. On the other hand, there are some other types of concession services which may be elicited by travel-related motivations. Geuens *et al.* (2004) find that there are specificities for airport shopping, such as motivation 'to contrast day-to-day' and 'to be out of place'. Several authors agree that the shopping and purchasing habits of a tourist often vary considerably from her normal pattern at home (Brown, 1992; Huang and Kuai, 2006). Another motivation is that travelers leaving a certain country shop in order to spend their remaining foreign currencies. Furthermore, the habit of buying souvenirs and presents motivates travelers to shop (Sulzmaier, 2001). Large international brands design new product lines exclusively for duty-free shops in order to seduce travelers to buy a unique souvenir (Vlitos-Rowe, 1999). Moreover, for some people traveling causes fear or feelings of insecurity, leading them to search for comforting and reassuring behaviors from shopping (Dube and Menon, 2000).

As a result, the social welfare function can be written as follows.

$$\begin{aligned} SW = & \sum_{h\in\{B,L\}} \int_0^{Q_h} \rho_h(y)dy - \theta Q \sum_{h\in\{B,L\}} v_h Q_h - \theta\omega Q^2 - (c+c_a)Q \\ & + \delta \sum_{h\in\{B,L\}} \int_{p_c}^{\bar{u}} Q_h \overline{G}_h(z;D(Q))dz + \delta(p_c - c_c) \sum_{h\in\{B,L\}} Q_h \overline{G}_h(p_c;D(Q)) \end{aligned} . \tag{5.17}$$

The airport maximizes social welfare with respect to p_c, the concession price, and τ, the charge for airlines. The first-order condition with respective to the concession price is

$$\frac{\partial SW}{\partial p_c} = (p_c - c_c) \sum_{h\in\{B,L\}} Q_h^* \frac{\partial \overline{G}_h\left(p_c; D(Q^*)\right)}{\partial p_c} = 0 . \tag{5.18}$$

Equation (5.18) is only satisfied when

$$p_c^W = c_c ,$$

where the superscript W is used to denote results for the welfare maximization case. Therefore, for a welfare maximizing airport, the optimal concession price is equal to the marginal concession cost. The welfare maximizing airport charge is characterized by

$$\frac{\partial SW}{\partial \tau} = \frac{\partial SW}{\partial p} \frac{dp^*}{d\tau} = 0 . \tag{5.19}$$

From Lemma 5.2, we have $dp^*/d\tau > 0$. Therefore, equation (5.19) is satisfied if and only if $\partial SW/\partial p = 0$, that is

$$\begin{aligned} p = c + c_a + \theta\Bigg[v_B Q_B^* + v_L Q_L^* + 2\omega Q^* - \delta \sum_{h\in\{B,L\}} Q_h^* \int_{p_c^W}^{\bar{u}} \frac{\partial \overline{G}_h\big(z; D(Q^*)\big)}{\partial D} dz \Bigg] \\ - \delta (p_c - c_c) \sum_{h\in\{B,L\}} Q_h^* \frac{\partial \overline{G}_h\big(p_c^W; D(Q^*)\big)}{\partial D} \\ - \delta \frac{dp}{dQ} \sum_{h\in\{B,L\}} \Bigg[\frac{dQ_h}{dp} \int_{p_c^W}^{\bar{u}} z g_h(z; d) dz - c_c \overline{G}_h\big(p_c^W; D(Q^*)\big) \Bigg] \end{aligned} \quad . \tag{5.20}$$

Substituting equation (5.10) into equation (5.20), we derive the optimal airport charge, τ^W:

$$\begin{aligned} \tau^W - c_a = \left(\left(1 - \frac{1}{n}\right) \omega Q^* + v_B Q_B^* + v_L Q_L^* - \frac{1}{n} \frac{b_L v_B + b_B v_L}{b_L + b_B} Q^* \right) \theta - \frac{1}{n} \frac{b_L b_B}{b_L + b_B} Q^* \\ - \theta\delta \sum_{h\in\{B,L\}} Q_h^* \int_{c_c}^{\bar{u}} \frac{\partial \overline{G}_h\big(z; D(Q^*)\big)}{\partial D} dz \\ - \delta \frac{dp}{dQ} \sum_{h\in\{B,L\}} \frac{dQ_h}{dp} \int_{c_c}^{\bar{u}} (z - c_c) g_h\big(z; D\big) dz \end{aligned} \quad . \tag{5.21}$$

The first line on the RHS of (5.21) is the sum of the uninternalized *MEC* for airlines, the *MEC* for passengers, a correction for the *MEC* for passengers which is internalized by airlines and a correction for airlines' market power. Note that as observed in Section 5.3, the larger the difference between business and leisure passengers' values of time, the less *MEC* for passengers will be internalized by individual airlines. Consequently, the conventional optimal airport charge which requires the component of congestion toll equal to $(1-1/n)MEC$ no longer applies here even if $\delta = 0$. In particular, if the value of time of business passengers is much higher than that of leisure passengers, the optimal charge should be set higher than *MEC* to protect business passengers; otherwise, the welfare-maximizing airport should charge below *MEC*. Similar to the case of a profit maximizing airport, the second line of (5.21) also contains two terms of interest when $\delta > 0$. The first term is again a downward correction for the congestion toll to internalize the positive externality of congestion on concessions, but this time it stems from the marginal increase in passenger concession surplus rather than the marginal increase in profit. Therefore, the airport can have incentives to reduce the congestion toll so as to increase the passenger volume and their waiting time. The above discussion can be summarized in Proposition 5.2.

Proposition 5.2: *In the case of a welfare maximizing airport, when concession services generate extra surplus, there is a downward correction for the congestion toll which is equal to the marginal passenger concession surplus due to the positive externality of delay. Therefore, it can be useful to decrease the airport charge so as to increase passengers' waiting time and so their consumption of concession goods.*

The last term accounts for the per passenger expected concession surplus, weighted for different passenger types. Unlike previous literature where this term is always negative, this is again no longer clear-cut when more than one type of passenger is considered. This can be seen as follows. Let $\Gamma(D) = \sum_{h \in \{B,L\}} (dQ_h / dp) \int_{c_c}^{\bar{u}} (z - c_c) g_h(z;D) dz$.

Consider the case in which $dQ_B / dp > 0$. Since $dQ/dp < 0$, from Lemma 5.2, we have $dQ_B / dp < -dQ_L / dp$.

It follows that $\Gamma(D) > 0$ when

$$\Lambda(D) = \frac{\int_{c_c}^{\bar{u}} (z - c_c) g_L(z;D) dz}{\int_{c_c}^{\bar{u}} (z - c_c) g_B(z;D) dz} < -\frac{dQ_B / dp}{dQ_L / dp}. \tag{5.22}$$

In other words, when (5.22) is satisfied the last term becomes a mark-up. Specifically, from the definition of $\Lambda(D)$ we have $\Lambda(D)$ decreases with the delay if and only if at the equilibrium

$$\int_{c_c}^{\bar{u}} \frac{\partial \overline{G}_L(z;D)}{\partial D} dz - \int_{c_c}^{\bar{u}} \frac{\partial \overline{G}_B(z;D)}{\partial D} dz < 0. \tag{5.23}$$

The left-hand side (LHS) of (5.23) is the difference between the impacts of delay on the expected concession surplus of one leisure passenger and one business passenger. When (5.23) is satisfied, condition (5.22) is more likely to be fulfilled. Therefore, for high levels of delay it is more likely to have a mark-up. As in the profit maximizing case, findings from Torres *et al.* (2005) support the idea that for these levels of delay it can be useful, for the welfare maximizing airport, to increase the airport charge to protect the business passengers from excessive congestion. This is consistent with Czerny and Zhang (2011) but from another perspective: it is welfare-enhancing to induce more business passengers and let them buy in the commercial area, gaining more extra surplus. Summarizing the above discussion leads to:

Observation 5.2: *In the case of a welfare maximizing airport and two types of travelers, when concession services generate extra surplus, the correction on the optimal airport charge due to the impact of changes in passenger volume on concession surplus is a mark-up, not a mark-down, for some levels of delay. Therefore, the socially efficient airport charge can be higher than what would prevail if passengers are treated as a single type.*

Comparing (5.16) and (5.22), Observations 5.1 and 5.2 differ in the following sense: the profit maximizing airport cares about the difference between the probability of purchase of business and leisure passengers at the monopoly concession price p_c^π; while the welfare maximizing airport cares about the difference between the concession surplus of business and leisure passengers.

5.4.3 Comparison between profit and welfare maximizing airports

In this section, we concentrate on the comparison between the pricing rules of profit and welfare maximizing airports derived above. Specifically, comparing equations (5.15) and (5.21), the first lines on the RHS of both equations are consistent with previous literature; therefore, we focus on the remaining parts – consisting of two terms – which highlight the effects of the positive externality of delay and passenger types on concessions. The first term takes into account the marginal increase in concession profit (passenger concession surplus) due to delay in the case of a profit (welfare) maximizing airport. This term is negative and comes from the positive externality of congestion on concessions. The second term takes into account the impact of different passenger types on the per passenger concession profit (expected concession surplus), in the case of a profit (welfare) maximizing airport. This term may be positive or negative, that is, a mark-up or a mark-down, according to the difference in the values of time between travelers and the level of delay.

Proposition 5.3:

(1) *There exists a* $\bar{\delta} \in (0,1)$ *such that* $\forall \delta \in [0, \bar{\delta})$ *the (downward) correction for the congestion toll due to the positive externality of delay is higher for a profit maximizing airport than a welfare maximizing airport;* $\forall \delta \in (\bar{\delta}, 1]$ *this correction is higher for a welfare maximizing airport than a profit maximizing airport.*

(2) *When the difference in the values of time between passenger types is small and there is mark-down due to concessions, there exists a* $\hat{\delta} \in (0,1)$ *such that* $\forall \delta \in [0, \hat{\delta})$ *the mark-down is higher for a profit maximizing airport;* $\forall \delta \in (\hat{\delta}, 1]$ *it is higher for a welfare maximizing airport. When the difference in the values of time between passenger types is large, the comparison is ambiguous.*

Proof: See Appendix C.1.

The first part of Proposition 5.3 suggests that in some situations a welfare maximizing airport can have more incentives to decrease the congestion toll and induce congestion - so as so to increase the passengers' probability of purchasing concession goods - than a profit maximizing airport. This is more likely to happen in those airports which provide unique and more desirable shopping experiences that are not available elsewhere and thus generate a sufficiently high proportion of extra surplus. The second part of Proposition 5.3 implies that in some situations a welfare maximizing airport can subsidize more than a profit maximizing airport, so as to decrease the aeronautical charge and increase the pool of passengers who are potential consumers of concession goods. This is true when the difference in passengers' values of time is small and the proportion of extra surplus generated by airport concession activities is sufficiently large. However, when the difference in passengers' values of time is large the comparison is no longer clear-cut. Specifically, we may have a charge or a subsidy for both types of airports and three different scenarios can happen depending on two conditions:

$$-\frac{dQ_B/dp}{dQ_L/dp} < \frac{E_L^W - E_L^\pi}{E_B^W - E_B^\pi} \ , \tag{5.24}$$

$$-\frac{dQ_B/dp}{dQ_L/dp} < \frac{E_L^\pi}{E_B^\pi} \ , \tag{5.25}$$

where $E_h^\pi = \left(p_c^\pi - c_c\right)\overline{G}_h\left(p_c^\pi; D\right)$ is the per passenger concession profit and $E_h^W = \int_{c_c}^{\bar{u}} (z - c_c) g_h(z; D) dz$ is the per passenger concession surplus. In the first scenario, when only (5.24) holds, the welfare maximizing airport charges less[58] than the profit maximizing airport. This happens because business passengers generate sufficiently high profit for concessions while leisure passengers generate sufficiently high consumer surplus from concessions. Therefore, the profit maximizing airport has higher incentives to retain business passengers than the welfare maximizing airport. In the second scenario, when only (5.25) holds, the profit maximizing airport charges less and the situation is just reversed. In the last scenario, when both (5.24) and (5.25) hold, there exists a $\tilde{\delta} \in (0,1)$ such that $\forall \delta \in [0, \tilde{\delta})$ the profit maximizing airport charges less; $\forall \delta \in (\tilde{\delta}, 1]$ the welfare maximizing airport charges less. This happens because leisure passengers generate sufficiently high profit in the profit maximizing case and sufficiently high consumer surplus in the welfare maximizing case; that is, when concession activities produce a sufficiently high proportion of extra surplus, the welfare maximizing airport has a stronger incentive to decrease the aeronautical charge and induce more leisure passengers.

[58] Note that it is possible that both airports subsidize, in which case 'charge less' means 'subsidize more'. It is also possible that one airport subsidizes while the other airport charges.

5.5 Concluding remarks

This paper focuses on the impact of concessions on airport congestion pricing. In particular, it adds to literature by taking into account the positive relationship between congestion and the consumption of concession goods, through dwell time, while incorporating the effect of passenger types.

We find that for both profit and welfare maximizing airports there is a downward correction for the congestion toll equivalent to the marginal concession profit and passenger concession surplus, respectively, due to the positive externality of delay. This correction may even turn the congestion toll into a subsidy, which is in contrast with previous literature on airport pricing. Therefore, the airport can have incentives to reduce the aeronautical charge so as to increase passengers' dwell time and their consumption of concession goods. Furthermore, we show that there is a correction on the optimal airport charge equal to the per passenger concession profit and expected concession surplus, weighted for different passenger types, for profit and welfare maximizing airports, respectively. We find that in the case of two types of travelers, for some levels of delay this correction may not be a mark-up rather than the traditional mark-down. Therefore, the optimal airport charge can be higher than what would prevail if passengers are treated as a single type. Finally, the comparison between privately and socially optimal airport charges highlights two results. First, when concession activities generate a sufficiently high proportion of extra surplus, the welfare maximizing airport can have more incentives to decrease the congestion toll and induce congestion, so as to increase the passengers' dwell time and the probability of purchasing concession goods. Second, the profit maximizing airport may impose a lower charge than the welfare maximizing airport, so as to adjust the impact of changes in the pool of potential consumers for concession services, depending on both the difference in the passengers' values of time and the proportion of extra surplus generated by airport concessions.

Non-aeronautical revenues have become the main income source of many airports and studies on the impact of commercial revenues on airport pricing are of increasing concern. Our findings, therefore, can be useful for both academics and practitioners because of their implications for the operation of the industry and the ensuing regulatory requirements. In this sense, further developments of the present work may go in two directions. First of all, in this paper we abstract away the possibility of price discrimination and assume that all passengers are charged a uniform airfare. Hence a natural extension is to check whether our results still hold when price discrimination is allowed. Second, within the scope of policy implications, the impact of different types of regulation, such as single-till or dual-till, should be investigated under our framework. It is of interest to explore whether considering the positive externality of congestion will contribute new insights to the policy debate.

6 Airport Pricing under the Separation of Terminal and Runway Congestion

6.1 Introduction

Airport congestion has become a growing phenomenon in the aviation industry. As a potential solution to relieve airport congestion, the imposition of congestion toll has been widely discussed and proposed by extensive literature. However, congestion toll is embedded in the airport charge, which is determined by an interaction between many factors. In order to achieve an efficient toll level, it is vital to make sure that every relevant aspect is under consideration. Airline market structure needs to be taken into account, since airlines with market power may well 'internalize' the congestion cost they impose to their own flights (Brueckner, 2002; Brueckner and Van Dender, 2008; Pels and Verhoef, 2004; Zhang and Zhang, 2006). Airport concession should also be in the picture, given that the number of passengers plays a different role in contributing to the airport's congestion level and concession revenues (Oum et al., 2004; Yang and Zhang, 2011; Zhang and Zhang, 2003, 2010). The types of passengers may also matter, because different types of passengers have different perceptions about congestion toll (Czerny and Zhang, 2011, 2012). In chapter 5, we investigated the interaction between these factors while incorporating the connection between congestion delay and concession consumption and derived corresponding optimal airport charges. Note that chapter 5 is based on the assumption that an increase in congestion delay will increase dwell time and hence induce higher probability of purchasing concession goods. This assumption does not reflect the fact that airport congestion may occur in the terminals or on the runways.

Interestingly, none of the previous studies has differentiated the congestion incurred in the terminals and that incurred in the runways, despite of the fact that these two types of congestion clearly have different implications to the airlines and the airport in the above-mentioned contexts. In particular, terminal congestion seems to be less of a concern to the airlines' operations, but it is likely to affect passenger behavior and airport concession activity to a large extent. On the other hand, runway congestion is more of an issue to the airlines and has less to do with airport concession. In other words, airport concession and passenger types are more closely related with terminal congestion than runway congestion. For example, there are cases that passengers will have to stay on the aircraft waiting for their turns to take-off when the runway is congested and hence is not able to purchase any concession goods during the waiting. Therefore, separating these two types of airport congestion may help clarify and deepen our understanding of the interactions between different factors in designing an optimal airport charge. Furthermore, these two types of airport congestion show different characteristics, with terminal congestion being totally ‘atomistic’ while runway congestion showing a certain degree of ‘internalization’. All in all, it would be beneficial to study the impacts of separating terminal congestion from runway congestion, both on our understanding of airport congestion and on the design of an optimal airport toll.

In this chapter, we modify the model in chapter 5 and propose a framework that treats terminal congestion and runway congestion separately. To capture the difference between these two types of airport congestion, we adopt a deterministic bottleneck model for the terminal and a simpler and more traditional congestion model for the runways. Bottleneck model is a more accurate structure based on queuing theory, but it only fits cases when players are all ‘atomistic’ such as highway traffic flow. Airport terminals face individual passengers who are by definition atomistic and will not take into account other passengers when making decisions, so it is a perfect context for the usage of bottleneck model. Runways, on the other hand, face airplanes operated by a few airlines that may have market power and hence may potentially internalize the congestion they impose on their own flights. Therefore, bottleneck model is not a good fit for runway congestion. Another difference with chapter 5 is that now we no longer assume airlines do not price discriminate passengers. Rather, in this chapter, airlines set airfares for business and leisure passengers independently. Our objective is to investigate the optimal airport charge given the separated treatment of runway and terminal congestions and compare our results with the traditional results found in the literature.

We find that oppose to Czerny and Zhang (2011, 2012), when terminal congestion is taken into account, welfare optimal uniform airfare *does not* yield the first-best outcome. First-best can be achieved through discriminative fares. The first-best fare charged on the business passengers is higher than that on the leisure passengers if and only if the relative schedule delay cost of business passengers is higher than leisure passengers. When the airport discriminates business and leisure passengers, increasing airport charge of one type of passengers will reduce the equilibrium quantity of this type but raise that of the other type. Although the total traffic volume decreases in leisure passenger airport charge, it may increase in business passenger airport charge. When both types of passengers are levied a uniform airport charge, an increase in airport charge will reduce the number of leisure passengers and the total number of passengers, but may reduce or increase the number of business passengers. Furthermore, we identify, under optimal uniform airport charge, various conditions under which terminal charges are above or under the externality a certain passenger imposes on the others. We also compare the airport pricing rule with the one derived in chapter 5 and find that under certain conditions to increase leisure passengers' dwell time and hence chance of purchasing concession goods, the airport will raise rather than reduce the airport charge.

The structure of the paper is as follows. Section 6.2 sets up the model. Sections 6.3 models passengers' equilibrium arrival pattern at the terminal. Section 6.4 derives the first-best outcomes. Sections 6.5 discusses, airlines' and airport's equilibrium behaviors. Section 6.6 contains the concluding remarks.

6.2 The model

Similar to chapter 5, we consider one congestible airport served by n identical airlines. Airlines compete in quantities to maximize their own profits. We only consider passengers departing from the airport.[59] There are two types of passengers: business and leisure. Business passengers have higher value of time than leisure passengers, i.e. $v_B > v_L$. Unlike chapter 5, now we assume that each airline makes separate decisions on the number of business passengers and the number of leisure passengers, leading to laissez-fair situation.

[59] Arriving passengers' concession activity is less likely to be correlated with terminal congestion. Besides, departure and arrival do not share same terminal procedures but may share runway.

