THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITE DE TOULOUSE

Délivré par l'Université Toulouse Capitole

École doctorale : **Sciences Economiques-Toulouse School of Economics**

Présentée et soutenue par

WANG Chunan

le 23 Juin 2017

Essays on Delay Reduction Contract, Airline Networks and Agricultural Land Marketization

---

Discipline : **Sciences Economiques**

Unité de recherche : **TSE-R (UMR CNRS 5314 – INRA 1415)**

Directeur de thèse : Monsieur Doh-Shin JEON, Professeur, Université Toulouse 1 Capitole

Co-directeur de thèse: Madame Estelle MALAVOLTI, Enseignant-Chercheur, Université Toulouse 1 Capitole et Ecole Nationale de l'Aviation Civile

JURY

Rapporteurs Monsieur Benny MANTIN, Professeur, University of Luxembourg
Monsieur Frédéric MARTY, CNRS Research Fellow, Université Côte d'Azur

Suffragants Madame Estelle MALAVOLTI, Enseignant-Chercheur, UT1 Capitole et ENAC
Madame Carole HARITCHABALET, Professeur, Université de Pau et des Pays de l'Adour

Abstract

This thesis consists of three self-contained papers, each of which corresponds to one chapter.

In the context of SESAR (Single European Sky ATM Research) project, the Air Navigation Service Provider (ANSP) can provide a delay reduction service to airlines. Thus, the first chapter, jointly written with Estelle Malavolti, studies the optimal design of delay reduction contract signed between an ANSP and a monopoly airline. In the contract design, we mainly consider the adverse selection problem, which comes from airlines' private information about their values of time. Then, we derive optimal contracts analytically considering both the welfare-maximizing and profit-maximizing ANSP. We find that, under incomplete information, the optimal degree of the delay reduction service for the airline with a low value of time may be distorted downwards to decrease the information rent of the airline with a high value of time. Moreover, because contracts should be adjusted over time according to the evolution of some relevant exogenous variables, we conduct comparative-static analysis to study the effects of safety standard and flight frequency on optimal contracts. Besides, we use numerical examples to study when a welfare-maximizing ANSP has to use public funds to provide the service.

The second chapter investigates the dual roles of congestion delays and horizontal product differentiation in airline network choice. A particular feature of this work is the incorporation of all possible network structures in a three-city network, including a hub-and-spoke network (HS), point-to-point network (PP), mixed network (MX), 2-hub network (2H), with a 3-hub network (3H) as an extension. More importantly, besides contributing to the limited amount of literature on the effects of congestion delays, for the

function of 2H, I focus on the exploitation of horizontal product differentiation rather than hub airport congestion reduction. In reality, the horizontal product differentiation arises as a result of different flight departure time slots and passengers' brand loyalty. I find that, first, because of the inclusion of congestion delays, the airline may choose PP even when the extra travel time disutility of one-stop services is relatively low. Second, without considering the airline's fixed investments of developing a hub airport, 2H will dominate the three other network structures as long as it is feasible, as it involves the horizontal product differentiation in more markets than the three other network structures. Third, comparative statics show that, under MX, for example, when the marginal congestion delay cost increases, the change of flight frequency between two spoke airports depends on the trade-off between the direct negative effect of a higher delay cost and the strategic redistribution of traffic among different routes. Finally, welfare analysis shows the airline's inefficient biases towards PP and 2H.

The third chapter, jointly written with Wanjun Yao and Shigeyuki Hamori, proposes that the agricultural land marketization affects the average output per unit of land, or average land productivity, not only through improving the land allocation efficiency but also through increasing the average operational farm size. The effect of the higher land allocation efficiency on average land productivity is positive. However, when there exists an inverse relationship between farm size and output per unit of land, or land productivity, the effect of the larger average operational farm size on average land productivity is negative. Then, the net effect of the agricultural land marketization on average land productivity depends on the comparison of these two channels. By using the agricultural land marketization reform in China in 2008 as the indicator of marketization and the China Health and Nutrition Survey (CHNS) database, this chapter empirically finds that: first, there exists an inverse relationship between farm size and land productivity in China; second, the agricultural land marketization in China improves the land allocation efficiency and increases the average operational farm size; and third, the higher land allocation efficiency improves the average land productivity by 29.1% and the larger average operational farm size reduces the average land productivity by 9.2%, implying that the

agricultural land marketization in China finally improves the average land productivity by 19.9%.

Acknowledgement

First and foremost, I would like to express my deepest appreciation to Prof. Estelle Malavolti, Prof. Doh-Shin Jeon and Prof. Yassine Lefouili. It was Estelle who first introduced me to the economics of air transportation. Her extensive knowledge and insightful understanding in this field benefit me a lot. More importantly, she gives me great support and encouragement during the hardest period of my doctoral studies. I would like to thank Doh-Shin deeply for his guidance in the fields of industrial organization and contract theory. His expertise not only makes me have a good understanding of these fields but also gives many critical comments for my research. I am also indebted to Yassine for his patience, enthusiasm and continuous support. I am deeply impressed by and learn a lot from his rigorous attitudes and hard-working spirits, which is one of the most valuable assets for my future career.

I owe special thanks to Prof. Wanjun Yao and Prof. Shigeyuki Hamori. Their interesting ideas and hard work make it possible to bring our co-authored paper to print. My sincere thanks also go to Prof. Patrick Rey, Prof. Christian Bontemps, Prof. Wilfried Sand-Zantman, Prof. Patrick Fève, Prof. Pierre Dubois, Prof. Jan K. Brueckner, Prof. Anming Zhang, Prof. Benny Mantin, Prof. Steve Lawford, Prof. Achim Czerny, Prof. Nicole Adler, Prof. Frank Fichert, Prof. Kenneth Button, Prof. Radosav Jovanović, Prof. Hangjun Yang, Prof. Changmin Jiang, Dr. Xiyan Wang, and and the participants in a number of seminars and conferences for their helpful discussions and insightful comments. Besides, I would like to thank the rest of my thesis committee: Prof. Carole Haritchabalet and Prof. Frédéric Marty for generously investing their time in reading my thesis and providing helpful suggestions.

My gratitude must also be extended to all the professors at the Toulouse School of Economics (TSE) and the French Civil Aviation University (ENAC) who provide me important suggestions in job market training. Special thanks to Aude Schloesing and Caroline Téjédor, who are always nice and patient for helping me solve numerous problems.

In particular, I would like to thank all my colleagues and friends at TSE who are too many to mention. Their company over the past seven years brings me so much fun and happiness.

Finally, I would like to thank my beloved parents and wife and dedicate this doctoral thesis to them. In the whole life of my father, Donglin Wang, he wished me to become a good scholar and gave me huge support and encouragement. It is him who spurs me on to never give up. My mother, Chunxia Wang, and my wife, Xiaoyu Wang, take the greatest possible care of me and encourage me all the time. Without their support, I would not have been able to finish this work.

This doctoral thesis is dedicated to my beloved family.

Contents

List of Tables

List of Figures

Introduction

This thesis aims to apply the economic theories and empirical methods in the fields of industrial organization and applied microeconomics to analyze the air transportation and agricultural land markets. Specifically, this thesis consists of three self-contained papers, each of which corresponds to one chapter. The first chapter mainly uses the methodology of contract theory to address the adverse selection problem in the design of delay reduction contract in Europe. The second chapter investigates the dual roles of congestion delays and horizontal product differentiation in airline network choice. Moreover, the third chapter tests empirically whether or not the agricultural land marketization will necessarily improve the average land productivity by examining the role of the inverse relationship between farm size and land productivity.

In the context of SESAR (Single European Sky ATM[1] Research) project, the Air Navigation Service Provider (ANSP) can provide a delay reduction service to airlines. Thus, the first chapter, jointly written with Estelle Malavolti, studies the optimal design of delay reduction contract signed between an ANSP and a monopoly airline.

Aviation in Europe is expected to experience a rapid growth and more delays in the future. According to STATFOR (2013), in the most-likely scenario, there will be 14.4 million flights in Europe in 2035, 50% more than 2012. Moreover, air traffic growth will be limited by the available airport capacity. When the capacity limits are reached, congestion at airports will increase quite rapidly, leading to more delays. To satisfy the development of EU air transport sector, in 2004, European Union and EUROCONTROL founded the SESAR project, in which satisfying future safety needs and reducing delays are important targets (see European Union and EUROCONTROL, 2015). In the context

[1] ATM is the abbreviation of Air Traffic Management.

of SESAR, in order to reduce delays, the Air Navigation Service Provider (ANSP) can provide a delay reduction service to airlines. In the short run, the service can be provided for free because of generous funds of SESAR. In the long run, however, the ANSP will face financial constraints. Then, a contract, the so-called delay reduction contract in this chapter, signed between the ANSP and airlines will be necessary. Therefore, this chapter aims to study the optimal design of delay reduction contract.

Specifically, we consider an ANSP, a monopoly airline, two airports, and a contract signed between an ANSP and a monopoly airline to reduce delays, in which contracting variables are the degree of the delay reduction service and the transfer from the airline to the ANSP. In the contract design, we mainly consider the adverse selection problem, which comes from airlines' private information about their values of time, and thus the trade-off between efficiency and rent extraction. In fact, airlines may have very different values of time, for example, Air France and easyJet, and airlines always know better about their values of time than the ANSP.

Then, we derive optimal contracts analytically considering both the welfare-maximizing and profit-maximizing ANSP. In particular, we find that, under incomplete information, the optimal degree of the delay reduction service for the airline with a low value of time may be distorted downwards to decrease the information rent of the airline with a high value of time. Moreover, because contracts should be adjusted over time according to the evolution of some relevant exogenous variables, we conduct comparative-static analysis to study the effects of safety standard and flight frequency on optimal contracts. Finally, we use numerical examples to study when a welfare-maximizing ANSP has to use public funds to provide the delay reduction service, and we find that, if the service is very effective, the ANSP may not have to use public funds; if the service is very ineffective, the ANSP has to use public funds; if the effectiveness of the service is intermediate, when the effectiveness decreases, the ANSP may not have to use public funds only when the passengers' value of time becomes higher.

This chapter is closely related to three branches of literature. In the incentives theory and regulation literature, Caillaud et al. (1988) summarize two types of the regulator's objective function, that is, distributional objectives and the cost of public funds. Baron and Myerson (1982) and Baron and Besanko (1984) use the distributional-objectives objective function, while Laffont and Tirole (1986) use the objective function with the cost of public funds. This chapter considers the cost of public funds when the ANSP acts as a social planner. Specifically, because passengers can benefit from the service but do not pay to the ANSP, it is possible that the service is socially desirable while the airline's benefit from the service is not as high as the total cost of providing the service. In this case, the ANSP has to use public funds to subsidize the service and thus consider the cost of public funds in the objective function.

The second branch of literature is the modeling of passenger utility. Some studies follow Dixit (1979) to use a quadratic passenger utility function, for example, Lin (2012) and Wang (2017). The main purpose of this kind of utility function is to include two substitutable air transport services. In this chapter, however, we follow Brueckner (2004), Brueckner and Flores-Fillol (2007), Flores-Fillol (2009), and Flores-Fillol (2010) and use a linear utility function. This kind of utility function can help us simplify the analysis.

The third branch of literature is the modeling of delay function. Among others, Brueckner (2002, 2005) models the delay cost as a non-decreasing function of the number of flights during the peak travel period of a day. Moreover, US Federal Aviation Administration (1969) models delays as a convex function of the number of flights. This delay function is estimated from steady-state queuing theory and has been used by Morrison (1987), Zhang and Zhang (1997, 2003, 2006) and Basso (2008). Pels and Verhoef (2004), De Borger and Van Dender (2006), Basso and Zhang (2007), and Yang and Zhang (2011) use a linear delay function. In this chapter, we model the delay function to capture the causes of delays. Specifically, our delay function consists of the delays due to exceptional events[2] in own slot and the delays induced by other flights, in which the number of exceptional events in a slot follows a Poisson distribution.

The second chapter investigates the dual roles of congestion delays and horizontal product differentiation in airline network choice.

[2]Exceptional events can be, for example, adverse weather conditions, aircraft defects and airport facilities limitations.

Air traffic delays remain a significant and worldwide reality. In Europe, from 2005 to 2015, the percentages of delayed flights for arrivals are approximately 40%, with an average delay per delayed flight for arrivals of approximately 29 minutes (see EUROCONTROL, 2010, 2011, 2016). As a result, and considering also the time values estimated by University of Westminster (2015) and Cook and Tanner (2015), delays are costly to both airlines and passengers. In fact, it has been well established that different airline's network structures may result in varying degrees of delays.[3] However, the means by which airlines respond to these costly delays by adjusting their network structure is, to date, little studied.[4] Therefore, this chapter aims to study how congestion delays shape airline network structure, within which the role of horizontal product differentiation will also be investigated.

In the model, I consider a monopoly airline and passengers in three markets. Passengers maximize their utility, which is a quadratic function of traffic of imperfectly substitutable non-stop and one-stop air transport services. This imperfect substitution indicates the horizontal product differentiation of services. Moreover, passengers value flight frequency because they dislike schedule delays. According to Douglas and Miller (1974) and Panzar (1979), schedule delay is the absolute difference between a passenger's most preferred departure time and that of his/her actual departure. The higher the flight frequency is, the shorter schedule delays will be. Thus, considering also the travel time, the difference in flight frequency and travel time indicates the vertical product differentiation of services.

The airline maximizes its profit by choosing a network structure, flight frequencies and passenger traffic. One feature of the model is the coverage of all possible network structures in a three-city network, that is, a hub-and-spoke network (HS), point-to-point network (PP), mixed network (MX), and 2-hub network (2H), as well as a 3-hub network (3H) as an extension. Under HS (e.g., Alitalia), passengers who travel between two spoke airports are required to connect at a hub airport. Differing from HS, under PP

[3] For example, Mayer and Sinai (2003) find that hubbing is the primary economic contributor to air traffic congestion.

[4] In the literature, only Fageda and Flores-Fillol (2015) and Silva et al. (2014) consider congestion delays in airline network choice.

(e.g., Ryanair and easyJet), passengers can travel directly from one airport to any other. Moreover, under MX (e.g., Air France), passengers who travel between two spoke airports can choose either one-stop or non-stop services, implying that MX is a combination of HS and PP. Finally, under 2H (e.g., Lufthansa and Air France-KLM group), two hubs are available for connection, while under 3H, each airport works as a hub.

In addition, I assume that the airline's cost function includes only the expected congestion delay cost. In fact, this cost specification excludes the fixed cost of operating a flight and the variable cost of serving a passenger. According to Doganis (2009), the short-run marginal cost of serving an extra passenger on a flight is close to zero. In addition, Smyth and Pearce (2007) and Pearce (2013) also claim a low marginal cost per passenger. Thus, we can omit this in order to simplify the analysis. Fixed cost, however, indeed accounts for a large share of the total cost. Previous literature models fixed cost to capture the economies of traffic density and then to explain the existence of the hub-and-spoke network. In fact, modeling the fixed cost will not provide further insights but will make the model intractable. More importantly, excluding fixed cost can also help us isolate the impact of the economies of traffic density on airline network choice.

The main trade-offs in the model are between congestion delays and schedule delays and between flight frequency and travel time. For the former, if the flight frequency in one route becomes higher, congestion delays in more than one route will increase. However, passengers' schedule delays in that route will decrease. For the latter, between two spoke airports, non-stop services always have lower flight frequency but shorter travel time, while one-stop services always have higher flight frequency but longer travel time.

When solving the model, I first consider HS, PP and MX alone in order to compare them with Lin (2012). Besides the common results, I find, surprisingly, that the airline may choose PP even when the extra travel time disutility of one-stop services is relatively low. In fact, this result arises from the inclusion of congestion delays, that is, a negative network externality. Specifically, on the one hand, because of the traffic concentration of HS and the convex cost function with respect to flight frequencies, the introduction of congestion delays, or negative network externality on the cost side, creates a cost disadvantage for HS, and thus makes HS less profitable. On the other hand, the introduction of an omitted negative network externality also reduces the region that makes MX feasible. Then, there will emerge an interval of extra travel time disutility, in which the disutility is too high to make HS more profitable and too low to make MX feasible, leaving PP as the airline's optimal network structure. The key insight from this result is that including the omitted negative network externality makes HS and MX less effective than previously understood as in Lin (2012). In fact, in addition to the commonly received advantage of PP, that is, saving the extra travel time of connecting at a hub airport, this result might provide another explanation for why some legacy airlines start to use PP in some local markets.

I then incorporate 2H into the analysis. I find that, without considering the airline's fixed investments of developing a hub airport, 2H will dominate the three other network structures as long as it is feasible. Because non-stop and one-stop services are imperfectly substitutable, passengers can obtain higher utility if both non-stop and one-stop services are available to choose than if only one of them is available. Under 2H, there are two markets, in which both non-stop and one-stop services are available. Under MX, there is only one such market. However, under HS and PP, passengers in any market cannot choose between non-stop and one-stop services. Therefore, because of the exploitation of horizontal product differentiation to a larger extent, 2H can generate higher passenger utility and then a higher airline profit than the three other network structures. To summarize, this result shows the role of horizontal product differentiation in improving passenger utility and airline profit.

In fact, this result can also help us understand the multi-hubbing and de-hubbing phenomena in the airline industry. In reality, we can observe that some airlines develop new hubs and then use 2H (multi-hubbing), while some others change from 2H to a single hub network (de-hubbing). Here comes a question: what is the motivation for airlines to use 2H? One answer is that when an airline's hub airport is congested, the airline can develop another hub to reduce the congestion of the previous hub. However, this answer might not be strong enough because besides 2H, PP and MX can also reduce hub airport congestion. The result above shows the role of horizontal product differentiation and thus provides another explanation for the use of 2H.

Moreover, I conduct comparative-static analysis and find that, for the airline under MX and 2H, there exist some strategic effects when the values of the parameters change, due to the division of local and connecting traffic in one market. For instance, under MX, when the marginal congestion delay cost increases, the change of flight frequency between two spoke airports depends on the trade-off between the direct negative effect of a higher delay cost and the strategic redistribution of traffic among different routes. Furthermore, in welfare analysis, I derive not only the first-best, but also the second-best, socially optimal network structure, and both show the airline's inefficient biases towards PP and 2H. Besides which, I extend the analysis to 3H and again show the role of horizontal product differentiation.