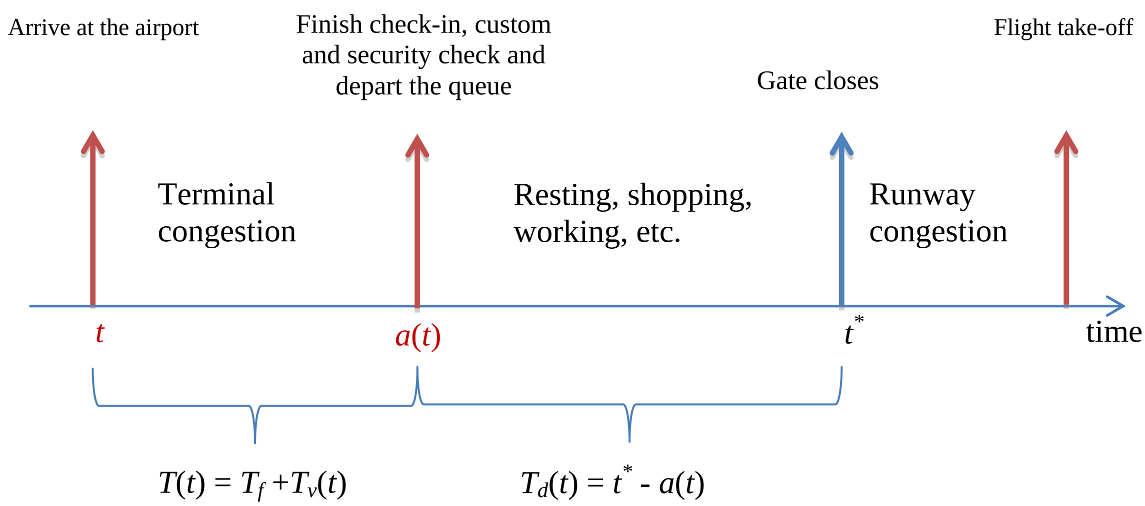
Further, we assume that airlines are able to price discriminate the two types of passengers with respect to the laissez-fair results. That is, airlines have effective mechanisms to prevent one type of passengers from mimicking the other type. We denote q_h^i as the number of type *h* passengers flying with airline *i* and $Q_h = \sum_{i=1}^{n} q_h^i$. The (inverse) demand function for passenger type *h* is $\rho_h(Q_h)$. We assume that $\rho_h(Q_h)$ is two times differentiable with $\rho_h'(Q_h) < 0$ and $\rho_h(Q_h) \cdot Q_h$ is strictly concave throughout the non-negative domain. In particular, we require that for any (Q_B, Q_L) pair, $\rho_h'' Q_h + 2\rho_h' < 0$. The latter assumption implies that the first-order condition leads to global maximum of the revenues summed across all airlines when there are no user costs.

Let t^* be the scheduled flight departure time. We assume that all the departing flights in concern are scheduled at this time and gates will close at this time as well. Although flights depart at different times over the day, this assumption may be analogous to flight banking behavior in hub airports which have been indicated in the literature (e.g. Daniel, 2001; Mayer and Sinai, 2003). In order to facilitate connecting passengers, hub carriers have incentives to schedule flights closer to each other, forming flight clusters or banks across the day. Within each bank, a large number of flights depart or arrive at times as close together as possible, leading to departure or arrival peaks. Between peaks, there are few landing or taking-off activities for hub carriers. Mayer and Sinai (2003) found that at Dallas-Fort Worth airport, hub carriers' flight arrival patterns are much smoother than departure patterns, suggesting departure banking is more pronounced and creates more severe delays than arrival banking. Therefore, our model will focus on departure passengers only. To simplify the analysis, we assume that the time between two peaks are long enough so that flights scheduled in a certain bank will not affect departures in other banks. Thus, we can consider t^* as the preferred departure time of one representative bank. Figure 6.1 illustrates the activity timeline for a particular departing passenger.[60] At the terminal side, the pre-departure procedure, including

[60] Here departing passengers only refer to originating passsengers, because in most of the cases connecting passengers do not follow the same terminal procedure as originating passengers. Sometimes connecting passengers are less affected by terminal congestion. For example, most connecting passengers will get their

check-in, security screening as well as passport control (if any), creates queues and will be considered collectively as a bottleneck. [61]

Figure 6.1 Activity timeline for a particular departure passenger who completes pre-departure procedures before scheduled gate closing time.



The passenger arrives at the airport at time t and has to wait in line for $T_v(t)$ units of time before being processed. Check-in and screening procedures take T_f units of time per passenger and have a capacity of s passengers per unit of time. We define $T(t) = T_v(t) + T_f$. The passenger completes the pre-departure procedure at time $a(t)$, or equivalently $t + T(t)$, and then proceeds to the departure lounge to wait for boarding at t^*. We assume that being late and missing the flight is never an acceptable option for passengers, i.e. $a(t) \leq t^*$. If the

boarding passes when checking-in for the first flight at their very origins and in the U.S. they do not need to go through security screening again at the transitting airport as long as they stay in the airside of the terminal. When passengers transit between an international leg and a domestic leg, however, substantial delays due to the crowded terminal can occur. As flights are scheduled by airlines and those connecting passengers have less control on their airport arrival time, these passengers are not particularly modeled in this paper.

[61] The present study assumes all passengers go through the same line. Some airports may provide dedicated express lines for frequent flyers or first-class and business passengers. To incoperate this, the present model can be modified by having two separate bottlenecks.

passenger finishes pre-departure procedures earlier than t^*, she will incur early schedule delay cost, $v_h \beta_h (t^* - a(t))$, $h \in \{B, L\}$, where $v_h \beta_h$ is the unit early schedule delay cost of type h passengers and β_h is the relative cost of early schedule delay to time delay. Here, we assume the early schedule cost is lower than pure waiting time cost, v_h. We assume that the waiting time in terminal queue and in runway queue incurs the same type of time cost and $\beta_h \in (0,1)$, because passengers feel more comfortable and can perform more activities when waiting in the lounge than lining up in a queue. Note that the dwell time of passenger arriving at the airport at time t is defined as: $T_d(t) = t^* - a(t)$. Following the common simplification in the literature, we assume that the fixed 'free flow' processing time T_f is constant and normalize it to zero. Thus, the terminal cost of passenger arriving at the airport at time t is:

$$c^h(t) = v_h T(t) + \beta_h v_h (t^* - a(t)) = v_h T(t) + \beta_h v_h (t^* - t - T(t)) \qquad (6.1)$$

Once being boarded on the aircrafts, as all the flights are scheduled and ready for take-off at t^*, they will form a queue and incur runway delays. Similar to chapter 5, we assume the seat capacity of each flight is the same and all flights are fully loaded. Hence, we denote runway delay as $D(Q)$ where $Q = Q_B + Q_L$, i.e. the total number of passengers served by all carriers across passenger types. Following the literature, $D' > 0$ and $D'' \geq 0$. In particular, if we assume all flights will form a random queue and the runway capacity is K passengers per unit of time, the first take-off will be at t^* and the last one will be at t^*+Q/K. Consequently, the expected runway delay is:

$$D(Q,K) = \int_0^{\frac{Q}{K}} z \frac{K}{Q} dz = \frac{Q}{2K} = \theta Q,$$

where $\theta = 1/2K$. The above is an example of deriving the congestion delay on the runway. In the rest of the chapter, we impose no functional form on runway congestion and consider D as a general function of Q while surpressing the notation of K. A type h passenger who arrives at the airport at time t incurs a generalized cost which equals to the sum of airfare, runway congestion cost and terminal cost:

$$f_h(t) = p_h + v_h D(Q) + c^h(t), \qquad (6.2)$$

where p_h is the airfare paid by the type *h* passenger.

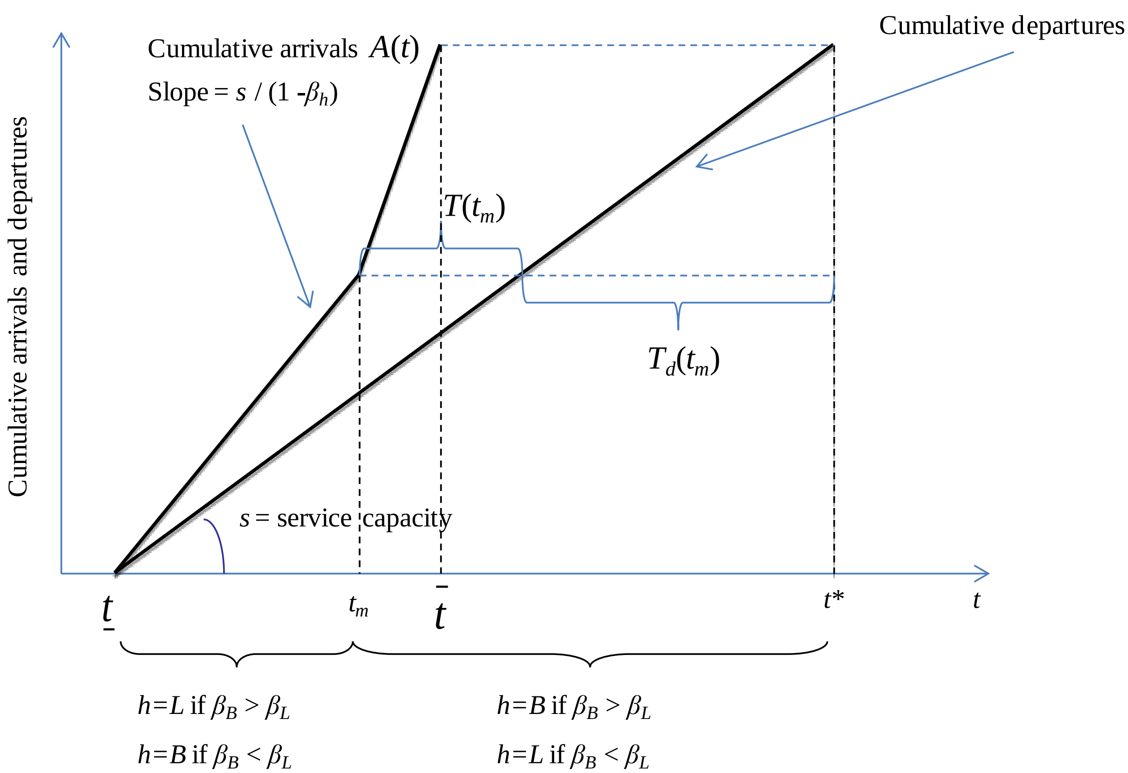
During the dwell time passengers will consume concession goods. Following chapter 5, we assume that the utility of purchasing concession goods, *u*, for a type *h* passenger who arrives at the airport at time *t*, follows cumulative distribution, $G_h(u;T_d(t))$ on the domain $[0,\bar{u}]$, which satisfies the first order stochastic dominance property (FOSD) with respect to the amount of dwell time. As the passenger will purchase only when the utility of purchasing concession goods exceeds the price, p_c, the probability of purchasing equals to $\bar{G}_h(p_c;T_d(t)) = 1 - G_h(p_c;T_d(t))$. The FOSD implies that $\bar{G}_h(p_c;T_d) \geq \bar{G}_h(p_c;\tilde{T}_d)\ \forall T_d > \tilde{T}_d$.

The sequence of the game is as follows. In the first stage, the airport sets the airport charge paid by airlines to maximize social welfare. In the second stage, airlines simultaneously determine output levels to maximize their respective profits. In the third stage, passengers determine whether to purchase the air ticket or not and which airline to travel with. In the last stage, passengers who decide to travel choose the time to arrive at the airport to minimize their respectively individual terminal cost. For simplicity, we normalize all the other costs incurred by airlines and the airport to be zero.

6.3 Passenger equilibrium

To obtain the sub-game perfect Nash equilibrium, we use backward induction and start from the last stage where passengers determine the airport arrival time given that they will take the flight at time t^*. According to the literature (e.g. Arnott, et al, 1994; Arnott and Kraus, 2003; and van den Berg and Verhoef, 2011), at the equilibrium, the cumulative number of passengers arriving at the airport and departing the bottleneck (i.e. completing the pre-departure procedures) follow the patterns shown in Figure 6.2. As indicated in conventional bottleneck models, we assume that passengers are perfectly informed of the congestion in the airport. This assumption is a little bit strong but still sensible for two reasons: first, frequent travelers usually have a good knowledge about the airport’s congestion level; second, although occassional travelers have much less experience with and knowledge about the airport, it is easy to learn about on-time performance of each airport or even a certain flight from the internet as long as the passengers care about such information.

Figure 6.2 Cumulative arrivals at the airport and departures from pre-departure processing sites at the terminal



All passengers will arrive at the airport during the time interval $[\underline{t}, \bar{t}]$. Passengers arrive at $\underline{t}$ incur zero queuing cost but only early schedule delay cost, while passengers who arrive at $\bar{t}$ will incur only queuing cost and leave the bottleneck exactly at time t^*, incurring zero early schedule delay cost. The bottleneck will always operate at its capacity until the last passenger is processed. Therefore, the rate of leaving the bottleneck is capacity *s* which defines the slope of the cumulative departures curve. As passengers minimize their individual terminal costs by choosing the time of arrival, we take the first-order condition of equation (6.1) with respective to *t* and obtain:

$$T'(t) = \frac{\beta_h}{1-\beta_h}.$$

Note that the relationship between $A(t)$, the total number of passengers arrive by t, and $T(t)$ is: $T(t) = \left(\mathrm{A(t)} - \mathrm{s(t} - \underline{\mathrm{t}})\right)/s$. Therefore, it is easy to find the slopes of the cumulative arrivals curve equal to

$$A'(t) = \frac{s}{1-\beta_h},$$

for $h = B$ and L respectively and indicate the arrival rates for type h passengers. If the relative schedule delay cost of business passengers is higher, i.e. $\beta_B > \beta_L$, business passengers are more sensitive to early schedule delay cost and willing to accept longer waiting time in return for shorter dwell time (e.g. Arnott, et al, 1994; Arnott and Kraus, 2003; and van den Berg and Verhoef, 2011). As a result, leisure passengers arrive during the interval $[\underline{t}, t_m]$ followed by business passengers who arrive during $[t_m, \bar{t}]$ and the arrival rate of the former is lower than the later, suggesting that the left segment of the cumulative departure curve is flatter than the right segment. Likewise, if $\beta_B < \beta_L$, business passengers will arrive earlier at a lower rate.

At equilibrium, passengers of the same type must face with the same terminal cost; otherwise, passengers with higher cost will adjust their arrival times to those associated with lower costs. Based on this equilibrium condition, terminal costs can be derived as follows:

$$c^B(t) = c^B(\bar{t}) = v_B T(\bar{t}), \quad c^L(t) = c^L(\underline{t}) = v_L \beta_L (t^* - \underline{t}) \quad \text{if} \quad \beta_B > \beta_L;$$

$$c^B(t) = c^B(\underline{t}) = v_B \beta_B (t^* - \underline{t}), \quad c^L(t) = c^L(\bar{t}) = v_L T(\bar{t}) \quad \text{if} \quad \beta_B < \beta_L. \qquad (6.3)$$

Because the queue starts at $\underline{t}$ and ends at t^*, we have $t^* - \underline{t} = (Q_B + Q_L)/s$. Because all passengers arrive at the airport by $\bar{t}$, we have

$$\bar{t} - \underline{t} = \frac{Q_B(1-\beta_B)}{s} + \frac{Q_L(1-\beta_L)}{s}.$$

The time spent by the last passenger waiting before being processed equals to

$$T(\bar{t}) = \frac{1}{s}\left[Q_B + Q_L - (\bar{t} - \underline{t})s\right] = \frac{Q_B\beta_B + Q_L\beta_L}{s}.$$

Then we can rewrite (6.3) into the following:

$$c^B(t) = v_B \frac{\beta_B Q_B + \beta_L Q_L}{s}, \quad c^L(t) = v_L \beta_L \frac{Q_B + Q_L}{s} \qquad \text{if} \quad \beta_B > \beta_L;$$

$$c^B(t) = v_B \beta_B \frac{Q_B + Q_L}{s}, \quad c^L(t) = v_L \frac{\beta_B Q_B + \beta_L Q_L}{s} \qquad \text{if} \quad \beta_B < \beta_L. \qquad (6.4)$$

In a word, the equilibrium terminal costs are functions of Q_B and Q_L. From now on, we replace $c^B(t)$ and $c^L(t)$ with $c^B(Q_B, Q_L)$ and $c^L(Q_B, Q_L)$ for terminal costs. We denote $\partial c^h / \partial Q_h$ as c_h^h and $\partial c^h / \partial Q_k$ as c_k^h $\forall h, k \in \{B, L\}$ and $h \neq k$. One interesting observation is that the structure of per passenger terminal costs depends on the relative sizes of β_B and β_L. The passenger type with higher relative schedule delay cost has higher own effect on the terminal cost than the cross effect. In particular, when $\beta_B > \beta_L$, the increase in business passenger terminal cost due to one additional business passenger is higher than one additional leisure passenger: $c_B^B > c_L^B$; however, regarding the leisure passenger terminal cost, the impact of increasing business passengers is the same as increasing leisure passengers: $c_L^L = c_B^L$. When $\beta_B < \beta_L$, we have $c_L^L > c_B^L$, while $c_B^B = c_L^B$. This property does not exist in the runway delay cost function or conventional airport congestion cost function which always assumes that it is the total traffic volume rather than the traffic volumes of individual passenger types that matters.

When passengers make air ticket purchasing decisions in the third stage, they take these terminal costs into consideration and equation (6.2) becomes

$$f_h(t) = f_h(Q_B, Q_L) = p_h + v_h D(Q) + c^h(Q_B, Q_L) \ . \tag{6.5}$$

The third-stage demand equilibrium requires $\rho_h(Q_h) = f_h(Q_B, Q_L)$, which leads to

$$p_h = \rho_h(Q_h) - v_h D(Q) - c^h(Q_B, Q_L) \quad \forall h \in \{B,L\}. \tag{6.6}$$

Thus, the inverse demand function with respect to airfare is a function of both Q_B and Q_L. Applying the Cramer's rule with respect to p_h, we obtain Lemma 6.1.

Lemma 6.1: *Under laissez-fair situation,* $\partial Q_h / \partial p_h < 0$ *and* $\partial Q_h / \partial p_k > 0 \quad \forall h \neq k$. *Under uniform pricing, where* $p_B = p_L = p$, $\partial Q_L / \partial p < 0$ *and* $\partial Q / \partial p < 0$, *but the sign of* $\partial Q_B / \partial p$ *is ambiguous.*

Proof: See Appendix D.1.

Under laissez-fair situation, an increase in type *h* airfare will suppress the volume of type *h* passengers while the reduced congestion levels will induce the demand of type *k* passengers, as the two types of passengers compete for terminal and runway resources. Regarding the case of uniform pricing, Czerny and Zhang (2011) obtained similar results when considering runway congestion alone.[62] They found that $\partial Q_B / \partial p$ is more likely to be positive if the difference between v_B and v_L is larger. In our model, both terminal congestion and runway congestion affect the sign of $\partial Q_B / \partial p$, but the terminal cost plays a role slightly different from the runway congestion cost. As indicated in Appendix D.1, suppose that there is no runway congestion and $\beta_B < \beta_L$, the sign of $\partial Q_B / \partial p$ does not depend on the difference between v_B and v_L but the difference between early schedule delay costs, $v_B \beta_B$ - $v_L \beta_L$, which could be negative though $v_B > v_L$.

[62] We also obtained similar results in chapter 5 based on linear demand function. Lemma 6.1 is proved with general demand functions.

6.4 First-best outcomes without time-varying terminal tolls

According to traditional bottleneck literature, social optimum is obtained when there is no queuing cost but only schedule delay costs. This is attainable if time-varying tolls are levied. However, in the case of airport, this is unlikely to happen. Thus, in this section, we only derive first-best outcomes in the sense that passenger queuing at the terminal persists and the terminal cost structure derived in Section 6.3 does not change.