The contributions of this work are threefold. First, the majority of the airline network choice literature (see Oum et al., 1995; Berechman and Shy, 1998; Kawasaki, 2008) compare HS and PP alone, with a few others including one more network structure, either MX (see Lin, 2012) or 2H (see Alderighi et al., 2005). In fact, Starr and Stinchcombe (1992) and Hendricks et al. (1995, 1999) use rather general models, allowing the network design to be endogenous. However, their frameworks focus mainly on airline cost but not on passenger demand, which thus leaves little room for the optimality of network structures other than HS and PP. Accordingly, this chapter contributes to the literature by incorporating all possible network structures in a three-city network, each of which has the potential to be an airline's optimal network structure.

Second, most of the previous studies explain airline network choice from the point of view of the economies of traffic density (see Bittlingmayer, 1990; Hendricks et al., 1995, 1999), demand uncertainty (see Barla and Constantatos, 2005) and schedule delays (see Brueckner, 2004), while only Fageda and Flores-Fillol (2015) and Silva et al. (2014) consider congestion delays. Therefore, this chapter contributes to the currently limited literature available on the effects of congestion delays on airline network choice.

The third is regarding the perspective of analyzing 2H. The conventional wisdom concerning the function of 2H (see Bilotkach et al., 2013) is to divert passengers from one hub to another and thus reduce hub airport congestion. In fact, a four-city network is the minimum requirement in order to show the congestion reduction function of 2H; however, this inevitably brings analytical difficulties. To make the analysis tractable, previous literature (see Bilotkach et al., 2013) has had to simplify many important elements. Nevertheless, in this chapter, I focus on the exploitation of horizontal product differentiation in 2H, which requires a three-city network only. In reality, the horizontal product differentiation can come from, for instance, different departure time slots (see Encaoua et al., 1996) and brand loyalty (see Brueckner and Whalen, 2000; Brueckner and Flores-Fillol, 2007).

This chapter is closely related to three branches of literature. Since the airline deregulation in the USA in 1978, there is a growing body of literature on airline network choice, in which the first branch is the scope of network structures. Previously, the literature has focused on the choice between HS and PP, but recently there has been a shift towards some "real-life" network structures, that is, MX and 2H. For MX, among others, Dunn (2008) empirically examines an airline's choice of providing non-stop services or not, given that the airline has (not) provided one-stop services. Moreover, Fageda and Flores-Fillol (2012) study hub airlines' incentives to provide non-stop services between spoke airports under two recent innovations, that is, the regional jet technology and low-cost business model. Further, Lin (2012) studies the network choice among HS, PP and MX under both monopoly and duopoly setups. For 2H, Bilotkach et al. (2013) use the utility function à la Mussa-Rosen (see Mussa and Rosen, 1978) to study the function of diverting traffic of 2H. Moreover, Wang (2016) theoretically examines the optimality of 2H in the spirit of Brueckner and Spiller (1991).

The second branch of literature is the theory explaining airline network choice. One theory is that airlines can better exploit the economies of traffic density under HS. According to Hendricks et al. (1995), the economies of traffic density arise when the cost per passenger on a route decreases with the number of passengers flying on that route. As a result, because HS has a higher traffic density than PP, as long as the cost of extra travel time of one-stop services is not high enough, the total cost for a given level of demand may be lower under HS than PP. Some empirical studies have confirmed the economies of traffic density under HS (see Brueckner et al., 1992; Brueckner and Spiller, 1994). Moreover, besides Hendricks, Piccione, and Tan (1995), theoretical studies explaining airline network choice from the point of view of the economies of traffic density include, for example, Bittlingmayer (1990), Oum et al. (1995) and Hendricks et al. (1999) [5].

Another theory concerns demand uncertainty. Barla and Constantatos (2005) show the flexibility of HS under uncertainty. Interestingly, they also find that both airlines may choose PP, because by committing not to enjoy the flexibility, airlines can avoid the spread of competition from one market to others. Furthermore, Hu (2010) also considers demand uncertainty but under a different setup.

Because passengers greatly value flight frequency (see Berry and Jia, 2010), schedule delays may be an important factor affecting airline network choice. Berechman and Shy (1998) and Brueckner and Zhang (2001) first connected airline network structure and scheduling. Then, Brueckner (2004) builds a framework that improves upon the previous two studies and shows that a high disutility of schedule delays would be conducive to HS. In addition, other relevant studies include Kawasaki (2008) and Flores-Fillol (2009).

Furthermore, some studies have introduced congestion delays into their models. In Fageda and Flores-Fillol (2015), they find that, even if there is a higher delay cost, duopoly airlines exhibit a preference for HS, which may be inefficient from the perspective of a welfare-maximizing social planner. Moreover, Silva et al. (2014) also consider congestion delays and show that a higher value of travel time favors PP. However, the frameworks of these two studies do not allow for the analysis of MX and 2H. Besides which, in Fageda and Flores-Fillol (2015), passenger demand is also perfectly inelastic.

The last theory is the horizontal product differentiation. Lin (2012) finds that HS will be the airline's optimal network structure if passengers do not differentiate between non-stop and one-stop services too much, and if the extra travel time disutility of one-stop services is low. However, if the passenger differentiation is substantial, MX (resp.

[5] Hendricks et al. (1999) also consider the nature of competition, that is, airlines compete aggressively or not.

PP) will be the airline's optimal network structure when the extra travel time disutility is low (resp. high).

The third branch of literature is the modeling of vertical and horizontal product differentiation. With respect to the vertical product differentiation, Flores-Fillol (2010) considers that airlines may compete in flight frequencies. Moreover, in the models of Brueckner (2004) and Kawasaki (2008), one-stop services always cost more time than non-stop services. With respect to the horizontal product differentiation, Encaoua et al. (1996) consider the difference of the departure time slot, while Brueckner and Whalen (2000) and Brueckner and Flores-Fillol (2007) consider brand loyalty. In addition, Lin (2012) uses the quadratic utility function in Dixit (1979) to capture the horizontal product differentiation.

The third chapter, jointly written with Wanjun Yao and Shigeyuki Hamori, tests empirically whether or not the agricultural land marketization will necessarily improve the average land productivity by examining the role of the inverse relationship between farm size and land productivity.

For developing countries, especially those in transition from agricultural to non-agricultural economy, on the one hand, the transition of economy reduces the amount of agricultural labor significantly and thus decreases the utilization rate of agricultural land. On the other hand, the transition increases the demand of agricultural products of urban areas and thus further aggravates the balance between supply and demand. Given the reality that the domestic farmland cannot be enlarged easily, governments in many countries try to improve the output per unit of land, or land productivity, to increase the supply of agricultural products.

According to economic theory, the agricultural land marketization can improve the land allocation efficiency. After the agricultural land marketization, less efficient agricultural producers can rent out or sell some of their land at a price higher than their marginal production, while more efficient producers can rent in or buy some land at a price lower than their marginal production. Finally, the agricultural land will be allocated more efficiently through market mechanism (see Yao, 2000; Benjamin and Brandt, 2002;

Carter and Yao, 2002; Deininger and Jin, 2005, 2008; Deininger et al., 2008a; Deininger et al., 2008b; Jin and Deininger, 2009; Barrett et al., 2010). Then, if the agricultural land marketization can improve the land allocation efficiency, can it also improve the average output per unit of land, or average land productivity? The conventional answer is affirmative because the higher land allocation efficiency implies the higher average land productivity (see, for example, Restuccia and Santaeulalia-Llopis, 2017[6]). However, if we consider the inverse relationship between farm size and land productivity, the answer is uncertain.

In many developing countries, there exists an inverse relationship between farm size and land productivity.[7] That is, compared to rural households with a large farm size, those with a small farm size have higher land productivity. This relationship has been found in the countries of Asia (see Sen, 1962; Lau and Yotopoulos, 1971; Bardhan, 1973; Rao and Chotigeat, 1981; Carter, 1984; Newell et al., 1997; Heltberg, 1998; Lamb, 2003), Africa (see Collier, 1983; Barrett, 1996; Byiringiro and Reardon, 1996; Kimhi, 2006; Carletto et al., 2013; Larson et al., 2014; Ali and Deininger, 2015), Europe (see Chayanov, 1926; Alvarez and Arias, 2004), and Latin America (see Berry and Cline, 1979; Cornia, 1985).[8]

After the agricultural land marketization, on the one hand, the previously unused land can be used again, and thus the total operational farm size may increase. Given the amount of rural households, the average operational farm size may also increase. On the other hand, rural households can obtain monetary incomes from land transactions, which can provide a basic guarantee for their migrations to urban areas. In this way, the amount of rural households may decrease, and the average operational farm size may

[6]Restuccia and Santaeulalia-Llopis (2017) use household-level data from Malawi and find that a reallocation of production factors to their efficient use will result in higher average total factor productivity (TFP) of farmers, in which the farm TFP and the output per unit of land are found to be strongly positively correlated across farms because the allocation of land is not related to productivity so many productive farmers are constrained by size.