The social welfare consists of two parts, aeronautic surplus and concession surplus. The former can be expressed in the following way:

$$S_a = \sum_{h=B,L} \int_0^{Q_h} \rho_h(x)dx - v_h D(Q)Q_h - c_h(Q_B, Q_L)Q_h \ .$$

From Section 6.3, we know that for each unit of time, s passengers will start their dwell time. Thus, the concession demand for each level of dwell time can be written as

$$x_h(p_c, T_d) = s \cdot \overline{G}_h(p_c; T_d)$$

Then, for any $\beta_h > \beta_k$, concession surplus can be written as

$$\begin{aligned} S_c = & \int_{T_d(t_m)}^{T_d(\underline{t})} \int_{p_c}^{\bar{u}} s\overline{G}_h(x;z)dxdz + \int_{T_d(\bar{t})}^{T_d(t_m)} \int_{p_c}^{\bar{u}} s\overline{G}_k(x;z)dxdz \\ & + (p_c - c_c)\left(\int_{T_d(t_m)}^{T_d(\underline{t})} s\overline{G}_h(p_c;z)dz + \int_{T_d(\bar{t})}^{T_d(t_m)} s\overline{G}_k(p_c;z)dz \right)' \end{aligned} \tag{6.7}$$

where c_c is the cost of one unit of concession goods and

$$T_d(\underline{t}) = t^* - \underline{t} = \frac{Q}{s}, \ T_d(t_m) = t^* - (t^* - \frac{Q_h}{s}) = \frac{Q_h}{s}, \ T_d(\bar{t}) = t^* - t^* = 0 \, .$$

Note that concession price only affects concession surplus and it is easy to show that

$$\frac{dS_c}{dp_c} = -s(p_c - c)\left(\int_{T_d(t_m)}^{T_d(\underline{t})} g_h(p_c;z)dz + \int_{T_d(\bar{t})}^{T_d(t_m)} g_k(p_c;z)dz \right). \tag{6.8}$$

The second term of (6.8) is always positive and hence to maximize concession surplus, the optimal concession price should always be set at the cost, a result consistent to Yang and Zhang (2011) and chapter 5. Therefore, to simplify the analysis, we assume $p_c = c_c$ for the rest of the paper. Consequently, (6.7) can be rewritten as

$$S_c = \int_{\frac{Q_h}{s}}^{\frac{Q}{s}} \int_{c_c}^{\bar{u}} s\overline{G}_h(x;z)dxdz + \int_0^{\frac{Q_h}{s}} \int_{c_c}^{\bar{u}} s\overline{G}_k(x;z)dxdz .$$

It remains a debate on whether concession surplus should be considered as part of the social welfare, as consumption on food and beverage would occur elsewhere even if passengers do not consume at the airport. Following chapter 5, we weight concession surplus by $\delta \in [0,1]$. Then, the social welfare equals

$$W = S_a + \delta \cdot S_c . \tag{6.9}$$

The first-best outcomes can be derived by taking first-order conditions of (6.9) with respect to Q_B and Q_L. That is,

$$\frac{\partial W}{\partial Q_B} = p_B^* - \Gamma^* - \phi_B^* + \delta \cdot \frac{\partial S_c^*}{\partial Q_B} = 0 , \tag{6.10}$$

$$\frac{\partial W}{\partial Q_L} = p_L^* - \Gamma^* - \phi_L^* + \delta \cdot \frac{\partial S_c^*}{\partial Q_L} = 0 , \tag{6.11}$$

where star stands for first-best outcomes and $\Gamma = (v_B Q_B + v_L Q_L) D'(Q)$, the marginal external congestion cost at runway; $\phi_h = Q_h c_h^h + Q_k c_h^k$, the marginal external terminal cost due to one additional type *h* passenger at terminal; and

$$\frac{\partial S_c}{\partial Q_h} = \begin{cases} \int_c^{\bar{u}} \overline{G}_k\left(x;\frac{Q}{s}\right) - \overline{G}_k\left(x;\frac{Q_h}{s}\right) + \overline{G}_h\left(x;\frac{Q_h}{s}\right) dx & \text{if } \beta_h > \beta_k \\ \int_c^{\bar{u}} \overline{G}_h\left(x;\frac{Q}{s}\right) dx & \text{if } \beta_h < \beta_k \end{cases} .$$

Let us first consider the situation where concession surplus is not a concern ($\delta = 0$). It is easy to show that the difference of the first-best airfares between business and leisure passengers is

$$p_B^* - p_L^* = \phi_B^* - \phi_L^* = (c_B^B - c_L^B)Q_B^* - (c_L^L - c_B^L)Q_L^*, \tag{6.12}$$

which is positive (negative) if the relative schedule delay cost of business passenger (β_{B}) is higher (lower) than that of leisure passengers (β_{L}) . Moreover, equation (6.12) will not be equal to zero unless β_{B} and β_{L} are equal, because when the relative schedule delay costs are not equal, one of the two terms on the right-hand side (RHS) of equation (6.12) will be zero and the other term will be either positive or negative. This is different from the finding of Czerny and Zhang (2011) which suggests that the social optimal can be achieved only by the welfare optimal uniform airfare. This difference is fundamentally driven by the inclusion of terminal costs into the picture. When runway congestion is the only concern, the impact on runway congestion due to one additional business passenger is the same as one additional leisure passenger. Thus, at social optimum, both types of passengers should be charged at the same price. However, this is no longer the case when terminal cost enters the picture. For example, when $\beta_B > \beta_L$, adding one more business passenger has more adverse impact on business passengers' terminal cost than adding one more leisure passenger. Consequently, business passengers will be charged higher than leisure passengers. The above analysis leads to the following proposition.

Proposition 6.1*: When taking into account terminal congestion but not concession activities, the welfare optimal uniform airfare does not yield the first-best outcome. First-best can be achieved through discriminative fares. The first-best fare charged on the business passengers is higher than that on the leisure passengers if and only if the relative schedule delay cost of business passengers is higher than leisure passengers.*

The inclusion of concession surplus adds another layer of complexity to the analysis. The FOSD property of $G_h(\cdot;T_d)$ suggests that the marginal change of concession surplus should always be positive regardless the difference between β_{B} and β_{L}, because $Q/s > Q_h/s$ and hence $\overline{G}_k\left(x;Q/s\right) \geq \overline{G}_k\left(x;Q_h/s\right)$. Thus, taking concession surplus into consideration will lead to a markdown on first-best airfares. However, the impact on the difference between p_B^* and p_L^* is less straightforward. Equation (6.12) now becomes

$$p_B^* - p_L^* = \begin{cases} (c_B^B - c_L^B)Q_B^* + \delta\int_c^{\bar{u}} \overline{G}_L\left(x;\dfrac{Q_B^*}{s}\right) - \overline{G}_B\left(x;\dfrac{Q_B^*}{s}\right)dx & if \quad \beta_B > \beta_L \\ -(c_L^L - c_B^L)Q_L^* + \delta\int_c^{\bar{u}} \overline{G}_L\left(x;\dfrac{Q_L^*}{s}\right) - \overline{G}_B\left(x;\dfrac{Q_L^*}{s}\right)dx & if \quad \beta_B < \beta_L \end{cases}. \tag{6.13}$$

When δ is positive, the difference between p_B^* and p_L^* depends not only on the relative schedule delay costs but the distributions of concession goods utility, $G_h(\cdot;T_d)$. As mentioned in chapter 5, Torres *et al*. (2005) empirically compare with the average concession goods expenditure of leisure passengers. They find that there exist $\hat{T}_{d1} < \hat{T}_{d2}$ such that business passengers on average spend more than leisure passengers if the dwell time is lower than $\hat{T}_{d1}$ or higher than $\hat{T}_{d2}$; leisure passengers spend more on average if the dwell time is between $\hat{T}_{d1}$ and $\hat{T}_{d2}$ minutes. Therefore, to carry on further analysis, we make some assumption on the functional form of $G_h(\cdot;T_d)$: for any level of *u*, $\overline{G}_B(u;T_d) \geq \overline{G}_L(u;T_d)$ if $T_d \leq \hat{T}_{d1}$ or $T_d \geq \hat{T}_{d2}$; $\overline{G}_B(u;T_d) < \overline{G}_L(u;T_d)$, otherwise. This assumption, together with equation (6.13), leads to Proposition 6.2.

Proposition 6.2: *Under the case of $\delta > 0$, (i) when $\beta_B > \beta_L$, $p_B^* > p_L^*$ if $\hat{T}_{d1} \leq Q_B^*/s \leq \hat{T}_{d2}$ while $p_B^* < p_L^*$ may occur otherwise; and (ii) when $\beta_B < \beta_L$, $p_B^* < p_L^*$ if $Q_L^*/s \leq \hat{T}_{d1}$ or $Q_L^*/s \geq \hat{T}_{d2}$ while $p_B^* > p_L^*$ may occur otherwise.*

Note that Q_h^*/s is the dwell time for passengers arriving the airport at t_m at first-best, given that the relative schedule delay cost of type *h* passengers is higher. Thus, the difference between first-best airfares does not depend on the dwell time of every passenger but only the passengers arriving at the point of time which sets the two types of passengers apart. A comparison with Proposition 6.1 shows that the inclusion of concession surplus may change the passenger type which should be charged higher at first-best to protect the other type. This is because the inclusion of concession surplus may create a trade-off between business and leisure passengers under certain circumstances. For example, when $\beta_B > \beta_L$, serving one more leisure passenger incurs lower external marginal terminal cost than serving one more business passenger; however, the increase in concession surplus due to one additional business passenger may be higher if $Q_B^* / s < \hat{T}_{d1}$ or $Q_B^* / s > \hat{T}_{d2}$. Given such trade-off, at the first-best, business passengers may or may not be charged higher than leisure passengers once concession surplus is in concern.

6.5 Equilibrium of the airline-airport game stages

6.5.1 Airline equilibrium behavior

In the second stage, airlines simultaneously choose output levels for business and leisure passengers while treating airport charge as given. Each airline's objective is to maximize profit

$$\begin{aligned} \pi^i(q_B^1,\ldots,q_B^n;q_L^1,\ldots,q_L^n) &= \sum_{h=B,L}\left(p_h - \tau_h\right)q_h^i \\ &= \sum_{h=B,L}\left(\rho_h(Q_h) - v_h D(Q) - c^h(Q_B,Q_L) - \tau\right)\cdot q_h^i \end{aligned}, \tag{6.14}$$

where τ_h is the airport charge per type h passenger. The first-order conditions of (6.14) with respect to q_h^i for any $i = 1..n$ and $h,k \in \{B,L\}$ with $k \neq h$ are

$$\frac{\partial \pi^i}{\partial q_h^i} = (\rho_h^{'} - v_h D^{'} - c_h^h)q_h^i + p_h - (v_k D^{'} + c_h^k)q_k^i - \tau_h = 0\,. \tag{6.15}$$

Since we assume the competing airlines are identical, by imposing symmetry and using superscript N to denote Nash equilibrium at the airline stage, (6.15) can be rewritten as the following

$$p_h^N + \frac{1}{n}(\rho_h^{'} Q_h^N - \Gamma^N - \phi_h^N) - \tau_h = 0 \quad \forall h \neq k. \tag{6.16}$$

Thus, at equilibrium, unlike chapter 5 which assumes no price discrimination on airfares across passengers, now as the two passengers types are charged independently, each airline will internalize exactly its own share of marginal external runway congestion cost. In addition, each airline will internalize its own share of marginal external terminal costs imposed by type *h* passengers and exercising its market power. The equilibrium outcomes can be written as a function of (τ_B, τ_L).

To guarantee that the first-order conditions give the local maximum, we assume that at equilibrium the Hessian matrix of π^i is negative definite. That is,

$$\pi_{hihi}^i = \rho_h^{''} \frac{Q_h}{n} + 2(\rho_h^{'} - v_h D^{'} - c_h^h) - (v_h Q_h + v_k Q_k) \frac{D^{''}}{n} < 0, \text{ and}$$

$$\pi_{BiBi}^i \pi_{LiLi}^i - \pi_{BiLi}^i \pi_{LiBi}^i > 0, \tag{6.17}$$

where $\pi_{hiki}^i = \pi_{kihi}^i = -(v_h D^{'} + c_k^h + v_k D^{'} + c_h^k) - (v_h Q_h + v_k Q_k) \frac{D^{''}}{n} < 0$.

The existence of a unique equilibrium requires that the stability condition is satisfied as well. To satisfy the stability condition, we further assume the maximum absolute eigenvalue of the $\partial q_i^R / \partial q_j$ matrix, i.e. the matrix representing the impacts of changes in airline *j*'s decision on airline *i's* best response output levels, is less than 1/ (*n* - 1). Following the approach in Zhang and Zhang (1996), this assumption leads to Lemma 6.2.

Lemma 6.2*: Given that at the Nash equilibrium the Hessian matrix of π^i is negative definite, then (i) the sufficient condition of local stability is that the maximum absolute eigenvalue of $\partial q_i^R / \partial q_j$ is less than* 1/ (*n* - 1)*; and (ii) this sufficient condition leads to*

$$\Delta = \begin{vmatrix} \pi^{i}_{BiBi} + (n-1)\pi^{i}_{BiBj} & \pi^{i}_{BiLi} + (n-1)\pi^{i}_{BiLj} \\ \pi^{i}_{LiBi} + (n-1)\pi^{i}_{LiBj} & \pi^{i}_{LiLi} + (n-1)\pi^{i}_{LiLj} \end{vmatrix} > 0 \,. \tag{6.18}$$

Proof: See Appendix D.2.

The impact of airport charges on the equilibrium traffic volumes depends on whether the airport charges based on passenger types or not. When the airport charge is discriminatory, differentiating both sides of (6.15) with respect to τ_h, imposing symmetry and using Cramer's rule, we have

$$\frac{\partial q^{iN}_{B}}{\partial \tau_B} = \frac{\pi^{i}_{LiLi} + (n-1)\pi^{i}_{LiLj}}{\Delta}, \; \frac{\partial q^{iN}_{L}}{\partial \tau_B} = -\frac{\pi^{i}_{LiBi} + (n-1)\pi^{i}_{LiBj}}{\Delta},$$

$$\frac{\partial q^{iN}_{B}}{\partial \tau_L} = -\frac{\pi^{i}_{BiLi} + (n-1)\pi^{i}_{BiLj}}{\Delta}, \; \frac{\partial q^{iN}_{L}}{\partial \tau_L} = \frac{\pi^{i}_{BiBi} + (n-1)\pi^{i}_{BiBj}}{\Delta}.$$

When there is a monopoly airline (n = 1), following (6.18), it is straightforward to show that $\partial Q^{N}_{h} / \partial \tau_h = \pi^{i}_{hihi} / \Delta < 0$ and $\partial Q^{N}_{h} / \partial \tau_k = -\pi^{i}_{kihi} / \Delta > 0$. When there is competition in the air carrier market, the assumption $\rho^{''}_{h} Q_h + 2\rho^{'}_{h} < 0$ (indicated in Section 6.2) implies that $\rho^{''}_{h} Q_h / n + \rho^{'}_{h} \leq 0$, for any $n \geq 2$. Therefore, we have

$$\pi^{i}_{hihj} = \rho^{''}_{h} \frac{Q_h}{n} + (\rho^{'}_{h} - v_h D^{'} - c^{h}_{h}) - (v_h Q_h + v_k Q_k) \frac{D^{''}}{n} < 0 \quad \forall h \neq k$$

$$\pi^{i}_{BiLj} = -(v_B D^{'} + c^{B}_{L}) - (v_B Q_B + v_L Q_L) \frac{D^{''}}{n} < 0$$

$$\pi^{i}_{LiBj} = -(v_L D^{'} + c^{L}_{B}) - (v_B Q_B + v_L Q_L) \frac{D^{''}}{n} < 0$$

Therefore, given that the part (ii) of Lemma 6.2 holds, we have $\partial Q^{N}_{h} / \partial \tau_h = n \partial q^{iN}_{h} / \partial \tau_h < 0$ and $\partial Q^{N}_{h} / \partial \tau_k = n \partial q^{iN}_{h} / \partial \tau_k > 0$. Further calculation leads to Proposition 6.3.

Proposition 6.3: *When the airport discriminates business and leisure passengers, (i) increasing airport charges of one type of passengers will reduce the Nash equilibrium quantities of this type but raise the output levels of the other type; (ii) an increase in leisure passenger airport charge will reduce the total traffic volume; and (iii) an increase in business passenger airport charge may reduce or increase the total traffic volume.*

Proof: Part (i) has been proved above. See Appendix D.3 for the proof of parts (ii) and (iii).

One interesting observation from Proposition 6.3 is that increasing airport charge of high value passengers may not suppress the total traffic volume. Again, runway congestion and terminal congestion play different roles. If runway congestion is the only concern, an increase in total traffic volume may occur if the time value of business passengers is sufficiently higher than that of leisure passengers. However, if there is no runway congestion but terminal congestion, the induced extra leisure passengers might dominate the reduction in business passengers only when $\beta_B < \beta_L$ and $v_B \beta_B > v_L \beta_L$.

Then we examine the impact of uniform airport charge on traffic volume and the results are presented in Proposition 6.4.

Proposition 6.4: *When both types of passengers are levied a uniform airport charge, i.e.* $\tau_B = \tau_L = \tau$ *, an increase in airport charge will reduce the number of leisure passengers and the total number of passengers, but may reduce or increase the number of business passengers.*

Proof: See Appendix D.4.

6.5.2 Equilibrium airport charge

From (6.10), (6.11) and (6.16), it is straightforward to show that if the airport charges discriminative tolls between business and leisure passengers, the first-best can be achieved with the following toll.

$$\tau_h^* = \frac{1}{n}\rho_h^{'} Q_h^* + \left(1 - \frac{1}{n}\right)\left(\Gamma^* + \phi_h^*\right) - \delta \frac{\partial S_c^*}{\partial Q_h^*} \quad \forall h \in \{B, L\}.$$

However, in reality, an airport is constrained by practical and legal barriers which make charging discriminative tolls infeasible. Even if an airport is allowed to do so, implementation of such toll scheme might be of even a greater challenge, as the airport usually is not able to distinguish business and leisure passengers while airlines have incentive to cheat as one type of passenger will be charged a lower toll than the other.[63] Therefore, the rest of the paper will focus only on uniform airport charge which leads to the second-best outcomes.

The second-best airport charge can be obtained by taking first-order condition of the social welfare function:

$$\begin{aligned} \frac{dW}{d\tau} &= \frac{\partial W}{\partial Q_B}\frac{\partial Q_B^N}{\partial \tau} + \frac{\partial W}{\partial Q_L}\frac{\partial Q_L^N}{\partial \tau} \\ &= \left(p_B^N - \Gamma^N - \phi_B^N + \delta \cdot \frac{\partial S_c^N}{\partial Q_B}\right)\frac{\partial Q_B^N}{\partial \tau} + \left(p_L^N - \Gamma^N - \phi_L^N + \delta \cdot \frac{\partial S_c^N}{\partial Q_L}\right)\frac{\partial Q_L^N}{\partial \tau} = 0 \end{aligned} . \qquad (6.19)$$

From (6.16), we know that

$$p_h^N = -\frac{1}{n}(\rho_h^{'} Q_h^N - \Gamma^N - \phi_h^N) + \tau_h \quad \forall h \neq k.$$

Replace p_B^N and p_L^N in (6.19) with the above expression and rearrange the equation. The optimal (second-best) uniform airport charge should satisfy the following pricing rule.[64]

[63] An airport might discriminate passengers by charging based on the value of air ticket. For example, the airport can charge at a fixed proportion of the air ticket price. However, this pricing scheme may not achieve first-best either and in many cases it may cause more distortions in congestion pricing than a uniform toll, because the first-best toll levied on business passengers – suppose they pay higher ticket price –is not necessarity higher than that levied on leisure passengers.