[7]The reasons explaining the existence of inverse relationship include, among others, land market imperfections (see Heltberg, 1998; Lamb, 2003), labor market imperfections (see Sen, 1966; Rosenzweig and Wolpin, 1985; Frisvold, 1994), credit market imperfections (see Feder, 1985; Eswaran and Kotwal, 1986; Carter, 1988), and risk (see Wiens, 1977; Rosenzweig and Binswanger, 1993; Kevane, 1996).

[8]Some studies show that, in USA and Japan, farm size is positively correlated with land productivity (see Sumner, 2014; Kawasaki, 2010).

then increase. Under the inverse relationship, the increase of average operational farm size will reduce the average land productivity.

To summarize, the agricultural land marketization affects the average land productivity not only through improving the land allocation efficiency but also through increasing the average operational farm size.[9] The improvement of land allocation efficiency has a positive effect on average land productivity. However, when there exists an inverse relationship between farm size and land productivity, the increase of average operational farm size has a negative effect on average land productivity. Therefore, the agricultural land marketization does not necessarily improve the average land productivity. Only when the positive effect of the higher land allocation efficiency dominates the negative effect of the larger average operational farm size, the marketization will finally improve the average land productivity.

In this chapter, we use the year 2008 as the indicator of the agricultural land marketization in China to test empirically the effect of the marketization on average land productivity. The empirical framework is the one for the study of inverse relationship (see Binswanger et al., 1995; Assunção and Braido, 2007; Barrett et al., 2010; Carletto et al., 2013) and the data we use is from the China Health and Nutrition Survey (CHNS) database[10]. Finally, we find that: first, there exists an inverse relationship between farm size and land productivity in China; second, the agricultural land marketization in China improves the land allocation efficiency and increases the average operational farm size; third, the higher land allocation efficiency improves the average land productivity by 29.1% and the larger average operational farm size reduces the average land productivity by 9.2%, implying that the agricultural land marketization in China finally improves the average land productivity by 19.9%.

[9]The agricultural land marketization affects the average land productivity also through, for example, influencing indirectly the amount of labor input and intermediate inputs. However, in this chapter, we focus our discussions on the direct effects of the marketization, that is, improving the land allocation efficiency and increasing the average operational farm size.

[10]http://www.cpc.unc.edu/projects/china

Chapter 1

Contract Design for EU Air Traffic Delay Reduction

1.1 Introduction

Aviation in Europe is expected to experience a rapid growth and more delays in the future. According to STATFOR (2013), in the most-likely scenario, there will be 14.4 million flights in Europe in 2035, 50% more than 2012. Moreover, air traffic growth will be limited by the available airport capacity. When the capacity limits are reached, congestion at airports will increase quite rapidly, leading to more delays. To satisfy the development of EU air transport sector, in 2004, European Union and EUROCONTROL founded the SESAR (Single European Sky ATM[1] Research) project, in which satisfying future safety needs and reducing delays are important targets (see European Union and EUROCONTROL, 2015). In the context of SESAR, in order to reduce delays, the Air Navigation Service Provider (ANSP) can provide a delay reduction service to airlines.

Under the delay reduction service, when facing potential delays, an airline will contact the ANSP to find a solution to reduce delays. After receiving the airline's request, the ANSP can find out several solutions satisfying all regulation constraints. Then, by costly calculation, evaluation and coordination, the ANSP can determine the solution which can reduce delays most and then implement it. In the short run, the service can be provided

[1] ATM is the abbreviation of Air Traffic Management.

for free because of generous funds of SESAR. In the long run, however, the ANSP will face financial constraints. Then, a contract, the so-called delay reduction contract in this chapter, signed between the ANSP and airlines will be necessary. Therefore, this chapter aims to study the optimal design of delay reduction contract.

Specifically, we consider an ANSP, a monopoly airline, two airports, and a contract signed between an ANSP and a monopoly airline to reduce delays, in which contracting variables are the degree of the delay reduction service and the transfer from the airline to the ANSP. In the contract design, we mainly consider the adverse selection problem, which comes from airlines' private information about their values of time, and thus the trade-off between efficiency and rent extraction. In fact, airlines may have very different values of time, for example, Air France and easyJet, and airlines always know better about their values of time than the ANSP.

Then, we derive optimal contracts analytically considering both the welfare-maximizing and profit-maximizing ANSP. In particular, we find that, under incomplete information, the optimal degree of the delay reduction service for the airline with a low value of time may be distorted downwards to decrease the information rent of the airline with a high value of time. Moreover, because contracts should be adjusted over time according to the evolution of some relevant exogenous variables, we conduct comparative-static analysis to study the effects of safety standard and flight frequency on optimal contracts. Finally, we use numerical examples to study when a welfare-maximizing ANSP has to use public funds to provide the delay reduction service, and we find that, if the service is very effective, the ANSP may not have to use public funds; if the service is very ineffective, the ANSP has to use public funds; if the effectiveness of the service is intermediate, when the effectiveness decreases, the ANSP may not have to use public funds only when the passengers' value of time becomes higher.

Related Literature

This chapter is closely related to three branches of literature. In the incentives theory and regulation literature, Caillaud et al. (1988) summarize two types of the regulator's objective function, that is, distributional objectives and the cost of public funds. Baron and Myerson (1982) and Baron and Besanko (1984) use the distributional-objectives objective function, while Laffont and Tirole (1986) use the objective function with the cost of public funds. This chapter considers the cost of public funds when the ANSP acts as a social planner. Specifically, because passengers can benefit from the service but do not pay to the ANSP, it is possible that the service is socially desirable while the airline's benefit from the service is not as high as the total cost of providing the service. In this case, the ANSP has to use public funds to subsidize the service and thus consider the cost of public funds in the objective function.

The second branch of literature is the modeling of passenger utility. Some studies follow Dixit (1979) to use a quadratic passenger utility function, for example, Lin (2012) and Wang (2017). The main purpose of this kind of utility function is to include two substitutable air transport services. In this chapter, however, we follow Brueckner (2004), Brueckner and Flores-Fillol (2007), Flores-Fillol (2009), and Flores-Fillol (2010) and use a linear utility function. This kind of utility function can help us simplify the analysis.

The third branch of literature is the modeling of delay function. Among others, Brueckner (2002, 2005) models the delay cost as a non-decreasing function of the number of flights during the peak travel period of a day. Moreover, US Federal Aviation Administration (1969) models delays as a convex function of the number of flights. This delay function is estimated from steady-state queuing theory and has been used by Morrison (1987), Zhang and Zhang (1997, 2003, 2006) and Basso (2008). Pels and Verhoef (2004), De Borger and Van Dender (2006), Basso and Zhang (2007), and Yang and Zhang (2011) use a linear delay function. In this chapter, we model the delay function to capture the causes of delays. Specifically, our delay function consists of the delays due to exceptional events[2] in own slot and the delays induced by other flights, in which the number of exceptional events in a slot follows a Poisson distribution.

The rest of the chapter is organized as follows. Section 1.2 introduces the model. Section 1.3 derives optimal delay reduction contracts. Section 1.4 studies the adjustments

[2]Exceptional events can be, for example, adverse weather conditions, aircraft defects and airport facilities limitations.

of optimal contracts. That is, this section conducts comparative-static analysis to study the effects of safety standard and flight frequency on optimal contracts. Section 1.5 uses numerical examples to study when a welfare-maximizing ANSP has to use public funds to provide the delay reduction service. Section 1.6 concludes.

1.2 The Model

We consider an ANSP, a monopoly airline, passengers with a mass N, and an air transport market connecting two airports.

Conditional on the use of the airline, following Brueckner (2004), Brueckner and Flores-Fillol (2007), Flores-Fillol (2009), and Flores-Fillol (2010), the passenger utility is:

$$v = y - p + b + a(s) - \alpha D(s). \tag{1.1}$$

In (1.1), y is the passengers' income; p is the fare; b is the passengers' travel benefit which is uniformly distributed on the support $[\zeta, \xi]$; $a(s)$ is the passengers' utility gain from a safety standard s with $a^{'}(s) \geqslant 0$; α is the passengers' value of time[3]; and $D(s)$ is expected delays per flight (in time units). Moreover, the safety standard s is exogenous and can vary within $[\underline{s}, \overline{s}]$. Note that $\underline{s}$ is far higher than the minimum safety requirement.

[3] According to University of Westminster (2015), three types of passenger costs of delay may be considered: "hard" costs (borne by the airline, such as re-booking and compensation), "soft" costs (borne by the airline, such as the loss of market share due to passenger dissatisfaction) and "internalized" costs (borne by the passenger and not passed on to the airline, such as potential loss of business due to late arrival at meeting). The passengers' value of time in this model mainly refers to the "internalized" costs in University of Westminster (2015).

The specification of expected delays per flight[4] is:

$$D(s) = 2\left[\sum_{k=0}^{+\infty} \frac{\left(\frac{\beta T}{f}\right)^k e^{-\left(\frac{\beta T}{f}\right)}}{k!} k g(s) + \gamma\beta \left(\frac{T}{f}\right)^{-1} g(s)\right]. \tag{1.2}$$

In (1.2), the first term in the square brackets is the delays due to exceptional events in own slot. We assume that the number of exceptional events in a slot follows a Poisson distribution with parameter $\frac{\beta T}{f}$, in which β is the exceptional event arriving rate; T is the number of available hours; and f is the flight frequency. Assume that flights are evenly spaced during available hours. Thus, the duration of a slot is $\frac{T}{f}$. k is the number of exceptional events. $g(s)$ is the amount of delays caused by an exceptional event. $g'(s) \geqslant 0$ captures the fact that the higher the safety standard is, the longer delays will be. The second term in the square brackets is the delays induced by other flights, which decreases with $\frac{T}{f}$ and increases with β and $g(s)$. Moreover, the parameter $\gamma > 0$ is the so-called delay externality parameter in this chapter. A greater γ implies a severer effect from other flights. In fact, the second term can represent the delays caused by airport congestion. The square brackets times two because there are two airports.

Passengers also have an outside option, for instance, traveling by train. Conditional on the use of the outside option, the passenger utility is:

$$v_0 = y + z, \tag{1.3}$$

In (1.3), z is the net benefit of the outside option.

Thus, a passenger chooses to travel by plane when:

$$y - p + b + a(s) - \alpha D(s) \geqslant y + z, \tag{1.4}$$

[4] According to Cook and Tanner (2011), the cost of delays for airlines is calculated for strategic delays (those accounted for in advance) and tactical delays (those incurred on the day of operations and not accounted for in advance). Strategic delays are for adding buffer to the airline schedule. Tactical delays include primary delays and secondary or reactionary delays, in which original delays caused by one aircraft (primary delays) cause "knock-on" effects in the rest of the network (known as secondary or reactionary delays). In fact, air traffic delays in this model mainly refer to tactical delays in Cook and Tanner (2011). Moreover, the delays due to exceptional events in own slots and the delays induced by other flights in the delay function of this model correspond roughly to the primary and reactionary delays, as defined in Cook and Tanner (2011), respectively.

that is, $b \geqslant p - a(s) + \alpha D(s) + z$. The air traffic then equals:

$$q = \int_{p-a(s)+\alpha D(s)+z}^{\xi} \frac{N}{\xi - \zeta} db$$
$$= [\xi - p + a(s) - \alpha D(s) - z] \frac{N}{\xi - \zeta}. \quad (1.5)$$

The airline's cost is:

$$c_{airline} = \tau q + \delta f + \theta f D(s). \quad (1.6)$$

In (1.6), τq is the variable cost, in which τ is the marginal cost per seat; δf is the fixed cost, in which δ is the fixed operating cost of a flight; and $\theta f D(s)$ is the supply side delay cost, in which θ is the airline's value of time. θ may be unobservable to the ANSP. However, it is common knowledge that θ belongs to the set $\Theta = \{\overline{\theta}, \underline{\theta}\}$, in which $\overline{\theta}, \underline{\theta} > 0$ and $\Delta\theta = \overline{\theta} - \underline{\theta} > 0$. If θ is the airline's private knowledge, the airline can be the one with $\overline{\theta}$ or $\underline{\theta}$ with probabilities μ and $1 - \mu$, respectively.

The airline maximizes profit by choosing the fare[5], that is:

$$\max_p \pi = pq(p) - [\tau q(p) + \delta f + \theta f D(s)]. \quad (1.7)$$

Letting $\eta = \frac{N}{\xi - \zeta}$, the optimal solution of (1.7) is:

$$p^*(s) = \frac{1}{2}[\xi + \tau + a(s) - z - \alpha D(s)], \quad (1.8)$$
$$q^*(s) = \frac{1}{2}\eta[\xi - \tau + a(s) - z - \alpha D(s)], \quad (1.9)$$
$$\pi^*(\theta, s) = \underbrace{-\frac{1}{4}\eta\{2[\xi - \tau + a(s) - z] - \alpha D(s)\}\alpha D(s)}_{\text{demand side delay cost}} - \underbrace{\theta f D(s)}_{\text{supply side delay cost}}$$
$$+ \frac{1}{4}\eta[\xi - \tau + a(s) - z]^2 - \delta f. \quad (1.10)$$

In (1.10), we can find both the airline's demand and supply side delay costs. Then, the

[5]In this model, flight frequency is not a endogenous decision variable of the airline. One reason can be the slot control in Europe. That is, at all major European airports, take-off and landing slots are allocated through grandfather right and "use it or lose it" rule. However, in the comparative-static analysis, we will study how the change of flight frequency affects optimal contracts.

passenger utility and surplus are:

$$v^*(s) = y - p^*(s) + b + a(s) - \alpha D(s), \tag{1.11}$$

$$\begin{aligned} ps^*(s) &= \underbrace{\int_{p^*(s)-a(s)+\alpha D(s)+z}^{\xi} [y - p^*(s) + b + a(s) - \alpha D(s)]\,\eta db}_{\text{from air traffic}} + \underbrace{\int_{\zeta}^{p^*(s)-a(s)+\alpha D(s)+z} (y+z)\,\eta db}_{\text{from outside option traffic}} \\ &= \underbrace{-\frac{1}{8}\eta\{2[\xi - \tau + a(s) - z] - \alpha D(s)\}\alpha D(s)}_{\text{passenger delay cost}} + \frac{1}{8}\eta[\xi + \tau - a(s) + z]^2 \\ &\quad - \frac{1}{2}\eta\xi[\xi + \tau - a(s) - z] + \frac{1}{2}\eta\xi^2 - \eta\zeta(y+z) + \eta\xi y, \end{aligned} \tag{1.12}$$

respectively. In (1.12), we can find the total passenger utility loss resulting from delays.

The ANSP signs a contract with the airline, in which contracting variables are r and t. $r \in [0,1]$ is the degree of the delay reduction service that the ANSP provides to the airline and t is the transfer from the airline to the ANSP. After signing the contract, expected delays per flight reduce from $D(s)$ to $D(s)[1 - \sigma \ln(1+r)]$. $\sigma \ln(1+r)$ is the fraction of delay reduction, in which $\sigma \in \left[0, \frac{1}{\ln 2}\right]$ measures the effectiveness of the service and $\ln(1+r)$ captures that the marginal value of the service is positive but decreasing with the degree. Then, the fare, air traffic, airline profit, passenger utility, and passenger surplus will be:

$$P^*(s,r) = p^*(s) + \frac{1}{2}\alpha D(s)\sigma \ln(1+r), \tag{1.13}$$

$$Q^*(s,r) = q^*(s) + \frac{1}{2}\eta\alpha D(s)\sigma \ln(1+r), \tag{1.14}$$

$$\begin{aligned} \Pi^*(\theta,s,r) = &\underbrace{\pi^*(\theta,s)}_{\text{initial profit}} + \underbrace{\frac{1}{4}\eta\alpha^2 D(s)^2 \sigma^2 [\ln(1+r)]^2 + q^*(s)\alpha D(s)\sigma\ln(1+r)}_{\text{demand side delay reduction benefit}} \\ &+ \underbrace{\theta f D(s)\sigma\ln(1+r)}_{\text{supply side delay reduction benefit}}, \end{aligned} \tag{1.15}$$

$$V^*(s,r) = v^*(s) + \frac{1}{2}\alpha D(s)\sigma\ln(1+r), \tag{1.16}$$

$$PS^*(s,r) = \underbrace{ps^*(s)}_{\text{initial passenger surplus}} + \underbrace{\frac{1}{2}\left\{\frac{1}{4}\eta\alpha^2 D(s)^2\sigma^2[\ln(1+r)]^2 + q^*(s)\alpha D(s)\sigma\ln(1+r)\right\}}_{\text{passenger delay reduction benefit}}, \tag{1.17}$$

respectively. According to (1.13), (1.15) and (1.17), we can find that the airline can enjoy both demand and supply side delay reduction benefits and passengers can enjoy higher surplus even though the fare increases.

Finally, the ANSP's cost of providing the service is:

$$C_{ANSP}(s,r) = m(s)\,r. \tag{1.18}$$

In (1.18), $m(s)$ is the marginal cost of the service, which increases with the safety standard, that is, $m'(s) \geqslant 0$. In fact, when providing the service, the ANSP has to spend more time on evaluation and coordination for satisfying a higher safety standard, which will inevitably result in a higher cost.[6]

The timeline of the model is shown in Figure 1.1.

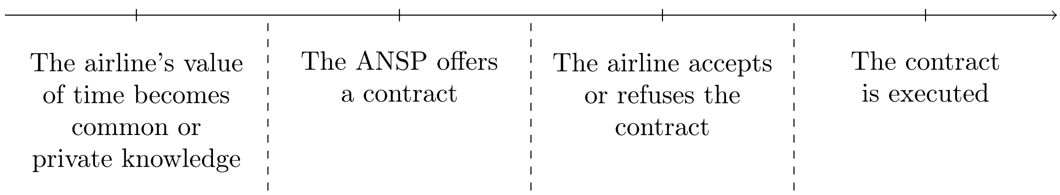

Figure 1.1: Timeline

1.3 Optimal Contracts

Next, we will derive optimal contracts by considering both the welfare-maximizing and profit-maximizing ANSP.