[64] Note that in this section, since we are denoting the second best airport charge, in terms of magnitudes, all the Nash equilibrium outcomes are at the second best levels.

$$\tau^{sb} = \left(1-\frac{1}{n}\right)\Gamma^N + \left(1-\frac{1}{n}\right)\frac{\frac{\partial Q_B^N}{\partial \tau}\phi_B^N + \frac{\partial Q_L^N}{\partial \tau}\phi_L^N}{\frac{\partial Q^N}{\partial \tau}} + \frac{1}{n}\frac{\frac{\partial Q_B^N}{\partial \tau}\rho_B^{'} Q_B^N + \frac{\partial Q_L^N}{\partial \tau}\rho_L^{'} Q_L^N}{\frac{\partial Q^N}{\partial \tau}} - \delta\frac{\frac{\partial Q_B^N}{\partial \tau}\frac{\partial S_c^N}{\partial Q_B} + \frac{\partial Q_L^N}{\partial \tau}\frac{\partial S_c^N}{\partial Q_L}}{\frac{\partial Q^N}{\partial \tau}}. \qquad (6.20)$$

The optimal airport charge has four components. The first term on the right-hand side of (6.20) is the part of the marginal external runway congestion cost which is not internalized by airlines. This component is consistent to the literature. The second term is new. It is the weighted marginal external terminal cost not internalized by airlines. The last two components are adjustment on market power and concession surplus, respectively. They are both weighted terms and the weights are determined by the marginal impact of airport charge on business and leisure passengers' traffic volumes. Because $\partial Q_B^N / \partial \tau$, the marginal impact of airport charge on business passenger volume, can be either positive or negative, while $\partial Q_L^N / \partial \tau$ is always negative, the sign of these weighted terms are ambiguous. For example, the market power adjustment is negative (a downward correction) when $\partial Q_B^N / \partial \tau$ is negative, but it can be positive when $\partial Q_B^N / \partial \tau$ is positive, leading to a markup on airport charge in order to protect business travelers from overcrowded runway and terminal by squeezing out leisure passengers.

A closer look at the second and the fourth terms reveals that their magnitudes and signs rely on the interaction between $\partial Q_h^N / \partial \tau$ and other parameters, such as the relative schedule delay cost and the cut-off dwell times, $\hat{T}_{d1}$ and $\hat{T}_{d2}$. Denoting Φ as the weighted external terminal cost which is part of the second term, we can rewrite it as below:

$$\Phi = \frac{\frac{\partial Q_B^N}{\partial \tau}\phi_B^N + \frac{\partial Q_L^N}{\partial \tau}\phi_L^N}{\frac{\partial Q^N}{\partial \tau}} = \phi_L^N + (\phi_B^N - \phi_L^N)\frac{\frac{\partial Q_B^N}{\partial \tau}}{\frac{\partial Q^N}{\partial \tau}} = \phi_B^N - (\phi_B^N - \phi_L^N)\frac{\frac{\partial Q_L^N}{\partial \tau}}{\frac{\partial Q^N}{\partial \tau}}. \qquad (6.21)$$

Note first that the sign of $\phi_B^N - \phi_L^N$ is the same as the sign of $\beta_B - \beta_L$. The second equal sign of (6.21) suggests that $\Phi > \phi_B^N$ if and only if $\beta_L > \beta_B$ and $\Phi < \phi_B^N$ if and only if $\beta_L < \beta_B$. However, according to the first equal sign, the relationship between Φ and ϕ_L^N depends on the sign of $\partial Q_B^N / \partial\tau$ as well. In particular, given $\partial Q_B^N / \partial\tau > 0$, $\Phi > \phi_L^N$ if and only if $\beta_L > \beta_B$ and $\Phi < \phi_L^N$ if and only if $\beta_L < \beta_B$; when $\partial Q_B^N / \partial\tau < 0$, the opposite will hold. We summarize the results of such analysis in Table 6.1.

Table 6.1 The ranges of Φ at second best

	$\beta_L < \beta_B$	$\beta_L > \beta_B$
$\partial Q_B^N / \partial\tau > 0$	$\Phi < \phi_L^N < \phi_B^N$ $\Phi < 0$ is possible	$\Phi > \phi_L^N > \phi_B^N$
$\partial Q_B^N / \partial\tau < 0$	$\phi_L^N < \Phi < \phi_B^N$	$\phi_L^N > \Phi > \phi_B^N$

Ideally, each passenger should be charged at the uninternalized cost they impose on the others. This is exactly how passengers are charged for runway congestion, the first component of airport charge. Moreover, for terminal tolls, it is also true under the first-best situation where the airport set tolls for business and leisure passengers separately. However, this is not the case for the terminal charge in the second-best case, as passengers of different types can not be distinguished for charges at the terminal. As a result, a particular passenger will have to pay more or less than the amount of externality she imposes on the others. In the former situation, the passenger is 'over charged' while in the latter situation, she is 'under charged'. The details are stated in Proposition 6.5.

Proposition 6.5: *When $\partial Q_B^N / \partial\tau < 0$, the passenger type with higher relative schedule delay cost will be over charged relative to the uninternalized terminal cost imposed on other passengers, while the other type under charged. When $\partial Q_B^N / \partial\tau > 0$ and $\beta_B > \beta_L$, all passengers will be under charged and, if the values of time differ dramatically between the two passenger types, they may be subsidized instead. When $\partial Q_B^N / \partial\tau > 0$ and $\beta_B < \beta_L$, all passengers will have to pay more than the uninternalized terminal cost they bring about from* *travelling.*

When $\beta_B = \beta_L$, we can show that $\Phi = \phi_B^N = \phi_L^N$. That is, regardless there is one additional business or leisure passenger, the marginal external terminal cost of this extra passenger is the same. Then the first and the second terms in equation (6.20) can be combined into one single uninternalized airport congestion cost, which has been widely derived from the literature. Thus, the conventional modeling approach which does not separate passenger runway and terminal costs will lose important features unless the relative schedule delay costs are the same across passenger types.

Similar analysis can be applied to the weighed marginal concession surplus, i.e. the fraction of the last term in (6.20), which is denoted as Σ. That is,

$$\Sigma = -\frac{\frac{\partial Q_B^N}{\partial \tau} S_{cB}^N + \frac{\partial Q_L^N}{\partial \tau} S_{cL}^N}{\frac{\partial Q^N}{\partial \tau}} = -S_{cL}^N - \left(S_{cB}^N - S_{cL}^N\right)\frac{\frac{\partial Q_B^N}{\partial \tau}}{\frac{\partial Q^N}{\partial \tau}} = -S_{cB}^N + \left(S_{cB}^N - S_{cL}^N\right)\frac{\frac{\partial Q_L^N}{\partial \tau}}{\frac{\partial Q^N}{\partial \tau}},$$

where $S_{ch}^N = \partial S_c^N / \partial Q_h > 0 \quad \forall h \in \{B, L\}$.

As discussed in Section 6.4, the sign of $S_{cB}^N - S_{cL}^N$ depends on the dwell time for passengers arriving the airport at t_m, i.e. $T_d(t_m)$. In particular, if $T_d(t_m) \leq \hat{T}_{d1}$ or $T_d(t_m) \geq \hat{T}_{d2}$, $S_{cB}^N - S_{cL}^N \geq 0$; otherwise, $S_{cB}^N - S_{cL}^N < 0$. Then, we can show that the inclusion of concession surplus does not always lead to a markdown in airport charge (Table 6.2).

Table 6.2 The ranges of Σ at second best

	$T_d(t_m) < \hat{T}_{d1}$ or $T_d(t_m) > \hat{T}_{d2}$	$\hat{T}_{d1} < T_d(t_m) < \hat{T}_{d2}$
$\partial Q_B^N / \partial \tau > 0$	$\Sigma > -S_{cL}^N > -S_{cB}^N$ $\Sigma > 0$ is possible	$\Sigma < -S_{cL}^N < -S_{cB}^N$
$\partial Q_B^N / \partial \tau < 0$	$-S_{cL}^N > \Sigma > -S_{cB}^N$	$-S_{cL}^N < \Sigma < -S_{cB}^N$

Given that $\partial Q_B^N / \partial\tau > 0$, when $T_d(t_m)$ is below $\hat{T}_{d1}$ or above $\hat{T}_{d2}$, Σ may become positive, leading to an upward correction on the airport charge. In other cases, taking concession surplus into account will impose a downward pressure on the airport charge. This is consistent with the findings in chapter 5, but the underlying reasoning differs. Since terminal and runway congestions are not separated in chapter 5, the direction of the correction due to concession surplus is affected by the level of congestion delay which is determined by total traffic volume. In the present setting, however, as $T_d(t_m)$ = Q_h/s for any $\beta_h > \beta_k$, the direction of this correction is solely related to the traffic volume of the passengers who have higher relative early schedule delay cost but not the volume of the other passenger type.

In equation (5.21), the impact of concession surplus on airport charge consists of two components: a downward correction on congestion toll given in the second line of (5.21) and a correction indicated by the last term of (5.21) which is equal to the expected concession surplus weighted across passenger types. A closer look at the last term of (6.20) reveals two components as well. For example, when $\beta_B > \beta_L$, S_{cB}^N equals to the expected concession surplus of one additional business passenger, $\int_c^{\bar{u}} \overline{G}_B(x; Q_B/s)dx$, plus the increment in concession surplus of leisure passengers due to adding one more business passenger into the system, $\int_c^{\bar{u}} \overline{G}_L(x; Q/s) - \overline{G}_L(x; Q_B/s)dx$. As business passengers arrive later than leisure passengers, one more business passenger contributes its own expected surplus from concession purchase and at the same time pushes the arrival time of each leisure passenger a little bit earlier and hence leads to higher dwell time of all the leisure passengers. However, S_{cL}^N simply equals to the expected concession surplus of one additional leisure passenger, $\int_c^{\bar{u}} \overline{G}_L(x; Q/s)dx$. Thus, when $\beta_B > \beta_L$, the last term of (6.20) can be rewritten as:

$$-\delta \frac{\frac{\partial Q_B^N}{\partial\tau} \int_c^{\bar{u}} \overline{G}_L\left(x; \frac{Q}{s}\right) - \overline{G}_L\left(x; \frac{Q_B}{s}\right)dx}{\frac{\partial Q^N}{\partial\tau}} - \delta \frac{\frac{\partial Q_B^N}{\partial\tau} \int_c^{\bar{u}} \overline{G}_B\left(x; \frac{Q_B}{s}\right)dx + \frac{\partial Q_L^N}{\partial\tau} \int_c^{\bar{u}} \overline{G}_L\left(x; \frac{Q}{s}\right)dx}{\frac{\partial Q^N}{\partial\tau}}. \quad (6.22)$$

Although the first term of (6.22) corresponds to the second line of (5.21), indicating the correction due to changes of dwell time, it is not a clear-cut downward adjustment. Instead, when $\partial Q_B^N / \partial \tau$ is positive, this term becomes a markup.[65] That is, to increase the dwell time of leisure passengers, the airport has incentive to charge more (less) and attract more business passengers when $\partial Q_B^N / \partial \tau$ is positive (negative). The second term of (6.22) corresponds to the last term of (5.21), indicating the correction due to changes of expected concession surplus as the composition of the two passenger types changes, which can be either positive or negative.

6.6 Concluding remarks

In this chapter, we propose a framework that treats terminal congestion and runway congestion separately, and study its implication on the design of optimal airport charge. To capture the difference between these two types of congestion, we adopt a deterministic bottleneck model for the terminal and a conventional congestion model for the runways. The inclusion of terminal congestion leads to first-best results different from the literature. In particular, the welfare (excluding concession surplus) optimal uniform fare *does not* yield the first-best outcome. First-best can only be achieved through discriminative fares. The first-best fare charged on the business passengers is higher than that on the leisure passengers if and only if the relative schedule delay cost of business passengers is higher than leisure passengers. If concession surplus is in concern, the comparison between first-best fares charged on the two passenger types depends on their relative early schedule delay costs as well as the dwell time of passengers arriving at the airport at time t_m.

We also study the impact of airport charge on equilibrium traffic levels and derive the optimal uniform airport charge. When the airport discriminates business and leisure passengers, increasing airport charge of one type of passengers will reduce the equilibrium quantity of this type but raise that of the other type. It can also be shown that an increase in

[65] When $\beta_B < \beta_L$, this term will be a markdown.

leisure passenger airport charge will reduce the total traffic volume but an increase in business passenger airport charge may reduce or increase the total traffic volume. On the other hand, when both types of passengers are levied a uniform airport charge, an increase in airport charge will reduce the number of leisure passengers and the total number of passengers, but may reduce or increase the number of business passengers. Furthermore, the structure of optimal uniform airport charge suggests that in terms of the terminal charge, passengers will be under charged or over charged in various conditions. When the number of business passengers decreases in airport charge, the passenger type with higher relative schedule delay cost will be over charged, while the other type under charged. However, when the number of business passengers increases in airport charge, both passenger types will be under charged (or even subsidized) if the relative schedule delay cost of business passengers is higher than that of leisure passengers. When comparing the airport pricing rule with the one found in chapter 5, we find that when the volume of business passengers increases in airport charge, to lengthen leisure passengers' dwell time and hence increase their chance of purchasing concession goods, the airport will raise rather than reduce the airport charge.

As clearly shown in this chapter, the separation of terminal congestion and runway congestion gives rise to some new insights with respect to the optimal airport pricing rule. One advantage of this new framework is that it explicitly models passengers' behavior in the terminal and relates concession purchase with dwell time rather than congestion delay. Given the increasing importance of airport concession, it will be beneficial to have a clearer picture of the behavior of different passenger groups in the terminal. This paper offers a first step of this attempt. Potential future studies may lie in more general settings such as allowing multiple flight departure times, incorporating flight scheduling as endogenous decisions and including connecting flights into the picture. Another direction might be to empirically test whether the predictions of passenger behavior fit the reality.

7 Conclusions

For the topic of container port competition and hinterland access conditions, this dissertation contributes to the related literature by providing more comprehensive and complete analysis for previously identified problem. First of all, we find the impacts of road capacity differ between modes of competition and it might be more reasonable to assume that ports compete in quantities rather than in prices. Our quantity competition model predicts that an increase in road capacity by an intermodal chain will likely benefit its port while negatively affecting the rival port. The price competition model established in the literature however predicts that an increase in road capacity by an intermodal chain will reduce its port's profit while the rival port's profit may increase. The policy implication of this finding is that if the mode of competition is in quantities, the regional government is less likely to receive objection to invest in roads; whilst, ports are less likely to advocate road expansion projects if they compete in price. We discuss in chapter 2 why quantity competition is more likely to prevail and the empirical evidence presented in chapter 3 is also consistent to the prediction of quantity competition model. In addition, the quantity competition model provides an explanation for the prevailing observation that port-related capacities are 'on the high side' in port ranges where ports compete vigorously. The high capacities may nevertheless be socially desirable globally due to the positive externality on roads in the rival's region.

Second, in the literature, evidence on the adverse impact of road congestion on port performance is limited to the statements from stakeholders in surveys. As an initial effort to measure and quantify such impacts, chapter 3 empirically explores the association between container ports' throughputs and road congestion as well as road supply around the ports and their respective rivals. A port's container throughput is negatively associated with the congestion delays on its own urban roads, but positively associated with delays on its rival's roads; and the impact of its own roads is stronger than the rival's roads. This finding conforms to the modeling base of chapter 2. As mentioned above, the impact of road capacity expansion is consistent to our model prediction: via the change of road congestion, an increase in road capacity implies an increase in throughput by the port nearby but a decrease in throughput by its rival port. Meanwhile, we discover an existent of channels other than road congestion through which additional road supply could adversely affect port throughputs, which leads to a potential opportunity for future research provided that better data are available. The estimated relationship between road congestion and port technical efficiency confirms to our hypothesis, i.e. road congestion is negatively correlated with port efficiency. However, the impact is much stronger for small ports but negligible for large ports. Therefore, when considering expanding roads around a port, the local government and port management should be cautious that the effectiveness of adding roads might be jeopardized if road congestion is not the dominant driving force for container throughput reduction and if the port's operational scale is very large and possibly serves as the primary port of entry. Partly due to the data constraint, we do not obtain a clear-cut association between rail services and port efficiency, but it is consistent to the literature that provision of on-dock facility is negatively correlated with port efficiency. Thus, investing in on-dock facility might be a bad idea for ports located in the center of a city where land space is highly constrained.

Third, as another heatedly discussed congestion mitigation method, road toll has not been studied in the context of port competition. Thus, chapter 2 provides a relatively complete analysis on the impact of two road toll systems, discriminative tolls and fixed-ratio toll schedule, on port competition and road congestion. Similar to road capacity expansion, increasing commuter toll alone helps the port to win an advantage over it rival. However, raising the truck toll will do exactly the opposite. As a result, the impact of fixed-ratio toll schedule depends on the relative values of time between commuters and shippers. Unlike capacity expansion, while a high toll by a region relieves its road congestion, it may increase road congestion in the rival region. From the standpoint of local governments, marginal external cost pricing on roads will no longer be optimal when ports compete. The deviation depends on the toll systems used as well as the time values of road users. If a discriminative toll system is used, marginal external cost pricing applies to commuters but trucks must be charged below this amount to compete for port-related traffic with the rival chain. If a fixed-ratio toll schedule is employed, the road users will both be charged above the marginal external cost if and only if the shippers' value of time is so high that port-related traffic would increase as the toll increases.

Fourth, noting that chapter 2 is based on the assumption that local governments of competing intermodal chains make independent and simultaneous decisions, chapter 4 goes one step further by investigating the issue of inter-governmental coordination on their investment decisions. By including the captive catchment area of each port into the model, we discover that improvement in transport infrastructure in a captive catchment area affects public ports and private ports differently. This is the driving force for differentiated strategic interaction among regional governments when we compare our results across port ownership types. An increase in investment in the port region will reduce the welfare of the rival port region but improve the welfare of the common inland region if ports are public. The opposite may occur if ports are private. For regional governments of public ports, their incentive of infrastructure investment is the lowest when two port regions coordinate. They will invest more once at least one of them coordinates with the inland region. The inland region, on the other hand, always has higher incentive to invest at lower level of coordination. Given private ports, the port regions' incentive of investment may be the highest when they coordinate while investment may be at the low end if the port region is coordinated with the inland. In terms of future study, a natural extension is to examine mixed-ownership ports and local governments' incentives to form various types of coalitions and to predict with the theoretical model whether and in which forms coalition will occur.

Abstracting away the competition between two congested transportation facilities, chapters 5 and 6 focus on a stand-alone congestible airport and the competition among users, such as airlines and passengers of different types, for scarce resources. These two chapters contribute to the literature by incorporating concession consumption, heterogeneous passengers and most importantly the relationship between concession activities and passengers' waiting time. We discuss how optimal airport charges derived in the literature should be modified to accommodate the missing pieces. Chapter 5 discovers a downward correction on congestion toll due to positive externality on concession, as lower charges can induce more traffic and hence enlarge the pool of potential concession buyers. In addition, the traditional 'negative component' on airport charge to subsidize concession consumption may become a positive charge when two types of passenger are under consideration, business passengers' value of time is much higher and the congestion delay is high. This is because raising the airport charge could increase the share of business passengers by discouraging leisure passengers.

Thus, doing so will be beneficial to the airport if business passengers generate more concession profits or surplus. The comparison between profit-maximizing and welfare-maximizing airport reveals that welfare-maximizing airport can have more incentives to induce congestion and increase the demand of concession goods under certain conditions.

Chapter 6 takes a more innovative approach by separately modeling terminal and runway congestions and applying the bottleneck model to describe passengers' airport arrival patterns. Given that passengers have differentiated relative early schedule delay costs, the marginal terminal cost of one additional business passenger differs from one additional leisure passengers. This feature, which is omitted in the literature, is the underlying reason of the main differences between the findings in chapter 6 and those in the literature. For example, given that concession is excluded from the analysis, first-best outcomes can only be achieved through discriminative airfares, together with airport charges which distinguish passengers of different types, rather than via the welfare optimal uniform airfare and airport charge. As a result, given a uniform airport charge, although the runway congestion can be fully internalized, the marginal terminal external costs will never be fully internalized and hence either over-internalization or under-internalization will occur depending on passengers' values of time and relative early schedule delay costs. This is fundamentally different from chapter 6 in which congestion externality (excluding the impacts on concession activities) will eventually internalized by airlines together with uniform welfare optimal airport charges.

Note that both chapters 2 and 3 investigate heterogeneous users of congestible facilities and discover that users can be charged more than the amount of externality they impose on the facility. In chapter 2, ignoring ports' market power, under the fixed-ratio toll, the shippers will eventually pay more than the congestion externality they impose on the road system. However, the over-internalization observed in chapter 2 is mainly because local commuters are not considered by port but by local government and hence it occurs regardless the relative values of time between commuters and shippers, while the over internalization of chapter 6, as mentioned above, is due to different terminal cost structure of two passenger types since airlines as well as the airport take into account both types when setting airfares and airport charge. Moreover, in chapter 6, it is possible that terminal externality imposed by the two passenger types will both be under-internalized, while this is not the case in chapter 2. Thus, chapters 2 and 3 provide two different scenarios in which over or under-internalization of externality would occur when user heterogeneity exists.

In the future, chapters 5 and 6 may be extended to embed the link between concession and congestion into the discussion of various types of regulations. Empirical test on passengers' airport arrival pattern would be also crucial in examining the degree that the model proposed in chapter 6 is close to reality. As chapter 6 considers only one scheduled departure time, it would be interesting to study the inter-temporal interaction between airlines as well as passengers by introducing multiple scheduled departure times and allowing flight scheduling as endogenous variables of airlines.

References

Appold, S.J., and Kasarda, J.D. (2006) The Appropriate Scale of US Airport Retail Activities, *Journal of Air transport Management*, 12, 277-287.

Arnott, R., de Palma, A. and Lindsey, R. (1990) Economics of a Bottleneck, *Journal of Urban Economics*, 27(1), 111-130.

Arnott, R., de Palma, A. and Lindsey, R. (1993) A Structural Model of Peak-Period Congestion: A Traffic Bottleneck with Elastic Demand, *The American Economic Review*, 83(1), 161-179.

Arnott, R., de Palma, A., and Lindsey, R. (1994) The Welfare Effects of Congestion Tolls with Heterogeneous Commuters, *Journal of Transport Economics and Policy*, 28 (2), 139–161.

Arnott, R. and Kraus, M. (2003) Principles of Transport Economics, in Hall, R. (ed.), *Handbook of Transportation Science*, Kluwer Academic Publishers, Dordecht, pp. 700-737.

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.

Basso, L.J. (2008) Airport Deregulation: Effects on Pricing and Capacity, *International Journal of Industrial Organization*, 26 (4), 1015-1031.

Basso, L.J., and Zhang, A. (2007) Congestible Facility Rivalry in Vertical Structures, *Journal of Urban Economics*, 61 (2), 218–237.

Berechman, J. (2007) The Social Costs of Global Gateway Cities: The Case of the Port of New York, Paper presented at the *International Conference on Gateways and Corridors*, Vancouver, BC, May 2007.

Bichou, K. and Gray, R. (2005) A Critical Review of Conventional Terminology for Classifying Seaports, *Transportation Research Part A*, 39, 75-92.

Brander, J.A. and Zhang, A. (1990) Market Conduct in the Airline Industry: An Empirical Investigation, *RAND Journal of Economics*, 21, 567-583.

Brander J.A., and Zhang, A. (1993) Dynamic Behaviour in the Airline Industry, *International Journal of Industrial Organization*, 11, 407–435.

Bresnahan, T.F. (1989) Empirical Studies of Industries with Market Power, in Schmalensee, R. and R. Willig (eds.), *Handbook of Industrial Organization*, Vol. 2, pp. 1011-1057, North-Holland, Amsterdam.

Brown, G. (1992) Tourism and Symbolic Consumption, in Johnson, P. and Thomas, B. (ed.) *Choice and Demand in Tourism*, Mansell, London, 51–71.

Brueckner, J.K. (2002) Airport Congestion When Carriers Have Market Power, *American Economic Review*, 92(5), 1357-1375.

Brueckner, J.K., and Van Dender, K. (2008) Atomistic Congestion Tolls at Concentrated Airports? Seeking a Unified View in the Internalization Debate, *Journal of Urban Economics*, 64 (2), 288–295.

Buendia, M., and de Barros, A.G. (2008) Critical Update of the Post-Security Retail Space Design in Passenger Departure Terminals, *Aerlines Magazine*, 40.

Bulow, J.I., Geanakoplos, J.D. and Klemperer, P.D. (1985) Multimarket Oligopoly: Strategic Substitutes and Complements, *Journal of Political Economy*, 93, 488-511.

Cantos-Sanchez, P., Moner-Colonques, R., Sempere-Monerris, J.J. and Alvarez-SanJaime, O. (2010) Vertical Integration and Exclusivities in Maritime Freight Transport, *Working paper*, University of Valencia, Spain.

Castillo-Manzana, J.I. (2010) Determinants of Commercial Revenues at Airports: Lessons Learned from Spanish Regional Airports, *Tourism Management*, 31 (6), 788 – 796.

Cervero, R. (2002) Induced Travel Demand: Research Design, Empirical Evidence, and Normative Policies, *Journal of Planning Literature,* 17(3), 2-20.

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.

Cullinane, K., Ji, P. and Wang, T. (2005) The Relationship Between Privatization and DEA Estimates of Efficiency in the Container Port Industry, *Journal of Economics and Business*, 57(5), 433–462

Cullinane, K., Song, D., Ji, P. and Wang, T. (2004) An Application of DEA Windows Analysis to Container Port Production Efficiency, *Review of Network Economics*, 3(2), 184-206

Cullinane, K. and Song, D. (2006) Estimating the Relative Efficiency of European Container Ports: A Stochastic Frontier Analysis, in Cullinane, K. and W. Talley (eds.), *Port Economics: Research in Transportation Economics*, Vol. 16, pp. 85-115, Elsevier JAI Press, Oxford, UK.

Czerny, I.A. (2006) Price-Cap Regulation of Airports: Single-Till versus Dual-Till, *Journal of Regulatory Economics*, 30 (1), 85–97

Czerny, I.A. (2007) Maritime Ports, Capacities and Door-to-Door Services, Paper presented at the *International Conference on Gateways and Corridors*, Vancouver, BC, May 2007.

Czerny, I.A. (2011) Airport Pricing and Concession Types, WHU – Working Paper Series in Economics.

Czerny, I. A., Hoffler, F. and Mun, S. (2013) Port Competition and Welfare Effect of Strategic Privatization, Working paper, WHU-Otto Beisheim School of Management.

Czerny, I.A., and Zhang, A. (2011) Airport Congestion Pricing and Passenger Types, *Transportation Research Part B*, 45 (3), 595 – 604.

Czerny, A.I. and Zhang, A. (2012) Airport Congestion Pricing When Airlines Price Discriminate, Working paper.

Containerization International Yearbook (2001~2010), National Magazine Company, London.

Daniel, J. I. (2001) Distributional Consequences of Airport Congestion Pricing, *Journal of Urban Economics*, 50, 230-258.

De Borger, B. and De Bruyne, D. (2011) Port Activities, Hinterland Congestion, and Optimal Government Policies: The Role of Vertical Integration in Logistic Operations, *Journal of Transport Economics and Policy*, 45(2), 247-275.

De Borger, B., Dunkerley, F. and Proost, S. (2007) Strategic Investment and Pricing Decisions in a Congested Transport Corridor, *Journal of Urban Economics*, 62(2), 294-316.

De Borger, B. and Proost, S. (2012) Transport Policy Competition between Governments: A Selective Survey of the Literature, *Economics of Transportation*, 1, 35-48

De Borger, B., Proost, S. and Van Dender, K. (2005) Congestion and Tax Competition on a Parallel Network, *European Economic Review*, 49(8), 2013-2040.

De Borger, B., Proost, S. and Van Dender, K. (2008) Private Port Pricing and Public Investment in Port and Hinterland Capacity, *Journal of Transport Economics and Policy*, 42(3), 527-561.

De Borger, B. and Van Dender, K. (2006) Prices, Capacities and Service Levels in a Congestible Bertrand Duopoly, *Journal of Urban Economics*, 60, 264-283.

De Langen, P. W. (2008) Ensuring Hinterland Access: The Role of Port Authorities, OECD/ITF discussion paper No. 2008-11, Paris

Downs, A. (1962) The Law Of Peak-Hour Express-Way Congestion, *Traffic Quarterly,* 16(3), 393-409.

Dube, L., and Menon, K. (2000) Ensuring Greater Satisfaction by Engineering Salesperson Response to Customer Emotions, *Journal of Retailing*, 76, 285–308.

Duranton, G., and Turner, M.A. (2009) The Fundamental Law of Road Congestion: Evidence from U.S. Cities, NBER working paper No. w15376.

Entwistle, M. (2007) Customer Service and Airport Retail: Stimulate Passenger Spending, *Journal of Airport Management*, 1, 151–157.

Fan, L., Koehler, M.M. and Wilson, W.W. (2011) Intermodalism and New Trade Flows, in: Talley, W.K. (ed.), *The Blackwell Companion to Maritime Economics*, pp. 121-137, Wiley-Blackwell, West Sussex, U.K.

FHWA (2000) Chapter IX Traffic Operations, *Comprehensive Truck Size and Weight Study*, Vol. III, Federal Highway Administration (FHWA), the United States.

FHWA (2004) *Traffic congestion and reliability: linking solutions to problems*, Federal Highway Administration (FHWA), the United States.

Fleming, D.K. and Baird, A.J. (1999) Some Reflections on Port Competition in the United States and Western Europe, *Maritime Policy and Management*, 26(4), 383-394.

Garcia-Alonso, L., and Sanchez-Soriano, J. (2009) Port Selection from a Hinterland Perspective, *Maritime Economics and Logistics*, 11(3), 260-269.

Geuens, M., Vantomme, D., and Brengman, M. (2004) Developing a Typology of Airport Shoppers, *Tourism Management*, 25, 615-622.

Golob, T.F., and Regan, A.C. (2000) Freight Industry Attitudes towards Policies to Reduce Congestion, *Transportation Research Part E*, 36E(1), 55-77.

Goodwin, P. B. (1996) Empirical Evidence on Induced Traffic, *Transportation*, 23(1), 23-54.

Graham, A. (2008) *Managing Airports: An International Perspective*, Butterworth Heinemann, Oxford.

Graham, A. (2009) How Important Are Commercial Revenues to Today's Airports? *Journal of Air Transport Management*, 15, 106–111.

Heaver, T.D. (2002) Supply Chain and Logistics Management: Implications for Liner

Shipping, in Grammenos, C. (ed.), *Maritime Economics and Business,* Chapter 17, London, Lloyds of London Press.

Heaver, T.D. (2006) The Evolution and Challenges of Port Economics, in Cullinane, K. and W. Talley (eds.), *Port Economics: Research in Transportation Economics*, Vol. 16, pp. 11-41, Elsevier JAI Press, Oxford, UK.

Hoff, A. (2007) Second Stage DEA: Comparison of Approaches for Modeling the DEA Score, *European Journal of Operations Research*, 181, 425–435.

Huang, W. H. and Kuai, L. (2006) The In-Flight Shopper, *Journal of Air Transport Management*, 12, 207–211.

Jones, I., Viehoff, I., and Marks, P. (1993) The Economics of Airports Slot, *Fiscal Studies*, 14, 37-57.

Jula, P. and Leachman, R. C. (2011) Long- and Short-Run Supply-Chain Optimization Models for the Allocation and Congestion Management of Containerized Imports from Asia to the United States, *Transportation Research Part E*, 47, 593-608.

Kitamura, R. (2009) The Effects of Added Transportation Capacity on Travel: A Review of Theoretical and Empirical Results, *Transportation*, 36, 645-762.

Kratzsch, U., and Sieg, G. (2011) Non-Aviation Revenues and Their Implications for Airport Regulation, *Transportation Research Part E*, 47, 755-763.

Kreps, D.M. and Scheinkman, J.A. (1983) Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes, *Bell Journal of Economics*, 14(2), 326-337.

Kreukels, T. and Wever, E. (1998) North Sea Ports in Competition, in Kreukels, T. and E.Wever (eds.), *North Sea Ports in Transition: Changing Tides*, Royal Van Gorcum.

Levinson, M. (2006) *The Box: How the Shipping Container Made the World Smaller and the World Economy Bigger,* Princeton University Press, Princeton, New Jersey.

Lindsey, R. (2007) Congestion Relief: Assessing the Case for Road Tolls in Canada, *C.D. Howe Institute Commentary*, No. 248.

Lindsey, R. (2008) Prospects for Urban Road Pricing in Canada, *Brookings-Wharton Papers on Urban Affairs*, 2008, pp. 235-293, Brookings Institution Press.

Lirn, T.C., Thanopoulou, H.A., Beynon, M.J. and Beresford, A. (2004) An Application of AHP on Transhipment Port Selection: A Global Perspective, *Maritime Economics and Logistics*, 6(1), 70-91.

Lovejoy, W.S. (1998) Integrated Operations: A Proposal for Operations Management

Teaching and Research, *Production and Operations Management*, 7 (2), 106–124.

Lozano, S. and Gutlerrez, E. (2011) Slacks-Based Measure of Efficiency of Airports with Airplanes Delays as Undesirable Outputs, *Computers & Operations Research*, 38(1), 131-139.

Luo, M. and Grigalunas, T. (2003) A Spatial-Economic Multimodal Transportation Simulation Model for US Coastal Container Ports, *Maritime Economics and Logistics*, 5, 158-178.

Maloni, M. and Jackson, E.C. (2005) North American Container Port Capacity: An Exploratory Analysis, *Transportation Journal*, 44(3), 1-22.

Mayer, C. and Sinai, T. (2003) Network Effects, Congestion Externalities, and Air Traffic Delays: Or Why Not All Delays Are Evil, *American Economic Review*, 93, 1194-1215.

McDonald, J. (2009) Using Least Squares and Tobit in Second Stage Dea Efficiency Analyses. European Journal of Operations Research, 197, 792–798.

Menezes, F.M., Pracz, M. and Tyers, R. (2007) Strategic Interaction amongst Australia's East Coast Ports, *The Australian Economic Review*, 40(3), 267-278.

Moffatt & Nichol and BST Associates (2007) *Container Diversion and Economic Impact Study: Effects of Higher Drayage Costs at San Pedro Bay Ports*. Report prepared for the Port of Los Angeles and the Port of Long Beach.

Morrison, S. A. (1987) The Equity and Efficiency of Runway Pricing, *Journal of Public Economics*, 34 (1), 45–60.

Morrison, W. G. (2009) Real Estate, Factory Outlets and Bricks: A Note on Non-Aeronautical Activities at Commercial Airports, *Journal of Air Transport Management*, 15, 112–115.

Neven, D., Roller, L-H. and Zhang, Z. (1999) Union Power and Product Market Competition: Evidence from the Airline Industry, CIG Working Paper Series FS , 4, 99-05.

Nir, A., Lin, K. and Liang, G. (2003) Port Choice Behaviour – from the Perspective of the Shippers, *Maritime Policy and Management*, 30(2), 165-173.

Noland, R.B., and Lem, L.L. (2002) A Review of the Evidence for Induced Travel and Changes in Transportation and Environmental Policy in the US and the UK, *Transportation Research Part D*, 7, 1-26.

Notteboom, T.E. (1997) Concentration and the Load Center Development in the European Container Port System, *Journal of Transport Geography*, 5, 99-115.

Notteboom, T.E. (2004) Container Shipping and Ports: An Overview, *Review of Network Economics*, 3(2), 86-106.

Notteboom, T.E. (2006a) Chapter 19 Concession Agreements as Port Governance Tools, *Research in Transport Economics*, 17, 437-455.

Notteboom, T.E. (2006b) The Time Factor in Liner Shipping Services, *Maritime Economics and Logistics*, 8, 19-39.

Notteboom, T.E. and Rodrigue, J-P. (2005) Port Regionalization: Towards a New Phase in Port Development, *Maritime Policy and Management* 32(3), 297-313.

Notteboom, T.E. and Rodrigue, J-P. (2008) Containerisation, Box Logistics and Global Supply Chains: The Integration of Ports and Liner Shipping Networks, *Maritime Economics and Logistics*, 10, 152-174.

Oum, T.H., Zhang, A., and Zhang, Y. (2004) Alternative Forms of Economic Regulation and Their Efficiency Implication for Airports, *Journal of Transport Economics and Policy*, 38(2), 217-246.

Panayides, P. M., Maxoulis, C. N., Wang, T.-F. and Ng, K. Y. A. (2009) A Critical Analysis of DEA Applications to Seaport Economic Efficiency Measurement, *Transport Reviews*, 29(2), 183-206

Pels, E., Nijkamp, P., and Rietveld, P. (2003) Access to and Competition between Airports: A Case Study for the San Francisco Bay area, *Transportation Research Part A*, 37 (1), 71–83.

Pels, E., and Verhoef, E.T. (2004) The Economics of Airport Congestion Pricing, *Journal of Urban Economics*, 55 (2), 257–277.

Pettit, S.J., and Beresford, A.K.C. (2008) An Assessment of Long-Term United Kingdom Port Performance: A Regional Perspective, *Maritime Economics and Logistics*, 10, 53-74.

Pope, J.A., T.R. Rakes, Rees, L.P. and Crouch, I.W.M. (1995) A Network Simulation of High-Congestion Road-Traffic Flows in Cities with Marine Container Terminals, *The Journal of the Operational Research Society*, 46(9), 1090-1101.

Port of Los Angeles (2009) Intermodal Logistics & Ports of Los Angeles/Port of Long Beach Rail Infrastructure, *Port of Los Angeles Public Rail Workshop*.

Port of Melbourne (2005) Submission to Victorian Competition and Efficiency Commission (VCEC) inquiry into managing transport congestion, Submission No. 53, website:

http://www.vcec.vic.gov.au/CA256EAF001C7B21/0/E6F4A4748F936D23CA2570840004A41A?OpenDocument, last accessed in June 2012.

Port of New York and New Jersey (2003) Loaded Containers in TEUs and Total ExpressRail Lifts by Month, website: http://www.panynj.gov/port/monthly-loaded-containers.html, last accessed in August 2013.

Quinet, E. and Vickerman, R. (2004) *Principles of Transport Economics*, Edward Elgar.

Rowley, J., and Slack, F. (1999) The Retail Experience in Airport Departure Lounges: Reaching for Timelessness and Placelessness, *International Marketing Review*, 16, 363–375.

Robinson, R. (2002) Ports as Elements in Value-Driven Chain Systems: The New Paradigm, *Maritime Policy and Management*, 29, 241-255.

Rodrigue, J.P. and Comtois, C. (2009) Chapter 7: Transportation and the Economy, in Rodrigue, J.-P., Comtois, C. and Slack, B. (Eds.), *The Geography of Transport Systems*, 2nd ed., Routledge, New York.

Rodrigue, J.P. and Notteboom, T. (2009) The Terminalisation of Supply Chains: Reassessing the Role of Terminals in Port/Hinterland Logistical Relationships, *Maritime Policy and Management*, 36(2), 165-183.

Schmidt, G. (2005) The OM triangle, *Operations Management Education Review*, 1 (1), 87–104.

Sherman, H. D. and Zhu, J. (2006) Benchmarking with Quality-Adjusted DEA (Q-DEA) to Seek Lower-Cost High-Quality Service: Evidence from a US Bank Application, *Annals of Operations Research*, 145(1), 301-319.

Simar, L. and Wilson, P. (2011) Two-Stage DEA: Caveat Emptor, *Journal of Productivity Analysis*, 36(2), 205-218.

Slack, B. (1985) Containerization, Inter-Port Competition and Port Selection, *Maritime Policy and Management*, 12(4), 293-303.

Slack, B (2009) Chapter 9: Transport Planning and Policy, in Rodrigue, J.-P., Comtois, C. and Slack, B. (Eds.), *The Geography of Transport Systems*, 2nd ed., Routledge, New York.

Small, K.A. and Verhoef, E.T. (2007) *The Economics of Urban Transportation*, Routledge, London.

Starkie, D. (2001) Reforming UK Airport Regulation, *Journal of Transport Economics and Policy*, 35 (1), 119–135.

Starkie, D. (2002) Airport Regulation and Competition, *Journal of Air Transport Management*, 8, 63–72.

Starkie, D. (2008) *Aviation Markets: Studies in Competition and Regulatory reform*, Ashgate, London.

Sulzmaier, S. (2001) *Consumer-Oriented Business Design: The Case of Airport Management*, Physica-Verlag Heidelberg, New York.