1.3.1 Welfare-Maximizing ANSP

The welfare-maximizing ANSP's objective is to maximize social welfare, that is:

$$\max_{\{(r,t)\}} W = PS^*(s,r) + \Pi^*(\theta,s,r) - C_{ANSP}(s,r) - \lambda\left[C_{ANSP}(s,r) - t\right]\mathbb{1}_{t<C(s,r)}, \tag{1.19}$$

[6] Another feasible setup is the linear benefit and convex cost, which is essentially equivalent to our setting, that is, the concave benefit and linear cost.

in which λ is the shadow cost of public funds.

Next, we discuss the model solution according to the passengers' value of time α.[7]

1.3.1.1 Scenario 1: $\alpha = 0$

When $\alpha = 0$, passengers cannot enjoy any benefit from the service, implying that the airline in fact delegates the provision of the service to the ANSP. Thus, the optimization problem of the ANSP, or the airline, is:

$$\max_r W = \Pi^* (\theta, s, r) - C_{ANSP} (s, r) . \tag{1.20}$$

Then, by taking the first-order and second-order conditions of (1.20), we can obtain that optimal contracts for the airline with $\overline{\theta}$ and $\underline{\theta}$ are $\left\{ \left(\overline{r}^*, \overline{t}^* \right) \right\}$ and $\{(\underline{r}^*, \underline{t}^*)\}$, respectively, in which:

$$\overline{r}^* = \frac{\overline{\theta} f D (s) \sigma}{m (s)} - 1, \tag{1.21}$$

$$\overline{t}^* \in \left[m (s) \overline{r}^*, \Pi^* \left(\overline{\theta}, s, \overline{r}^* \right) - \pi^* \left(\overline{\theta}, s \right) \right] , \tag{1.22}$$

$$\underline{r}^* = \frac{\underline{\theta} f D (s) \sigma}{m (s)} - 1, \tag{1.23}$$

$$\underline{t}^* \in [m (s) \underline{r}^*, \Pi^* (\underline{\theta}, s, \underline{r}^*) - \pi^* (\underline{\theta}, s)] . \tag{1.24}$$

Because $\overline{\theta} > \underline{\theta}$, we have $\overline{r}^* > \underline{r}^*$, that is, the airline with a high value of time will enjoy a higher degree of the service.

Moreover, by checking $\overline{r}^* > 0$ and $\underline{r}^* > 0$, we can find that the delay reduction service will be provided to the airline with $\overline{\theta}$ and $\underline{\theta}$ if and only if $\sigma > \frac{m(s)}{\overline{\theta} f D(s)}$ and $\sigma > \frac{m(s)}{\underline{\theta} f D(s)}$ hold, respectively, that is, the service is effective enough.

[7]We focus our following discussions on interior solutions. Moreover, these solutions can be implemented only when the social values of the service are non-negative, that is, $[PS^* (s, r) - ps^* (s)] + [\Pi^* (\theta, s, r) - \pi^* (\theta, s)] - C_{ANSP} (s, r) - \lambda [C_{ANSP} (s, r) - t] \mathbb{1}_{t<C(s,r)} \geqslant 0$.

1.3.1.2 Scenario 2: $\alpha > \frac{\xi-\tau+a(s)-z}{D(s)}$

According to the passenger utility, we have:

$$\xi \geqslant b \geqslant p - a(s) + \alpha D(s) + z \geqslant \tau - a(s) + \alpha D(s) + z$$
$$\Rightarrow \alpha \leqslant \frac{\xi - \tau + a(s) - z}{D(s)}. \tag{1.25}$$

Thus, when $\alpha > \frac{\xi-\tau+a(s)-z}{D(s)}$, no passenger will choose to travel by plane. Then, the air transport market will close down and there will be no such contract.

1.3.1.3 Scenario 3: $0 < \alpha \leqslant \frac{\xi-\tau+a(s)-z}{D(s)}$ and High Airline Benefit

In this scenario, both passengers and the airline can benefit from the service and the airline's benefit is higher than the ANSP's cost of providing the service.

Complete Information Under complete information, it is optimal for the ANSP to set the transfer at least as the cost of providing the service. Thus, the optimization problem of the ANSP is:

$$\max_r W = PS^*(s,r) + \Pi^*(\theta,s,r) - C_{ANSP}(s,r). \tag{1.26}$$

Then, by taking the first-order condition of (1.26), we can obtain that the first-best optimal contracts for the airline with $\overline{\theta}$ and $\underline{\theta}$ are $\left\{\left(\overline{r}^{FB}, \overline{t}^{FB}\right)\right\}$ and $\left\{\left(\underline{r}^{FB}, \underline{t}^{FB}\right)\right\}$, respectively. Specifically, as shown in (1.27) and (1.29), $\overline{r}^{FB}$ and $\underline{r}^{FB}$ are determined by $\overline{\Omega}$ and $\underline{\Omega}$, respectively. For both $\overline{\Omega}$ and $\underline{\Omega}$, the optimal degree is determined by the intersection of a logarithmic and linear function of r. According to (1.27) and (1.29), because $\overline{\theta} > \underline{\theta}$, we also have $\overline{r}^{FB} > \underline{r}^{FB}$. A detailed discussion about the second-order condition of (1.26) is in Section 1.7.1.1. Discussions about other second-order conditions in this chapter are similar to Section 1.7.1.1 and thus are omitted hereafter.[8] Moreover,

[8] All omitted discussions about second-order conditions are available upon request.

$\bar{t}^{FB}$ and $\underline{t}^{FB}$ are given by (1.28) and (1.30), respectively.

$$\overline{\Omega} = 3\eta\alpha^2 D(s)^2 \sigma^2 \ln\left(1+\bar{r}^{FB}\right) - \left\{4m(s)\left(1+\bar{r}^{FB}\right)\right.$$
$$\left. -3\eta\left[\xi - \tau + a(s) - z - \alpha D(s)\right]\alpha D(s)\sigma - 4\bar{\theta} f D(s)\sigma\right\} = 0, \quad (1.27)$$
$$\bar{t}^{FB} \in \left[m(s)\bar{r}^{FB}, \Pi^*\left(\bar{\theta}, s, \bar{r}^{FB}\right) - \pi^*\left(\bar{\theta}, s\right)\right], \quad (1.28)$$
$$\underline{\Omega} = 3\eta\alpha^2 D(s)^2 \sigma^2 \ln\left(1+\underline{r}^{FB}\right) - \left\{4m(s)\left(1+\underline{r}^{FB}\right)\right.$$
$$\left. -3\eta\left[\xi - \tau + a(s) - z - \alpha D(s)\right]\alpha D(s)\sigma - 4\underline{\theta} f D(s)\sigma\right\} = 0, \quad (1.29)$$
$$\underline{t}^{FB} \in \left[m(s)\underline{r}^{FB}, \Pi^*\left(\underline{\theta}, s, \underline{r}^{FB}\right) - \pi^*\left(\underline{\theta}, s\right)\right]. \quad (1.30)$$

Incomplete Information Under incomplete information, the first-best optimal degrees of the service can be feasible and public funds may still not be used.

The ANSP maximizes the expected social welfare, that is:

$$\max_{\{(\bar{r},\underline{r})\}} W = \mu\left[PS^*(s,\bar{r}) + \Pi^*\left(\bar{\theta}, s, \bar{r}\right) - C_{ANSP}(s,\bar{r})\right]$$
$$+ (1-\mu)\left[PS^*(s,\underline{r}) + \Pi^*\left(\underline{\theta}, s, \underline{r}\right) - C_{ANSP}(s,\underline{r})\right]. \quad (1.31)$$

For feasible separating contracts, the airline's incentive compatibility and participation constraints must be satisfied, that is:

$$\Pi^*\left(\bar{\theta}, s, \bar{r}\right) - \bar{t} \geqslant \Pi^*\left(\bar{\theta}, s, \underline{r}\right) - \underline{t}, \quad (1.32)$$
$$\Pi^*\left(\underline{\theta}, s, \underline{r}\right) - \underline{t} \geqslant \Pi^*\left(\underline{\theta}, s, \bar{r}\right) - \bar{t}, \quad (1.33)$$
$$\Pi^*\left(\bar{\theta}, s, \bar{r}\right) - \bar{t} \geqslant \pi^*\left(\bar{\theta}, s\right), \quad (1.34)$$
$$\Pi^*\left(\underline{\theta}, s, \underline{r}\right) - \underline{t} \geqslant \pi^*\left(\underline{\theta}, s\right). \quad (1.35)$$

Denoting the information rent of the airline with $\bar{\theta}$ and $\underline{\theta}$ by $\bar{u}$ and $\underline{u}$, respectively, in which:

$$\bar{u} = \Pi^*\left(\bar{\theta}, s, \bar{r}\right) - \pi^*\left(\bar{\theta}, s\right) - \bar{t}, \quad (1.36)$$
$$\underline{u} = \Pi^*\left(\underline{\theta}, s, \underline{r}\right) - \pi^*\left(\underline{\theta}, s\right) - \underline{t}. \quad (1.37)$$