Suykens, F. and Van De Voorde, E. (1998) A Quarter a Century of Port Management in Europe: Objectives and Tools, *Maritime Policy and Management*, 25(3), 251-261.

Sydney Ports (2012) PBLIS Industry Update February 2012, website: http://www.sydneyports.com.au/port_development/?a=17309, last accessed in June 2012.

Takahashi, T. (2004) Spatial Competition of Governments in the Investment of Public Facilities, *Regional Science and Urban Economics*, 34, 455-488.

Talley, W. K. and Ng, M. W. (2013) Maritime Transport Chain Choice by Carriers, Ports and Shippers, *International Journal of Production Economics*, 142, 311-316.

Tirole, J. (1988) *The Theory of Industrial Organization*, MIT Press, Cambridge.

Tobin, J. (1958) Estimation of Relationships for Limited Dependent Variables, *Econometrica*, 26(1), 24-36.

Torres, E., Domınguez, J. S., Valdes, L., and Aza, R. (2005) Passenger Waiting Time in an Airport and Expenditure Carried out in the Commercial Area, *Journal of Air Transport Management*, 11, 363–367.

Turner, H., R. Windle and Dresner, M. (2004) North American Container Port Productivity: 1984-1997, *Transportation Research Part E*, 40, 339-356.

TTI (2010) *Urban Mobility Report 2010*, Texas Transportation Institute (TTI).

TTI (2012) Appendix A: Methodology – Basic Urban Mobility Report Calculation, *Urban Mobility Report 2012*, Texas Transportation Institute (TTI).

Ubbels, B. and Verhoef, E.T. (2008) Governmental Competition in Road Charging and Capacity Choice, *Regional Science and Urban Economics*, 38, 174-190.

Ugbonma, C., Ugboma, O. and Ogwude, I.C. (2006) An Analytic Hierarchy Process (AHP) Approach to Port Selection Decisions — Empirical Evidence from Nigerian Ports, *Maritime Economics and Logistics*, 8(3), 251-266.

US GAO (2003) *Freight Transportation: Strategies Needed to Address Planning and Financing Limitations*, U.S. General Accounting Office, Washington, DC.

Van De Voorde, E. and Vanelslander, T. (2009) Market Power and Vertical and Horizontal Integration in the Maritime Shipping and Port Industry, *OECD/ITF discussion paper 2009-2*.

van den Berg, V. and Verhoef, E.T. (2011) Congestion Tolling in the Bottleneck Model with Heterogeneous Values of Time, *Transportation Research Part B*, 45, 60–78.

van Klink, H.A. and van den Berg, G.C. (1998) Gateways and Intermodalism, *Journal of Transport Geography*, 6(1), 1-9.

Van Reeven, P. (2010) The Effect of Competition on Economic Rents in Seaports, *Journal of Transport Economics and Policy*, 44(1), 79-92.

Veldman, S.J., and Bückmann, E.H. (2003) A Model on Container Port Competition: An Application for the West European Container Hub-Ports, *Maritime Economics and Logistics*, 5, 3-22.

Verbeek, M. (2005) *A Guide to Modern Econometrics*, John Wiley and Sons, West Sussex, U.K.

Vlitos-Rowe, I. (1999) *The Future of Duty-free Retailing: Developments and trends in Europe and the Middle East*, Financial Times Energy, London.

Weisbrod, G. and Fitzroy, S. (2011) Chapter 16: Traffic Congestion Effects on Supply Chains: Accounting for Behavioral Elements in Planning and Economic Impact Models, in Renko, S. (Ed.): *Supply Chain Management - New Perspectives*, InTech.

Yang, H., and Zhang, A. (2011) Price-cap Regulation of Congestion Airports, *Journal of regulatory Economics*, 39 (3), 293-312.

Yuen, A., Basso, L.J. 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.

Yuen, A., Zhang, A. and Cheung, W. (2012) Port Competitiveness from the Users' Perspective: An Analysis of Major Container Ports in China and Its Neighboring Countries, *Research in Transportation Economics*, 35(1), 34–40.

Yuen, A., Zhang, A., and Cheung, W. (2013) Foreign Participation and Competition: A Way to Improve Container Port Efficiency in China? *Transportation Research Part A*, 49, 220-231.

Zhang, A. (2008) The Impact of Hinterland Access Conditions on Rivalry between Ports, OECD/ITF discussion paper 2008-8.

Zhang, A. and Czerny, A. I. (2012) Airports and Airlines Economics and Policy: An Interpretive Review of Recent Research, Economics of Transportation, 1, 15-34.

Zhang, A. and Zhang, Y. (1996) Stability of a Cournot-Nash Equilibrium: The Multiproduct Case, *Journal of Mathematical Economics*, 26(4), 441-462.

Zhang, A., and Zhang, Y., (2003) Airport Charges and Capacity Expansion: Effects of Concessions and Privatization, *Journal of Urban Economics*, 53 (1), 54–75.

Zhang, A., and Zhang, Y. (2006) Airport Capacity and Congestion When Carriers Have Market Power, *Journal of Urban Economics*, 60, 229-247.

Zhang, A., and Zhang, Y. (2010) Airport Capacity and Congestion Pricing with Both Aeronautical and Commercial Operations, *Transportation research Part B*, 44 (3), 404–413.

Zongdag, B., Bucci, P. Gutzkow, P. and De Jong, G. (2010) Port Competition Modeling Including Maritime, Port, and Hinterland Characteristics, *Maritime Policy and Management*, 37(3), 179-194.

Appendices

Appendix A: Appendix for chapter 2

A.1 Regularity condition and comparative statics for profit-maximizing ports

Under the linear case, from (2.10), we can solve for $Y_i(X_i, K_i, t_i)$. That is,

$$Y_i = \frac{K_i(M_i - t_i) - \alpha c X_i}{K_i m_i + \alpha c} \text{ and } V_i = \frac{K_i(M_i - t_i) + K_i m_i X_i}{K_i m_i + \alpha c}$$

$$H_i = \frac{\partial p_i}{\partial X_i} = \frac{\partial \rho_i}{\partial X_i} - \frac{\partial D_i}{\partial V_i}\frac{\partial V_i}{\partial X_i} = -a_i - \frac{\mu_i}{\alpha} < 0 \text{ where } \mu_i = \frac{\alpha c m_i}{K_i m_i + \alpha c} > 0$$

$$\frac{\partial p_i}{\partial X_j} = \frac{\partial \rho_i}{\partial X_j} = -b_i$$

$$\pi_i^i = H_i X_i + p_i$$

$$\pi_j^i = \frac{\partial p_i}{\partial X_j} X_i = -b_i X_i < 0$$

The negative definite Slusky matrix of demand functions in (2.1) implies:

$$a_1 a_2 - b_1 b_2 \geq 0 \tag{A.1}$$

Therefore, from (A.1), regularity conditions of Cournot competition are satisfied, because: (1) $\pi_{ii}^i = 2H_i < 0$; (2) quantities are strategic substitutes, as $\pi_{ij}^i = -b_i < 0$ and (3) stability condition is satisfied, as:

$$\Delta = |\,\pi_{11}^1 \pi_{22}^2| - |\,\pi_{12}^1 \pi_{21}^2| = \pi_{11}^1 \pi_{22}^2 - \pi_{12}^1 \pi_{21}^2 = 4H_1 H_2 - b_1 b_2 > 0$$

Let L_1 be any of road capacity (K_1), commuter toll (t_1), truck toll (T_1), or fixed-ratio road toll (τ_1) in chain 1. Differentiate both sides of $\pi_i^i = 0$ (i = 1, 2) with respect to L_1 and solve the system of equations by Cramer's rule. Then, we can get:

$$\frac{dX_1^*}{dL_1} = \frac{-2H_2}{\Delta}\frac{\partial\pi_1^1}{\partial L_1} \text{ and } \frac{dX_2^*}{dL_1} = \frac{-b_2}{\Delta}\frac{\partial\pi_1^1}{\partial L_1} \tag{A.2}$$

A.2 Proof of Proposition 2.1

The first two terms in equation (2.11) equals to:

$$\frac{\partial p_1}{\partial K_1} = -\frac{\partial D_1}{\partial K_1} - \frac{\partial D_1}{\partial V_1}\frac{\partial Y_1}{\partial K_1} = -\frac{\partial D_1}{\partial K_1}\frac{\partial V_1}{\partial X_1} = \frac{\mu_1 V_1}{\alpha K_1} > 0$$

The last two terms in equation (2.11) equals to:

$$\frac{\partial H_1}{\partial K_1}X_1 = \frac{{\mu_1}^2}{\alpha^2 c}X_1 > 0$$

Thus, $\frac{\partial\pi_1^1}{\partial K_1} = \frac{\mu_1 V_1}{\alpha K_1} + \frac{{\mu_1}^2}{\alpha^2 c}X_1 > 0$ and, according to (A.2), $\frac{dX_1^*}{dK_1} > 0$ and $\frac{dX_2^*}{dK_1} < 0$. Consequently,

$$\frac{dp_1^*}{dK_1} = \frac{\partial p_1}{\partial X_1}\frac{dX_1^*}{dK_1} + \frac{\partial p_1}{\partial X_2}\frac{dX_2^*}{dK_1} + \frac{\partial p_1}{\partial K_1} = \left(H_1 - \frac{b_1 b_2}{2H_2}\right)\frac{-2H_2}{\Delta}\left(\frac{\mu_1 V_1}{\alpha K_1} + \frac{{\mu_1}^2}{\alpha^2 c}X_1\right) + \frac{\mu_1 V_1}{\alpha K_1}$$

$$= 3b_1 b_2 \frac{{\mu_1}^2}{\alpha^2 c\Delta}X_1 + (2H_1H_2 + b_1b_2)\frac{\mu_1}{\alpha\Delta}\frac{M_i - t_i}{K_i m_i + \alpha c} > 0$$

$$\frac{dp_2^*}{dK_1} = \frac{\partial p_2}{\partial X_2}\frac{dX_2^*}{dK_1} + \frac{\partial p_2}{\partial X_1}\frac{dX_1^*}{dK_1} = -\frac{b_2}{2}\frac{dX_1^*}{dK_1} < 0$$

$$\frac{dD_1^*}{dK_1} = \frac{\partial D_1}{\partial K_1} + \frac{\partial D_1}{\partial V_1}\frac{\partial Y_1}{\partial K_1} + \frac{\partial D_1}{\partial V_1}\frac{\partial V_1}{\partial X_1}\frac{dX_1^*}{dK_1} = -\frac{\partial p_1}{\partial K_1} + \frac{\mu_1}{\alpha}\frac{dX_1^*}{dK_1}$$

$$= \left(-4a_1a_2 + b_1b_2 - 4a_1\frac{\mu_1}{\alpha}\right)\frac{{\mu_1}^2}{\alpha^2 c\Delta}X_1 + \left(-2H_1H_2 + b_1b_2 - 2a_1a_2 - 2a_1\frac{\mu_2}{\alpha}\right)\frac{\mu_1}{\alpha\Delta}\frac{M_i - t_i}{K_i m_i + \alpha c}$$

< 0 (since both brackets are negative)

$$\frac{dD_2^*}{dK_1} = \frac{\partial D_2}{\partial V_2}\frac{\partial V_2}{\partial X_2}\frac{dX_2^*}{dK_1} = \frac{\mu_2}{\alpha}\frac{dX_2^*}{dK_1} < 0$$

Let commuter surplus $S_{Yi} = \int_0^{Y_i}\rho_{Li}(y)dy - \alpha D_i Y_i - t_i Y_i$. Then,

$$\frac{dS_{Y1}^*}{dK_1} = -\alpha \frac{dD_1^*}{dK_1} Y_1 > 0 \text{ , and } \frac{dS_{Y2}^*}{dK_1} = -\alpha \frac{dD_2^*}{dK_1} Y_2 > 0.$$

A.3 Proof of Proposition 2.2

Equation (2.13) equals to $\frac{\partial p_1}{\partial t_1} = -\frac{\partial D_1}{\partial V_1}\frac{\partial Y_1}{\partial t_1} = \frac{c}{K_1 m_1 + \alpha c} > 0$

Therefore, $\frac{\partial \pi_1^1}{\partial t_1} = \frac{\partial p_1}{\partial t_1} > 0$ and hence $\frac{dX_1^*}{dt_1} > 0$ and $\frac{dX_2^*}{dt_1} < 0$.

$$\frac{d\pi^{1*}}{dt_1} = \pi_2^1 \frac{dX_2^*}{dt_1} + \frac{\partial \pi^1}{\partial t_1} = -b_1 X_1 \frac{dX_2^*}{dt_1} + \frac{\partial p_1}{\partial t_1} X_1 > 0$$

$$\frac{d\pi^{2*}}{dt_1} = \pi_1^2 \frac{dX_1^*}{dt_1} = -b_2 X_2 \frac{dX_1^*}{dt_1} < 0$$

$$\frac{dp_1^*}{dt_1} = \frac{\partial p_1}{\partial X_1}\frac{dX_1^*}{dt_1} + \frac{\partial p_1}{\partial X_2}\frac{dX_2^*}{dt_1} + \frac{\partial p_1}{\partial t_1} = \left(H_1 - \frac{b_1 b_2}{2H_2}\right)\frac{-2H_2}{\Delta}\frac{\partial p_1}{\partial t_1} + \frac{\partial p_1}{\partial t_1} = \frac{2}{\Delta}\frac{\partial p_1}{\partial t_1}(6H_1H_2 - b_1b_2) > 0$$

$$\frac{dp_2^*}{dt_1} = \frac{\partial p_2}{\partial X_2}\frac{dX_2^*}{dt_1} + \frac{\partial p_2}{\partial X_1}\frac{dX_1^*}{dt_1} = \frac{b_2 H_2}{\Delta}\frac{\partial p_1}{\partial t_1} < 0$$

$$\frac{dD_1^*}{dt_1} = \frac{\partial D_1}{\partial V_1}\frac{\partial Y_1}{\partial t_1} + \frac{\partial D_1}{\partial V_1}\frac{\partial V_1}{\partial X_1}\frac{dX_1^*}{dt_1} = -\frac{\partial p_1}{\partial t_1} + \frac{\mu_1}{\alpha}\frac{-2H_2}{\Delta}\frac{\partial p_1}{\partial t_1} = \frac{1}{\Delta}\frac{\partial p_1}{\partial t_1}(-2H_1H_2 + b_1b_2 + 2a_1H_2) < 0$$

$$\frac{dD_2^*}{dt_1} = \frac{\partial D_2}{\partial V_2}\frac{\partial V_2}{\partial X_2}\frac{dX_2^*}{dt_1} = \frac{\mu_2}{\alpha}\frac{dX_2^*}{dt_1} < 0$$

$$\frac{dS_{Y1}^*}{dt_1} = -\left(1 + \alpha\frac{dD_1^*}{dt_1}\right)Y_1 = -Y_1\frac{\mu_1}{\alpha m_1 \Delta}\left(-2\mu_1 H_2 + \frac{K_1 m_1}{c}\Delta\right) < 0, \text{ and } \frac{dS_{Y2}^*}{dt_1} = -\alpha\frac{dD_2^*}{dt_1}Y_2 > 0$$

A.4 Proof of Proposition 2.3

Since $\frac{\partial \pi_1^1}{\partial T_1} = -1$ and $\frac{\partial \pi_2^2}{\partial T_1} = 0$, we have $\frac{dX_1^*}{dT_1} < 0$ and $\frac{dX_2^*}{dT_1} > 0$.

$$\frac{d\pi^{1*}}{dT_1} = \pi_2^1 \frac{dX_2^*}{dT_1} + \frac{\partial \pi^1}{\partial T_1} = -b_1 X_1 \frac{dX_2^*}{dT_1} - X_1 < 0$$

$$\frac{d\pi^{2*}}{dT_1} = \pi_1^2 \frac{dX_1^*}{dT_1} = -b_2 X_2 \frac{dX_1^*}{dT_1} > 0$$

$$\frac{dp_1^*}{dT_1} = \frac{\partial p_1}{\partial X_1}\frac{dX_1^*}{dT_1} + \frac{\partial p_1}{\partial X_2}\frac{dX_2^*}{dT_1} + \frac{\partial p_1}{\partial T_1} = \left(H_1 - \frac{b_1 b_2}{2H_2}\right)\frac{2H_2}{\Delta} - 1 = \frac{-2H_1H_2}{\Delta} < 0$$

$$\frac{dp_2^*}{dT_1} = \frac{\partial p_2}{\partial X_2}\frac{dX_2^*}{dT_1} + \frac{\partial p_2}{\partial X_1}\frac{dX_1^*}{dT_1} = \frac{-b_2H_2}{\Delta} > 0$$

$\frac{dD_1^*}{dT_1} = \frac{\partial D_1}{\partial V_1}\frac{\partial V_1}{\partial X_1}\frac{dX_1^*}{dT_1} < 0$, and, thus, $\frac{dS_{Y1}^*}{dT_1} = -\alpha\frac{dD_1^*}{dT_1}Y_1 > 0$

$\frac{dD_2^*}{dT_1} = \frac{\partial D_2}{\partial V_2}\frac{\partial V_2}{\partial X_2}\frac{dX_2^*}{dT_1} > 0$, and, thus, $\frac{dS_{Y2}^*}{dT_1} = -\alpha\frac{dD_2^*}{dT_1}Y_2 < 0$

A.5 Proof of Proposition 2.4

Let $T_i = t_i = \tau_i$

$$\frac{\partial \pi_1^1}{\partial \tau_1} = \frac{\partial p_1}{\partial \tau_1} + \frac{\partial H_1}{\partial \tau_1}X_1 = \left(-1 - \frac{\partial D_1}{\partial V_1}\frac{\partial Y_1}{\partial t_1}\right) + 0 = \frac{-K_1m_1 - \alpha c + c}{K_1m_1 + \alpha c}\begin{cases} > 0,\ if\ \alpha < 1 - \frac{K_1m_1}{c} \\ = 0,\ if\ \alpha = 1 - \frac{K_1m_1}{c} \\ < 0,\ if\ \alpha > 1 - \frac{K_1m_1}{c} \end{cases}$$

Therefore, if and only if $\alpha < 1 - \frac{K_1m_1}{c}$, we will obtain the following results:

$\frac{dX_1^*}{d\tau_1} > 0$ and $\frac{dX_2^*}{d\tau_1} < 0$

$$\frac{d\pi^{1*}}{d\tau_1} = \pi_2^1\frac{dX_2^*}{d\tau_1} + \frac{\partial \pi^1}{\partial \tau_1} = -b_1X_1\frac{dX_2^*}{d\tau_1} + \frac{\partial p_1}{\partial \tau_1}X_1 > 0$$

$$\frac{d\pi^{2*}}{d\tau_1} = \pi_1^2\frac{dX_1^*}{d\tau_1} = -b_2X_2\frac{dX_1^*}{d\tau_1} < 0$$

$$\frac{dp_1^*}{d\tau_1} = \frac{\partial p_1}{\partial X_1}\frac{dX_1^*}{d\tau_1} + \frac{\partial p_1}{\partial X_2}\frac{dX_2^*}{d\tau_1} + \frac{\partial p_1}{\partial \tau_1} = \frac{2}{\Delta}\frac{\partial p_1}{\partial \tau_1}(6H_1H_2 - b_1b_2) > 0$$

$$\frac{dp_2^*}{d\tau_1} = \frac{\partial p_2}{\partial X_2}\frac{dX_2^*}{d\tau_1} + \frac{\partial p_2}{\partial X_1}\frac{dX_1^*}{d\tau_1} = \frac{b_2H_2}{\Delta}\frac{\partial p_1}{\partial \tau_1} < 0$$

$\frac{dD_2^*}{d\tau_1} = \frac{\partial D_2}{\partial V_2}\frac{\partial V_2}{\partial X_2}\frac{dX_2^*}{d\tau_1} = \frac{\mu_2}{\alpha}\frac{dX_2^*}{d\tau_1} < 0$, and therefore $\frac{dS_{Y2}^*}{d\tau_1} = -\alpha\frac{dD_2^*}{d\tau_1}Y_2 > 0$

However, in terms of the road congestion in chain 1, we have $\forall\alpha$

$$\frac{dD_1^*}{d\tau_1} = \frac{\partial D_1}{\partial V_1}\frac{\partial Y_1}{\partial \tau_1} + \frac{\partial D_1}{\partial V_1}\frac{\partial V_1}{\partial X_1}\frac{dX_1^*}{d\tau_1} = -\frac{\partial p_1}{\partial t_1} + \frac{\mu_1}{\alpha}\frac{-2H_2}{\Delta}\left(-1 + \frac{\partial p_1}{\partial t_1}\right)$$

$$= \frac{1}{\Delta}\frac{\partial p_1}{\partial t_1}(-2H_1H_2 + b_1b_2 + 2a_1H_2) + \frac{\mu_1}{\alpha}\frac{2H_2}{\Delta} < 0 \quad \text{(as } \frac{\partial p_1}{\partial t_1} > 0 \text{ from proof of Proposition 2.2)}$$

$$\frac{dS_{Y1}^*}{d\tau_1} = -\left(1 + \alpha\frac{dD_1^*}{d\tau_1}\right)Y_1 = -Y_1\frac{\mu_1 K_1}{\alpha c\Delta}\left(-H_2\left(3a_1 + \frac{\mu_1}{\alpha}\right) + H_1H_2 - b_1b_2\right) < 0$$

A.6 Regularity condition and comparative statics for social welfare-maximizing ports ($0 < \alpha \leq 1$)

$$W_i^i = F_i^i + \pi_i^i + T_i = \left(-t_i\frac{\mu_i}{m_i} - \mu_i Y_i\right) + H_i X_i + p_i + T_i = \left(-t_i\frac{\mu_i}{m_i} - \mu_i Y_i\right) + H_i X_i + \rho_i - D_i$$

$$W_j^i = \pi_j^i = -b_i X_i < 0$$

The regularity conditions of Cournot competition are satisfied:

$$W_{ii}^i = F_{ii}^i + \pi_{ii}^i = \frac{{\mu_i}^2}{m_i} + 2H_i = -2a_i - \gamma_i\mu_i < 0, \text{ where } \gamma_i = \frac{2}{\alpha} - \frac{\alpha c}{K_i m_i + \alpha c} > \frac{1}{\alpha} \geq 1$$

$$W_{ij}^i = \pi_{ij}^i = -b_i < 0 \text{ (strategic substitutes)}$$

$$\Omega = |\, W_{11}^1 W_{22}^2| - |\, W_{12}^1 W_{21}^2| = W_{11}^1 W_{22}^2 - W_{12}^1 W_{21}^2$$

$$= (2a_1 + \gamma_1\mu_1)(2a_2 + \gamma_2\mu_2) - b_1b_2 > 0 \text{ (stability condition)}$$

Therefore, $\frac{dX_1^*}{dL_1} = \frac{2a_2 + \gamma_2\mu_2}{\Omega}\frac{\partial W_1^1}{\partial L_1}$ and $\frac{dX_2^*}{dL_1} = \frac{-b_2}{\Omega}\frac{\partial W_1^1}{\partial L_1}$

A.7 Proof of Proposition 2.5

$$\frac{\partial W_1^1}{\partial K_1} = \left(1 + \alpha\frac{-\alpha c}{K_1 m_1 + \alpha c}\right)\frac{\partial p_1}{\partial K_1} + (\alpha Y_1 + X_1)\frac{\partial H_1}{\partial K_1} + t_1\frac{\mu_1}{K_1 m_1 + \alpha c} \qquad \text{(A.3)}$$

Since $0 < \alpha \leq 1$, the first bracket of (A.3) is positive. From Proposition 2.1, we have shown that $\frac{\partial p_1}{\partial K_1} > 0$ and $\frac{\partial H_1}{\partial K_1} > 0$. Therefore, $\frac{\partial W_1^1}{\partial K_1} > 0$ and hence $\frac{dX_1^*}{dK_1} > 0$ and $\frac{dX_2^*}{dK_1} < 0$.

Since $\frac{\partial W^1}{\partial K_1} = \frac{\partial F^1}{\partial K_1} + \frac{\partial \pi^1}{\partial K_1} = t_1 \frac{\partial Y_1}{\partial K_1} + (\alpha Y_1 + X_1)\frac{\partial p_1}{\partial K_1} > 0$, we can get:

$$\frac{dW^{2*}}{dK_1} = W_1^2 \frac{dX_1^*}{dK_1} = -b_2 X_2 \frac{dX_1^*}{dK_1} < 0$$

$$\frac{dp_1^*}{dK_1} = \frac{\partial p_1}{\partial X_1}\frac{dX_1^*}{dK_1} + \frac{\partial p_1}{\partial X_2}\frac{dX_2^*}{dK_1} + \frac{\partial p_1}{\partial K_1} = \left(-\left(a_1 + \frac{\mu_1}{\alpha}\right)(2a_2 + \gamma_2\mu_2) + b_1 b_2\right)\frac{1}{\Omega}\frac{\partial W_1^1}{\partial K_1} + \frac{\partial p_1}{\partial K_1}$$

$$= \frac{\mu_1}{\alpha\Omega}\left\{\frac{\mu_1}{\alpha c} X_1 \Theta_{\mathrm{p}} + \frac{M_1}{K_1 m_1 + \alpha c}\Psi_{\mathrm{p}} - \frac{t_1}{K_1 m_1 + \alpha c}\Phi_{\mathrm{p}}\right\}, \text{ where}$$

$$\Theta_{\mathrm{p}} = -\frac{\alpha\mu_1}{m_1} W_{22}^2 \left(2a_1 + \frac{\mu_1}{\alpha}\right) + 2\frac{K_1 m_1 + (1-\alpha)\alpha c}{K_1 m_1 + \alpha c} b_1 b_2 > 0$$

$$\Psi_{\mathrm{p}} = -W_{22}^2\left(\frac{(1-\alpha)K_1 m_1 + (1+\alpha)\alpha c}{K_1 m_1 + \alpha c} a_1 + \frac{(1-\alpha)K_1 m_1 + \alpha c}{K_1 m_1 + \alpha c}\frac{\mu_1}{\alpha}\right) + \frac{(1+\alpha)K_1 m_1 + (1-\alpha)\alpha c}{K_1 m_1 + \alpha c} b_1 b_2 > 0$$

$$\Phi_{\mathrm{p}} = -W_{22}^2\left(\frac{K_1 m_1 + (1+2\alpha)\alpha c}{K_1 m_1 + \alpha c} a_1 + \frac{K_1 m_1 + (1+\alpha)\alpha c}{K_1 m_1 + \alpha c}\frac{\mu_1}{\alpha}\right) + b_1 b_2 > 0$$

$$\Psi_{\mathrm{p}} - \Phi_{\mathrm{p}} = \alpha\left(\left(a_1 + \frac{\mu_1}{\alpha}\right)(2a_2 + \gamma_2\mu_2) + \frac{K_1 m_1 - \alpha c}{K_1 m_1 + \alpha c} b_1 b_2\right)$$

$$> \alpha\left(2b_1 b_2 + \frac{K_1 m_1 - \alpha c}{K_1 m_1 + \alpha c} b_1 b_2\right) = \alpha\frac{3K_1 m_1 + \alpha c}{K_1 m_1 + \alpha c} b_1 b_2 > 0$$

Since positive Y_i requires $M_i > t_i$, then $\frac{M_1}{K_1 m_1 + \alpha c}\Psi_{\mathrm{p}} > \frac{t_1}{K_1 m_1 + \alpha c}\Phi_{\mathrm{p}}$. Therefore, $\frac{dp^{1*}}{dK_1} > 0$.

$$\frac{dp_2^*}{dK_1} = \frac{\partial p_2}{\partial X_2}\frac{dX_2^*}{dK_1} + \frac{\partial p_2}{\partial X_1}\frac{dX_1^*}{dK_1} = \left(a_2 + \frac{K_2 m_2 + (1-\alpha)\alpha c}{K_2 m_2 + \alpha c}\frac{\mu_2}{\alpha}\right)\frac{dX_2^*}{dK_1} < 0$$

$$\frac{dD_1^*}{dK_1} = -\frac{\partial p_1}{\partial K_1} + \frac{\mu_1}{\alpha}\frac{dX_1^*}{dK_1} = \frac{\mu_1}{\alpha}\frac{2a_2 + \gamma_2\mu_2}{\Omega}\frac{\partial W_1^1}{\partial K_1} - \frac{\partial p_1}{\partial K_1}$$

$$= (-a_1(2a_2 + \gamma_2\mu_2) + b_1 b_2)\frac{1}{\Omega}\frac{\partial W_1^1}{\partial K_1} - \frac{dp_1^*}{dK_1} < 0$$

$$\frac{dS_{Y1}^*}{dK_1} = -\alpha \frac{dD_1^*}{dK_1} Y_1 > 0$$

$$\frac{dD_2^*}{dK_1} = \frac{\partial D_2}{\partial V_2}\frac{\partial V_2}{\partial X_2}\frac{dX_2^*}{dK_1} = \frac{\mu_2}{\alpha}\frac{dX_2^*}{dK_1} < 0$$

$$\frac{dS_{Y2}^*}{dK_2} = -\alpha \frac{dD_2^*}{dK_2} Y_2 > 0$$

If the port charges are positive, or equivalently the road tolls (especially, the truck tolls) are low (e.g., $\tau_i < MEC_i - \frac{\partial \rho_i/\partial X_i}{\partial V_i/\partial X_i} X_i \ or\ T_i < MEC_i \frac{\partial V_i}{\partial X_i} - \frac{\partial \rho_i}{\partial X_i} X_i - t_i \frac{\partial Y_i}{\partial X_i}$), we have:

$$\frac{d\pi^{1*}}{dK_1} = X_1^* \frac{dp_1^*}{dK_1} + p_1^* \frac{dX_1^*}{dK_1} > 0$$

$$\frac{d\pi^{2*}}{dK_1} = X_2^* \frac{dp_2^*}{dK_1} + p_2^* \frac{dX_2^*}{dK_1} < 0$$

A.8 Proof of Proposition 2.6

$$\frac{\partial W_1^1}{\partial \tau_1} = \frac{\partial W_1^1}{\partial t_1} = -\frac{\mu_1}{m_1} - \mu_1 \frac{\partial Y_1}{\partial t_1} - \frac{\partial D_1}{\partial V_1}\frac{\partial Y_1}{\partial t_1} = \frac{c(K_1 m_1 + (1-\alpha)\alpha c)}{(K_1 m_1 + \alpha c)^2} > 0 \quad \text{(since } 0 < \alpha \leq 1\text{)}$$

$$\frac{\partial W_1^1}{\partial T_1} = 0$$

Therefore, $\frac{dX_1^*}{d\tau_1} = \frac{dX_1^*}{dt_1} > 0$, $\frac{dX_2^*}{d\tau_1} = \frac{dX_2^*}{dt_1} < 0$, and $\frac{dX_1^*}{dT_1} = \frac{dX_2^*}{dT_1} = 0$.

$$\frac{dp_1^*}{d\tau_1} = \frac{1}{\Omega}\frac{\partial W_1^1}{\partial \tau_1}(H_1(2a_2 + \gamma_2\mu_2) + b_1 b_2) + \frac{\partial p_1}{\partial \tau_1}\begin{cases} < 0,\ if\ \alpha \geq 1 - \frac{K_1 m_1}{c} \\ undetermined,\ if\ \alpha < 1 - \frac{K_1 m_1}{c} \end{cases}$$

$$\frac{dp_1^*}{dt_1} = \frac{1}{\Omega}\frac{\partial W_1^1}{\partial t_1}(H_1(2a_2 + \gamma_2\mu_2) + b_1 b_2) + \frac{\partial p_1}{\partial t_1}$$

$$= \frac{1}{\Omega}\left(\frac{c}{K_1 m_1 + \alpha c}\right)\left[\left(1 - \frac{\alpha^2 c}{K_1 m_1 + \alpha c}\right)(H_1(2a_2 + \gamma_2\mu_2) + b_1 b_2) + \Omega\right]$$

Because $0 < -H_1(2a_2 + \gamma_2\mu_2) - b_1 b_2 < (2a_1 + \gamma_1\mu_1)(2a_2 + \gamma_2\mu_2) - b_1 b_2 = \Omega$ and

$0 < 1 - \frac{\alpha^2 c}{K_1 m_1 + \alpha c} < 1$, the square bracket is positive and $\frac{dp_1^*}{dt_1} > 0$.

$$\frac{dp_2^*}{d\tau_1} = \frac{dp_2^*}{dt_1} = -\frac{b_2}{\Omega}\frac{\partial W_1^1}{\partial t_1}\left((2a_2 + \gamma_2\mu_2) + H_2\right) < 0$$

Therefore, similar to Proposition 2.5, if the port charges are positive, or equivalently the road tolls (especially, the truck tolls) are low, we have:

$$\frac{d\pi^{1*}}{dt_1} = X_1^* \frac{dp_1^*}{dt_1} + p_1^* \frac{dX_1^*}{dt_1} > 0 \text{ , } \frac{d\pi^{2*}}{dt_1} = X_2^* \frac{dp_2^*}{dt_1} + p_2^* \frac{dX_2^*}{dt_1} < 0 \text{ , and } \frac{d\pi^{2*}}{d\tau_1} = X_2^* \frac{dp_2^*}{d\tau_1} + p_2^* \frac{dX_2^*}{d\tau_1} < 0.$$

$$\frac{dW^{2*}}{d\tau_1} = \frac{dW^{2*}}{dt_1} = -b_2 X_2 \frac{dX_1^*}{dt_1} < 0$$

$$\frac{dD_1^*}{d\tau_1} = \frac{dD_1^*}{dt_1} = -\frac{\partial p_1}{\partial t_1} + \frac{\mu_1}{\alpha}\frac{2a_2+\gamma_2\mu_2}{\Omega}\left(-\frac{(\alpha c)^2}{(K_1 m_1+\alpha c)^2} + \frac{\partial p_1}{\partial t_1}\right)$$

$$= \frac{1}{\Omega}\frac{\partial p_1}{\partial t_1}\left(-(2a_2 + \gamma_2\mu_2)\left(2a_1 + \left(\gamma_1 - \frac{1}{\alpha}\right)\mu_1\right) + b_1 b_2\right) - \frac{(\alpha c)^2}{(K_1 m_1+\alpha c)^2}\frac{\mu_1}{\alpha}\frac{2a_2+\gamma_2\mu_2}{\Omega}$$

< 0 (Since $\gamma_1 - \frac{1}{\alpha} > 0$)

$$\frac{dS_{Y1}^*}{d\tau_1} = \frac{dS_{Y1}^*}{dt_1} = -\left(1 + \alpha \frac{dD_1^*}{dt_1}\right) Y_1 = -Y_1 \frac{\mu_1}{m_1 \Omega}\left((2a_2 + \gamma_2\mu_2)\frac{\mu_1}{\alpha}\left(1 - \frac{\alpha^2 c}{K_1 m_1+\alpha c}\right) + \frac{K_1 m_1}{\alpha c}\Omega\right) < 0$$

$$\frac{dD_2^*}{d\tau_1} = \frac{dD_2^*}{dt_1} = \frac{\partial D_2}{\partial V_2}\frac{\partial V_2}{\partial X_2}\frac{dX_2^*}{dt_1} < 0 \text{ , and therefore } \frac{dS_{Y2}^*}{d\tau_1} = \frac{dS_{Y2}^*}{dt_1} = -\alpha \frac{dD_2^*}{dt_1} Y_2 > 0$$

$$\frac{dD_1^*}{dT_1} = \frac{\partial D_1}{\partial V_1}\frac{\partial V_1}{\partial X_1}\frac{dX_1^*}{dT_1} = 0\text{, and, thus, } \frac{dS_{Y1}^*}{dT_1} = -\alpha \frac{dD_1^*}{dT_1} Y_1 = 0$$

$$\frac{dD_2^*}{dT_1} = \frac{\partial D_2}{\partial V_2}\frac{\partial V_2}{\partial X_2}\frac{dX_2^*}{dT_1} = 0\text{, and, thus, } \frac{dS_{Y2}^*}{dT_1} = -\alpha \frac{dD_2^*}{dT_1} Y_2 = 0$$

Appendix B: Appendix for chapter 4

Proof of the statement: for public ports, at port stage equilibrium, Q_{BI} and Q_{NI} are both positive for any k_I , k_B and k_N > 0.

Proof:

$Q_{BI} = (d/2) + k_I(p_N - p_B) > 0$ holds if and only if

$$p_B - p_N < d/2k_I \ . \qquad \text{(B.1)}$$

Since at equilibrium $|p^B - p^N| = (d/2)(|k_N - k_B|/\Delta_W) < (d/2)(1/k_I)$, (B.1) holds for equilibrium port charges.

Appendix C: Appendix for chapter 5

C.1 Proof of Proposition 5.3

(1) Let $\varDelta_1$ be the difference between the first terms of the second line of equations (5.15) and (5.21).

$$\varDelta_1 = (p_c^{\Pi} - c_c) \sum_{h \in \{B,L\}} Q_h \frac{\partial \bar{G}_h\big(p_c^{\Pi}; D(Q)\big)}{\partial D} - \delta \sum_{h \in \{B,L\}} Q_h \int_{c_c}^{\bar{u}} \frac{\partial \bar{G}_h\big(z; D(Q)\big)}{\partial D} dz$$

If $\delta = 1$ we have

$$\varDelta_1 = \sum_{h \in \{B,L\}} Q_h \int_{c_c}^{p_c^{\Pi}} \frac{\partial \bar{G}_h\big(p_c^{\Pi}; D(Q)\big)}{\partial D} dz - \sum_{h \in \{B,L\}} Q_h \int_{p_c^{\Pi}}^{\bar{u}} \frac{\partial \bar{G}_h\big(z; D(Q)\big)}{\partial D} dz$$

$$- \sum_{h \in \{B,L\}} Q_h \int_{c_c}^{p_c^{\Pi}} \frac{\partial \bar{G}_h\big(z; D(Q)\big)}{\partial D} dz$$

$$< \sum_{h \in \{B,L\}} Q_h \int_{c_c}^{p_c^{\Pi}} \frac{\partial \bar{G}_h\big(p_c^{\Pi}; D(Q)\big)}{\partial D} dz - \sum_{h \in \{B,L\}} Q_h \int_{c_c}^{p_c^{\Pi}} \frac{\partial \bar{G}_h\big(z; D(Q)\big)}{\partial D} dz$$

Given $\partial^2 \bar{G}_h(p_c; D)/\partial p_c \partial D < 0$, we have that

$$\sum_{h \in \{B,L\}} Q_h \int_{c_c}^{p_c^{\Pi}} \frac{\partial \bar{G}_h\big(p_c^{\Pi}; D(Q)\big)}{\partial D} dz < \sum_{h \in \{B,L\}} Q_h \int_{c_c}^{p_c^{\Pi}} \frac{\partial \bar{G}_h\big(z; D(Q)\big)}{\partial D} dz$$

Therefore, $\varDelta_1 < 0$.

If $\delta = 0$ we have

$$\varDelta_1 = (p_c^{\Pi} - c_c) \sum_{h \in \{B,L\}} Q_h \frac{\partial \bar{G}_h\big(p_c^{\Pi}; D(Q)\big)}{\partial D} > 0$$

Therefore, since $\varDelta_1$ is linear in δ, there must exist some $\bar{\delta} \in (0,1)$ such that $\varDelta_1 = 0$.