Then, we can write the airline's incentive compatibility and participation constraints as:

$$\overline{u} \geqslant \underline{u} + \Delta\theta f D(s) \sigma \ln(1+\underline{r}), \tag{1.38}$$

$$\underline{u} \geqslant \overline{u} - \Delta\theta f D(s) \sigma \ln(1+\overline{r}), \tag{1.39}$$

$$\overline{u} \geqslant 0, \tag{1.40}$$

$$\underline{u} \geqslant 0. \tag{1.41}$$

In order to solve the ANSP's problem, first make (1.38) and (1.41) be binding and omit (1.39) and (1.40) and then check the omitted constraints after solving the problem. Thus, we have:

$$\overline{u} = \Delta\theta f D(s) \sigma \ln(1+\underline{r}), \tag{1.42}$$

$$\underline{u} = 0. \tag{1.43}$$

In fact, the maximization of (1.31) gives the same optimal degrees of the service as under complete information, that is, $\overline{r}^{FB}$ and $\underline{r}^{FB}$ determined by (1.27) and (1.29). Moreover, the second-best optimal degrees $\overline{r}^{SB} = \overline{r}^{FB}$ and $\underline{r}^{SB} = \underline{r}^{FB}$ can satisfy the omitted constraints (1.39) and (1.40). According to (1.42) and (1.43), the second-best optimal transfers are:

$$\bar{t}^{SB} = \Pi^*\left(\overline{\theta}, s, \overline{r}^{FB}\right) - \pi^*\left(\overline{\theta}, s\right) - \Delta\theta f D(s) \sigma \ln\left(1+\underline{r}^{FB}\right), \tag{1.44}$$

$$\underline{t}^{SB} = \Pi^*\left(\underline{\theta}, s, \underline{r}^{FB}\right) - \pi^*\left(\underline{\theta}, s\right), \tag{1.45}$$

as long as $\bar{t}^{SB} \geqslant C_{ANSP}\left(s, \underline{r}^{FB}\right)$. Furthermore, as shown in (1.44), the ANSP provides an information rent $\Delta\theta f D(s) \sigma \ln\left(1+\underline{r}^{FB}\right)$ to the airline with $\overline{\theta}$ in order to make it not mimic the other type.

If $\bar{t}^{SB} < C_{ANSP}\left(s, \underline{r}^{FB}\right)$, the ANSP may still propose separating contracts by using public funds or propose a pooling contract without using public funds instead. The separating contracts are better than the pooling contract in terms of efficiency, while the pooling contract saves the cost of public funds. As the optimality between these two types of contracts depends on parameter values, we will not discuss them in details.

1.3.1.4 Scenario 4: $0 < \alpha \leqslant \frac{\xi - \tau + a(s) - z}{D(s)}$ and Low Airline Benefit

In this scenario, both passengers and the airline can benefit from the service and the airline's benefit is lower than the ANSP's cost of providing the service. Therefore, as long as the social benefit of the service outweighs the social cost, it is optimal for the ANSP to use public funds to cover the part of cost which cannot be covered by the airline's transfer.

Because the analysis for $\overline{r}^{FB}$ and $\underline{r}^{FB}$ under complete information is similar to that in Scenario 3, we will only derive optimal contracts under incomplete information.

Incomplete Information Under incomplete information, the ANSP maximizes the expected social welfare, that is:

$$\max_{\left\{\left(\overline{r},\overline{t}\right);\left(\underline{r},\underline{t}\right)\right\}} W = \mu \left\{ PS^* \left(s, \overline{r}\right) + \Pi^* \left(\overline{\theta}, s, \overline{r}\right) - C_{ANSP} \left(s, \overline{r}\right) - \lambda \left[C_{ANSP} \left(s, \overline{r}\right) - \overline{t}\right] \right\}$$
$$+ \left(1 - \mu\right) \left\{ PS^* \left(s, \underline{r}\right) + \Pi^* \left(\underline{\theta}, s, \underline{r}\right) - C_{ANSP} \left(s, \underline{r}\right) - \lambda \left[C_{ANSP} \left(s, \underline{r}\right) - \underline{t}\right] \right\}, \tag{1.46}$$

subject to the airline's incentive compatibility and participation constraints as shown in (1.32) through (1.41).

When solving the problem, we also have (1.42) and (1.43). Next, plugging:

$$\overline{t} = \Pi^* \left(\overline{\theta}, s, \overline{r}\right) - \pi^* \left(\overline{\theta}, s\right) - \Delta\theta f D\left(s\right) \sigma \ln\left(1 + \underline{r}\right), \tag{1.47}$$

$$\underline{t} = \Pi^* \left(\underline{\theta}, s, \underline{r}\right) - \pi^* \left(\underline{\theta}, s\right), \tag{1.48}$$

into (1.46) and taking the first-order condition of (1.46), we can obtain the second-best optimal menu of contracts $\left\{\left(\overline{r}^{SB}, \overline{t}^{SB}\right), \left(\underline{r}^{SB}, \underline{t}^{SB}\right)\right\}$, as shown in (1.49) through (1.52). Again, $\overline{r}^{SB}$ and $\underline{r}^{SB}$ are determined by the intersection of a logarithmic and linear function and can satisfy omitted constraints. Moreover, the second-order condition can be satisfied.

$$\begin{aligned}
\overline{\Omega} = & \left(3+2\lambda\right)\eta\alpha^{2}D\left(s\right)^{2}\sigma^{2}\ln\left(1+\overline{r}^{SB}\right)-\left\{4\left(1+\lambda\right)m\left(s\right)\left(1+\overline{r}^{SB}\right)\right. \\
& -\left(3+2\lambda\right)\eta\left[\xi-\tau+a\left(s\right)-z-\alpha D\left(s\right)\right]\alpha D\left(s\right)\sigma \\
& \left.-4\left(1+\lambda\right)\overline{\theta}fD\left(s\right)\sigma\right\}=0, \qquad (1.49) \\
\overline{t}^{SB} = & \frac{1}{4}\eta\alpha^{2}D\left(s\right)^{2}\sigma^{2}\left[\ln\left(1+\overline{r}^{SB}\right)\right]^{2}+\left\{\frac{1}{2}\eta\left[\xi-\tau+a\left(s\right)-z-\alpha D\left(s\right)\right]\alpha\right. \\
& \left.+\overline{\theta}f\right\}D\left(s\right)\sigma\ln\left(1+\overline{r}^{SB}\right)-\Delta\theta fD\left(s\right)\sigma\ln\left(1+\underline{r}^{SB}\right), \qquad (1.50) \\
\underline{\Omega} = & \left(3+2\lambda\right)\eta\alpha^{2}D\left(s\right)^{2}\sigma^{2}\ln\left(1+\underline{r}^{SB}\right)-\left\{4\left(1+\lambda\right)m\left(s\right)\left(1+\underline{r}^{SB}\right)\right. \\
& -\left(3+2\lambda\right)\eta\left[\xi-\tau+a\left(s\right)-z-\alpha D\left(s\right)\right]\alpha D\left(s\right)\sigma \\
& \left.-4\left(1+\lambda\right)\underline{\theta}fD\left(s\right)\sigma+4\frac{\mu}{1-\mu}\lambda\Delta\theta fD\left(s\right)\sigma\right\}=0, \qquad (1.51) \\
\underline{t}^{SB} = & \frac{1}{4}\eta\alpha^{2}D\left(s\right)^{2}\sigma^{2}\left[\ln\left(1+\underline{r}^{SB}\right)\right]^{2}+\left\{\frac{1}{2}\eta\left[\xi-\tau+a\left(s\right)-z-\alpha D\left(s\right)\right]\alpha\right. \\
& \left.+\underline{\theta}f\right\}D\left(s\right)\sigma\ln\left(1+\underline{r}^{SB}\right). \qquad (1.52)
\end{aligned}$$

According to (1.49) and (1.51), for the optimal degrees of the service, we can obtain $\overline{r}^{SB}=\overline{r}^{FB}$, that is, there is no distortion for the airline with $\overline{\theta}$, while $\underline{r}^{SB}<\underline{r}^{FB}$, that is, there is a downward distortion for the one with $\underline{\theta}$. Here, we also have $\overline{r}^{SB}>\underline{r}^{SB}$. Moreover, only the airline with $\overline{\theta}$ can get a positive information rent $\Delta\theta fD\left(s\right)\sigma\ln\left(1+\underline{r}^{SB}\right)$. In fact, under incomplete information, the optimal degree for the airline with $\underline{\theta}$ is distorted downwards to decrease the information rent of the airline with $\overline{\theta}$, which reflects the trade-off between efficiency and rent extraction. The graphical comparison of optimal degrees between complete and incomplete information is in Section 1.7.1.2.

1.3.2 Profit-Maximizing ANSP

The profit-maximizing ANSP's objective is to maximize the difference between the transfer and the cost of providing the service, that is:

$$\max_{\{(r,t)\}}H=t-C_{ANSP}\left(s,r\right). \qquad (1.53)$$

Next, we will derive optimal contracts under $0 < \alpha \leqslant \frac{\xi-\tau+a(s)-z}{D(s)}$.