(2) Let $\varDelta_2$ be the difference between the second terms of the second line of equations (5.15) and (5.21).

$$\varDelta_2 = (p_c^{\Pi} - c_c) \sum_{h \in \{B,L\}} \frac{dQ_h}{dp} \bar{G}_h\big(p_c^{\Pi}; D(Q)\big) - \delta \sum_{h \in \{B,L\}} \frac{dQ_h}{dp} \int_{c_c}^{\bar{u}} (z - c_c)\, g_h(z; D) dz$$

If $\delta = 1$ we have

$$\Delta_2 = \sum_{h\in\{B,L\}} \frac{dQ_h}{dp} \int_{p_c^\Pi}^{\bar{u}} (p_c^\Pi - c_c)\, g_h(z;D)dz - \sum_{h\in\{B,L\}} \frac{dQ_h}{dp} \int_{c_c}^{\bar{u}} (z - c_c)\, g_h(z;D)dz$$

Let $E_h^\Pi = \int_{p_c^\Pi}^{\bar{u}} (p_c^\Pi - c_c)\, g_h(z;D)dz$ and $E_h^W = \int_{c_c}^{\bar{u}} (z - c_c)\, g_h(z;D)dz$. Then

$$\Delta_2 = \frac{dQ_B}{dp}(E_B^\Pi - E_B^W) + \frac{dQ_L}{dp}(E_L^\Pi - E_L^W)$$

Since

$$E_h^W = \int_{p_c^\Pi}^{\bar{u}} (z - c_c)\, g_h(z;D)dz + \int_{c_c}^{p_c^\Pi} (z - c_c)\, g_h(z;D)dz > \int_{p_c^\Pi}^{\bar{u}} (z - c_c)\, g_h(z;D)dz > E_h^\Pi$$

when $dQ_B/dp < 0$, that is, $(v_B - v_L)\theta < b_L$, we have $\Delta_2 > 0$. When $dQ_B/dp > 0$, $\Delta_2 > 0$ if and only if

$$-\frac{\frac{dQ_B}{dp}}{\frac{dQ_L}{dp}} < \frac{E_L^W - E_L^\Pi}{E_B^W - E_B^\Pi}$$

If $\delta = 0$ we have

$$\Delta_2 = \frac{dQ_B}{dp} E_B^\Pi + \frac{dQ_L}{dp} E_L^\Pi$$

when $dQ_B/dp < 0$, we have $\Delta_2 < 0$. When $dQ_B/dp > 0$, $\Delta_2 < 0$ if and only if

$$-\frac{\frac{dQ_B}{dp}}{\frac{dQ_L}{dp}} < \frac{E_L^\Pi}{E_B^\Pi}$$

Therefore, when $dQ_B/dp < 0$, since Δ_2 is linear in δ, there must exist some $\hat{\delta} \in (0,1)$ such that $\Delta_2 = 0$.

Appendix D: Appendix for chapter 6

D.1 Proof of Lemma 6.1

Under laissez-fair situation, differentiate both sides of (6.6) with respect to p_B and use the Cramer's rule. We obtain

$$\psi = \begin{vmatrix} \rho_B^{'} - v_B D' - c_B^B & -v_B D' - c_L^B \\ -v_L D' - c_B^L & \rho_L^{'} - v_L D' - c_L^L \end{vmatrix}$$

$$= \rho_B^{'}\rho_L^{'} - \rho_B^{'}\left(v_L D' + c_L^L\right) - \rho_L^{'}\left(v_B D' + c_B^B\right) + v_L D'\left(c_B^B - c_L^B\right) + v_B D'\left(c_L^L - c_B^L\right) + c_B^B c_L^L - c_L^B c_B^L > 0$$

$$\frac{\partial Q_B}{\partial p_B} = \frac{1}{\psi}\left(\rho_L^{'} - v_L D' - c_L^L\right) < 0 \text{ and } \frac{\partial Q_B}{\partial p_L} = \frac{1}{\psi}\left(v_L D' + c_B^L\right) > 0.$$

Similarly, $\frac{\partial Q_L}{\partial p_L} = \frac{1}{\psi}\left(\rho_B^{'} - v_B D' - c_B^B\right) < 0$ and $\frac{\partial Q_L}{\partial p_B} = \frac{1}{\psi}\left(v_B D' + c_L^B\right) > 0$.

Under uniform pricing, $p_B = p_L = p$. Differentiate both sides of (6.6) with respect to p and use the Cramer's rule. We obtain:

$$\frac{\partial Q_L}{\partial p} = \frac{\partial Q_L}{\partial p_B} + \frac{\partial Q_L}{\partial p_L} = \frac{1}{\psi}\left(\rho_B^{'} - (v_B - v_L)D' - (c_B^B - c_B^L)\right),$$

$$\frac{\partial Q_B}{\partial p} = \frac{\partial Q_B}{\partial p_B} + \frac{\partial Q_B}{\partial p_L} = \frac{1}{\psi}\left(\rho_L^{'} + (v_B - v_L)D' - (c_L^L - c_L^B)\right).$$

Note that

$$c_B^B - c_B^L = \begin{cases} \left(\beta_B v_B - \beta_L v_L\right)/s > 0 & \text{if } \beta_B > \beta_L \\ \left(v_B - v_L\right)\beta_B/s > 0 & \text{if } \beta_B < \beta_L \end{cases} \text{ and } c_L^L - c_L^B = \begin{cases} -\left(v_B - v_L\right)\beta_L/s < 0 & \text{if } \beta_B > \beta_L \\ -\left(\beta_B v_B - \beta_L v_L\right)/s & \text{if } \beta_B < \beta_L \end{cases}..$$

The sign of $\beta_B v_B - \beta_L v_L$ is ambiguous when $\beta_B < \beta_L$. Thus, $\frac{\partial Q_L}{\partial p} < 0$ and the sign of $\frac{\partial Q_B}{\partial p}$ is unclear.

$$\frac{\partial Q}{\partial p} = \frac{\partial Q_B}{\partial p} + \frac{\partial Q_L}{\partial p} = \frac{1}{\psi}\left(\rho_B^{'} + \rho_L^{'} - (c_B^B - c_L^B) - (c_L^L - c_B^L)\right) < 0 .$$

D.2 Proof of Lemma 6.2

Hessian matrix of π^i is negative definite is equivalent to $\pi_{hihi}^i < 0$ and $\pi_{BiBi}^i \pi_{LiLi}^i - \pi_{BiLi}^i \pi_{LiBi}^i > 0$.

(i) The sufficient condition of local stability

Following Zhang and Zhang (1996), we derive the sufficient condition for local stability. From the first order condition (6.15), we have for each airline

$$f_i = \begin{bmatrix} \pi_{Bi}^i \\ \pi_{Li}^i \end{bmatrix} = 0 .$$

To derive the best response functions at the Nash equilibrium, let $q_i = \begin{bmatrix} q_B^i \\ q_L^i \end{bmatrix}$ be the vector of equilibrium quantities of airline $i = 1,\dots,n$. Then, we have

$$\frac{\partial f_i}{\partial q_i} = \begin{bmatrix} \pi_{BiBi}^i & \pi_{BiLi}^i \\ \pi_{LiBi}^i & \pi_{LiLi}^i \end{bmatrix},$$

and due to symmetry, for any $j,k \neq i$,

$$\frac{\partial f_i}{\partial q_j} = \frac{\partial f_i}{\partial q_k} = \begin{bmatrix} \pi_{BiBj}^i & \pi_{BiLj}^i \\ \pi_{LiBj}^i & \pi_{LiLj}^i \end{bmatrix}.$$

Then,

$$(Df_i)(q_{-i};q_i)=\left[\underbrace{\frac{\partial f_i}{\partial q_1}\ \cdots\ \frac{\partial f_i}{\partial q_j}}_{\substack{(n-1) identical 2x2 matrix\\ j\neq i}}\quad \frac{\partial f_i}{\partial q_i}\right]=\left[\frac{\partial f_i}{\partial q_{-i}}\quad \frac{\partial f_i}{\partial q_i}\right].$$

Denote $q_i^R=\begin{bmatrix} q_B^{iR}\\ q_L^{iR}\end{bmatrix}$ as the vector of best response functions for airline *i*. Then,

$$\frac{\partial q_i^R}{\partial q_{-i}}=\left[\underbrace{\frac{\partial q_i^R}{\partial q_1}\ \cdots\ \frac{\partial q_i^R}{\partial q_j}}_{\substack{(n-1) identical 2x2 matrix\\ j\neq i}}\right]=-\left[\frac{\partial f_i}{\partial q_i}\right]^{-1}\left[\frac{\partial f_i}{\partial q_{-i}}\right].$$

For a certain rival indexed by $j\neq i$, its impact on airline *i*'s best response outputs is

$$\frac{\partial q_i^R}{\partial q_j}=\begin{bmatrix}\frac{\partial q_B^{iR}}{\partial q_B^j} & \frac{\partial q_B^{iR}}{\partial q_L^j}\\ \frac{\partial q_L^{iR}}{\partial q_B^j} & \frac{\partial q_L^{iR}}{\partial q_L^j}\end{bmatrix}=-\begin{bmatrix}\pi^i_{BiBi} & \pi^i_{BiLi}\\ \pi^i_{LiBi} & \pi^i_{LiLi}\end{bmatrix}^{-1}\begin{bmatrix}\pi^i_{BiBj} & \pi^i_{BiLj}\\ \pi^i_{LiBj} & \pi^i_{LiLj}\end{bmatrix},$$

$$=-\frac{1}{A}\begin{bmatrix}\pi^i_{LiLi} & -\pi^i_{BiLi}\\ -\pi^i_{LiBi} & \pi^i_{BiBi}\end{bmatrix}\cdot\begin{bmatrix}\pi^i_{BiBj} & \pi^i_{BiLj}\\ \pi^i_{LiBj} & \pi^i_{LiLj}\end{bmatrix}=\begin{bmatrix}\frac{-a}{A} & \frac{-b}{A}\\ \frac{-c}{A} & \frac{-d}{A}\end{bmatrix}$$

where

$A=\pi^i_{BiBi}\pi^i_{LiLi}-\pi^i_{BiLi}\pi^i_{LiBi}>0$ (due to negative definite Hessian of π^i)

$a=\pi^i_{BiBj}\pi^i_{LiLi}-\pi^i_{BiLi}\pi^i_{LiBj}$

$b=\pi^i_{BiLj}\pi^i_{LiLi}-\pi^i_{BiLi}\pi^i_{LiLj}$

$c=\pi^i_{BiBi}\pi^i_{LiBj}-\pi^i_{LiBi}\pi^i_{BiBj}$

$$d = \pi^i_{BiBi}\pi^i_{LiLj} - \pi^i_{LiBi}\pi^i_{BiLj}$$

Similar to Zhang and Zhang (1996), we denote the $2n \times 2n$ best response matrix as

$$T^{'} = \begin{bmatrix} 0 & \frac{\partial q_i^R}{\partial q_j} & \cdots & \frac{\partial q_i^R}{\partial q_j} \\ \frac{\partial q_i^R}{\partial q_j} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{\partial q_i^R}{\partial q_j} \\ \frac{\partial q_i^R}{\partial q_j} & \cdots & \frac{\partial q_i^R}{\partial q_j} & 0 \end{bmatrix} = \begin{bmatrix} \frac{\partial q_i^R}{\partial q_j} & 0 & \cdots & 0 \\ 0 & \frac{\partial q_i^R}{\partial q_j} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{\partial q_i^R}{\partial q_j} \end{bmatrix} \cdot \begin{bmatrix} 0 & I & \cdots & I \\ I & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & I \\ I & \cdots & I & 0 \end{bmatrix} = R \cdot E$$

Note that $\left\|T^{'}\right\|_p \le \left\|R\right\|_p \cdot \left\|E\right\|_p = (n-1) \cdot \left\|R\right\|_p$. Thus, $\left\|R\right\|_p < \frac{1}{n-1}$ implies $\left\|T^{'}\right\|_p < 1$, i.e. the local stability condition. Since that R is a diagonal matrix, then $\left\|R\right\|_p = \max_{1 \le k \le n} \left\|R_{k,k}\right\|_p = \left\|\frac{\partial q_i^R}{\partial q_j}\right\|_p$.

Therefore, we get the sufficient condition for local stability: $\exists p \;\; s.t. \left\|\frac{\partial q_i^R}{\partial q_j}\right\|_p < \frac{1}{n-1}$. This sufficient condition is equivalent to $\max|\lambda| < \frac{1}{n-1}$, in which λ is the eigenvalue of $\frac{\partial q_i^R}{\partial q_j}$.

(ii) This sufficient condition leads to $\Delta > 0$

Since $\det\left(\frac{\partial q_i^R}{\partial q_j} - \lambda I\right) = \begin{vmatrix} \frac{-a}{A} - \lambda & \frac{-b}{A} \\ \frac{-c}{A} & \frac{-d}{A} - \lambda \end{vmatrix} = 0$, we have $A^2\lambda^2 + A(a+d)\lambda + ad - bc = 0$.

Solving this function, we get the two eigenvalues: $\lambda = \frac{-(a+d) \pm \sqrt{(a-d)^2 + 4bc}}{2A}$. Thus, the sufficient condition is equivalent to

$$\max|\lambda| = \frac{|a+d| + \sqrt{(a-d)^2 + 4bc}}{2A} < \frac{1}{n-1}. \tag{D.1}$$

After some tedious algebra, we can show that

$$(a+d)^2 - \left((a-d)^2 + 4bc\right) = 4A(\pi^i_{BiBj}\pi^i_{LiLj} - \pi^i_{BiLj}\pi^i_{LiBj}).$$

Note that

$$\pi^i_{BiBj} - \pi^i_{BiLj} = \rho''_B \frac{Q_B}{n} + \rho'_B - (c^B_B - c^B_L) < 0$$

$$\pi^i_{LiLj} - \pi^i_{LiBj} = \rho''_L \frac{Q_L}{n} + \rho'_L - (c^L_L - c^L_B) < 0$$

Therefore, $\pi^i_{BiBj} < \pi^i_{BiLj} < 0$ and $\pi^i_{LiLj} < \pi^i_{LiBj} < 0$, which implies that $\pi^i_{BiBj}\pi^i_{LiLj} - \pi^i_{BiLj}\pi^i_{LiBj} > 0$. Then, we can rewrite

$$\pi^i_{BiBj}\pi^i_{LiLj} - \pi^i_{BiLj}\pi^i_{LiBj} = \frac{(a+d)^2 - \left((a-d)^2 + 4bc\right)}{4A} > 0.$$

We can also show that

$$\Delta = A + (n-1)(a+d) + (n-1)^2(\pi^i_{BiBj}\pi^i_{LiLj} - \pi^i_{BiLj}\pi^i_{LiBj})$$

If $a+d \geq 0$, it is obvious that $\Delta > 0$.

If $a+d < 0$, we have

$$
\begin{aligned}
\Delta &= A-(n-1)\,|\,a+d\,|+(n-1)^2(\pi^i_{BiBj}\pi^i_{LiLj}-\pi^i_{BiLj}\pi^i_{LiBj})\\
&= A-(n-1)\,|\,a+d\,|+(n-1)^2\frac{(a+d)^2-\left((a-d)^2+4bc\right)}{4A}\\
&= \frac{1}{A}\left(A^2-(n-1)A\,|\,a+d\,|+\frac{(n-1)^2\,|\,a+d\,|^2}{4}-\frac{(n-1)^2\left((a-d)^2+4bc\right)}{4}\right)\\
&= \frac{1}{A}\left(\left(A-\frac{(n-1)\,|\,a+d\,|}{2}\right)^2-\frac{(n-1)^2\left((a-d)^2+4bc\right)}{4}\right)\\
&= \frac{1}{A}\left(A-\frac{(n-1)\,|\,a+d\,|}{2}+\frac{(n-1)\sqrt{(a-d)^2+4bc}}{2}\right)\left(A-\frac{(n-1)\,|\,a+d\,|}{2}-\frac{(n-1)\sqrt{(a-d)^2+4bc}}{2}\right)
\end{aligned}
\quad .
$$

Inequality (D.1) implies that both brackets in the above expression are positive, which means that $\Delta > 0$.

D.3 Proof of Proposition 6.3 (ii) and 6.3 (iii)

Note that $\pi^i_{hihj}-\pi^i_{hikj}=\rho^{''}_h\frac{Q_h}{n}+\rho^{'}_h-(c^h_h-c^h_k)<0$.

(ii) Since $\pi^i_{BiBi}-\pi^i_{BiLi}=\rho^{''}_B\frac{Q_B}{n}+2\rho^{'}_B-(v_B-v_L)D^{'}-(c^B_B-c^B_L)-(c^B_B-c^L_B)<0$, we have

$$
\frac{\partial Q^N}{\partial\tau_L}=\frac{\partial Q^N_B}{\partial\tau_L}+\frac{\partial Q^N_L}{\partial\tau_L}=\frac{n}{\Delta}\left(\pi^i_{BiBi}-\pi^i_{BiLi}+(n-1)(\pi^i_{BiBj}-\pi^i_{BiLj})\right)<0.
$$

(iii) Since $\pi^i_{LiLi}-\pi^i_{LiBi}=\rho^{''}_L\frac{Q_L}{n}+2\rho^{'}_L+(v_B-v_L)D^{'}-(c^L_L-c^L_B)-(c^L_L-c^B_L)$, then we obtain

$$
\begin{aligned}
\frac{\partial Q^N}{\partial\tau_B}&=\frac{\partial Q^N_B}{\partial\tau_B}+\frac{\partial Q^N_L}{\partial\tau_B}=\frac{n}{\Delta}\left(\pi^i_{LiLi}-\pi^i_{LiBi}+(n-1)(\pi^i_{LiLj}-\pi^i_{LiBj})\right)\\
&=\frac{n}{\Delta}\left(\rho^{''}_L Q_L+n\rho^{'}_L-n(c^L_L-c^L_B)+\rho^{'}_L+(v_B-v_L)D^{'}-(c^L_L-c^B_L)\right)\\
&=\frac{n}{\Delta}\left(\rho^{''}_L Q_L+n\rho^{'}_L-n(c^L_L-c^L_B)+\psi\frac{\partial Q_B}{\partial p}\right)
\end{aligned}
$$

As shown in Lemma 6.1, the sign of $\frac{\partial Q_B}{\partial p}$ is unknown and so is the sign of $\frac{\partial Q^N}{\partial\tau_B}$.

D.4 Proof of Proposition 6.4

Since $\tau_B = \tau_L = \tau$, differentiating both sides of (6.15) with respect to τ, imposing symmetry and using Cramer's rule, we have

$$\frac{\partial q_L^{iN}}{\partial \tau} = \frac{1}{\Delta}\begin{vmatrix} \pi^i_{BiBi} + (n-1)\pi^i_{BiBj} & 1 \\ \pi^i_{LiBi} + (n-1)\pi^i_{LiBj} & 1 \end{vmatrix} = \frac{1}{\Delta}\left(\pi^i_{BiBi} - \pi^i_{LiBi} + (n-1)(\pi^i_{BiBj} - \pi^i_{LiBj})\right),$$

where $\pi^i_{BiBj} - \pi^i_{LiBj} = \rho_B^{''}\frac{Q_B}{n} + \rho_B^{'} - (v_B - v_L)D^{'} - (c_B^B - c_B^L) < 0$. Thus, $\frac{\partial Q_L^N}{\partial \tau} = n\frac{\partial q_L^{iN}}{\partial \tau} < 0$.

$$\frac{\partial q_B^{iN}}{\partial \tau} = \frac{1}{\Delta}\begin{vmatrix} 1 & \pi^i_{BiLi} + (n-1)\pi^i_{BiLj} \\ 1 & \pi^i_{LiLi} + (n-1)\pi^i_{LiLj} \end{vmatrix} = \frac{1}{\Delta}\left(\pi^i_{LiLi} - \pi^i_{BiLi} + (n-1)(\pi^i_{LiLj} - \pi^i_{BiLj})\right)$$
$$= \frac{1}{\Delta}\left(\rho_L^{''}Q_L + n\psi\frac{\partial Q_B}{\partial p} + \psi\frac{\partial Q}{\partial p_B}\right)$$

From Lemma 6.1, we know that the sign of $\frac{\partial Q_B}{\partial p}$ is ambiguous and hence the sign of $\frac{\partial Q_B^N}{\partial \tau} = n\frac{\partial q_B^{iN}}{\partial \tau}$ is ambiguous as well.

$$\frac{\partial Q^N}{\partial \tau} = \frac{\partial Q_B^N}{\partial \tau} + \frac{\partial Q_L^N}{\partial \tau} = \frac{1}{\Delta}\left(\rho_B^{''}Q_B + \rho_L^{''}Q_L + (n+1)\left(\rho_B^{'} + \rho_L^{'} - (c_B^B - c_L^B) - (c_L^L - c_B^L)\right)\right) < 0.$$