Complete Information Under complete information, the ANSP will set the transfer as the airline's benefit from the service, that is, $t = \Pi^*(\theta, s, r) - \pi^*(\theta, s)$. Thus, the optimization problem of the ANSP is:

$$\max_r H = \Pi^*(\theta, s, r) - \pi^*(\theta, s) - C_{ANSP}(s, r). \tag{1.54}$$

Then, by taking the first-order condition of (1.54), we can obtain that the first-best optimal contracts for the airline with $\overline{\theta}$ and $\underline{\theta}$ are $\left\{\left(\overline{r}^{FB}, \overline{t}^{FB}\right)\right\}$ and $\left\{\left(\underline{r}^{FB}, \underline{t}^{FB}\right)\right\}$, respectively, as shown in (1.55) through (1.58). Again, $\overline{r}^{FB}$ and $\underline{r}^{FB}$ are determined by the intersection of a logarithmic and linear function, and because $\overline{\theta} > \underline{\theta}$, we have $\overline{r}^{FB} > \underline{r}^{FB}$. Moreover, the second-order condition can be satisfied.

$$\begin{aligned}
\overline{\Omega} &= \eta\alpha^2 D(s)^2 \sigma^2 \ln\left(1+\overline{r}^{FB}\right) - \left\{2m(s)\left(1+\overline{r}^{FB}\right)\right. \\
&\quad \left. -\eta\left[\xi-\tau+a(s)-z-\alpha D(s)\right]\alpha D(s)\sigma - 2\overline{\theta} f D(s)\sigma\right\} = 0, &(1.55)\\
\overline{t}^{FB} &= \frac{1}{4}\eta\alpha^2 D(s)^2\sigma^2\left[\ln\left(1+\overline{r}^{FB}\right)\right]^2 + \left\{\frac{1}{2}\eta\left[\xi-\tau+a(s)-z-\alpha D(s)\right]\alpha\right. \\
&\quad \left. +\overline{\theta} f\right\} D(s)\sigma \ln\left(1+\overline{r}^{FB}\right), &(1.56)\\
\underline{\Omega} &= \eta\alpha^2 D(s)^2 \sigma^2 \ln\left(1+\underline{r}^{FB}\right) - \left\{2m(s)\left(1+\underline{r}^{FB}\right)\right. \\
&\quad \left. -\eta\left[\xi-\tau+a(s)-z-\alpha D(s)\right]\alpha D(s)\sigma - 2\underline{\theta} f D(s)\sigma\right\} = 0, &(1.57)\\
\underline{t}^{FB} &= \frac{1}{4}\eta\alpha^2 D(s)^2\sigma^2\left[\ln\left(1+\underline{r}^{FB}\right)\right]^2 + \left\{\frac{1}{2}\eta\left[\xi-\tau+a(s)-z-\alpha D(s)\right]\alpha\right. \\
&\quad \left. +\underline{\theta} f\right\} D(s)\sigma \ln\left(1+\underline{r}^{FB}\right). &(1.58)
\end{aligned}$$

Incomplete Information Under incomplete information, the ANSP maximizes the expected profit, that is:

$$\max_{\left\{\left(\overline{r},\overline{t}\right);\left(\underline{r},\underline{t}\right)\right\}} H = \mu\left[\overline{t} - C_{ANSP}(s, \overline{r})\right] + (1-\mu)\left[\underline{t} - C_{ANSP}(s, \underline{r})\right], \tag{1.59}$$

subject to the airline's incentive compatibility and participation constraints as shown in (1.32) through (1.41).

Then, by taking the first-order condition of (1.59), we can obtain the second-best optimal menu of contracts $\left\{\left(\overline{r}^{SB},\overline{t}^{SB}\right),\left(\underline{r}^{SB},\underline{t}^{SB}\right)\right\}$, as shown in (1.60) through (1.63). Again, $\overline{r}^{SB}$ and $\underline{r}^{SB}$ are determined by the intersection of a logarithmic and linear function and can satisfy omitted constraints. Moreover, the second-order condition can be satisfied.

$$\begin{aligned}
\overline{\Omega} &= \eta\alpha^2 D(s)^2\sigma^2\ln\left(1+\overline{r}^{SB}\right)-\left\{2m(s)\left(1+\overline{r}^{SB}\right)\right.\\
&\left.-\eta\left[\xi-\tau+a(s)-z-\alpha D(s)\right]\alpha D(s)\sigma-2\overline{\theta}fD(s)\sigma\right\}=0, \qquad (1.60)\\
\overline{t}^{SB} &= \frac{1}{4}\eta\alpha^2 D(s)^2\sigma^2\left[\ln\left(1+\overline{r}^{SB}\right)\right]^2+\left\{\frac{1}{2}\eta\left[\xi-\tau+a(s)-z-\alpha D(s)\right]\alpha\right.\\
&\left.+\overline{\theta}f\right\}D(s)\sigma\ln\left(1+\overline{r}^{SB}\right)-\Delta\theta fD(s)\sigma\ln\left(1+\underline{r}^{SB}\right), \qquad (1.61)\\
\underline{\Omega} &= \eta\alpha^2 D(s)^2\sigma^2\ln\left(1+\underline{r}^{SB}\right)-\left\{2m(s)\left(1+\underline{r}^{SB}\right)\right.\\
&-\eta\left[\xi-\tau+a(s)-z-\alpha D(s)\right]\alpha D(s)\sigma-2\underline{\theta}fD(s)\sigma\\
&\left.+2\frac{\mu}{1-\mu}\Delta\theta fD(s)\sigma\right\}=0, \qquad (1.62)\\
\underline{t}^{SB} &= \frac{1}{4}\eta\alpha^2 D(s)^2\sigma^2\left[\ln\left(1+\underline{r}^{SB}\right)\right]^2+\left\{\frac{1}{2}\eta\left[\xi-\tau+a(s)-z-\alpha D(s)\right]\alpha\right.\\
&\left.+\underline{\theta}f\right\}D(s)\sigma\ln\left(1+\underline{r}^{SB}\right). \qquad (1.63)
\end{aligned}$$

According to (1.60) and (1.62), for the optimal degrees of the service, we can obtain $\overline{r}^{SB}=\overline{r}^{FB}$, that is, there is no distortion for the airline with $\overline{\theta}$, while $\underline{r}^{SB}<\underline{r}^{FB}$, that is, there is a downward distortion for the one with $\underline{\theta}$. Here, we also have $\overline{r}^{SB}>\underline{r}^{SB}$. Moreover, only the airline with $\overline{\theta}$ can get a positive information rent $\Delta\theta fD(s)\sigma\ln\left(1+\underline{r}^{SB}\right)$.

1.4 Adjustments of Optimal Contracts

Because contracts should be adjusted over time according to the evolution of some relevant exogenous variables, we will study the effects of safety standard and flight frequency on optimal contracts. Moreover, we will analyze only the adjustments of optimal contracts under incomplete information in Scenario 4 of welfare-maximizing ANSP. For other optimal contracts, the analysis is similar and thus is omitted hereafter.

When $0 < \alpha \leqslant \frac{\xi-\tau+a(s)-z}{D(s)}$, we do not have explicit solutions for the optimal degree of the service. Thus, we will use derivatives of implicit functions. Specifically, for a variable x_l $(l = 1, 2, \cdots, L)$, we have:

$$\frac{\partial r}{\partial x_l} = -\frac{\partial \Omega / \partial x_l}{\partial \Omega / \partial r}. \tag{1.64}$$

Because:

$$\frac{\partial \Omega}{\partial r} = \text{slope}\,(LOGR) - \text{slope}\,(LR) < 0, \tag{1.65}$$

holds for any optimal degree, the sign of $\frac{\partial r}{\partial x_l}$ is the same as that of $\frac{\partial \Omega}{\partial x_l}$. That is, in order to see the effect of a variable on the optimal degree, we just need to study its effect on the implicit function determining the optimal degree.

1.4.1 Effect of Safety Standard

Undoubtedly, safety is the highest priority in air transport sector and the safety standard always becomes higher. Thus, it is worthy to study how the improvement of safety standard affects optimal degrees. We first give a definition.

Definition 1.1. *The safety elasticity of delays and the safety elasticity of cost are defined as, respectively:*

$$\varepsilon_{gs} \equiv \frac{dg\,(s)}{g\,(s)}\frac{s}{ds}, \tag{1.66}$$

$$\varepsilon_{ms} \equiv \frac{dm\,(s)}{m\,(s)}\frac{s}{ds}. \tag{1.67}$$

The safety elasticity of delays (resp. cost) measures the percentage change in delays caused by an exceptional event (the marginal cost of the service) in response to a one percent change in safety standard.

Let $\overline{\varepsilon}^{SB}\left(\overline{r}^{SB}\right)$ and $\underline{\varepsilon}^{SB}\left(\underline{r}^{SB}\right)$ denote two thresholds, in which:

$$\overline{\varepsilon}^{SB}\left(\overline{r}^{SB}\right) \equiv -\frac{(3+2\lambda)\,\eta\alpha D\,(s)\,\sigma s}{2\,(1+\lambda)\left(1+\overline{r}^{SB}\right)m\,(s)}\frac{\partial \overline{V}^{*}\left(s,\overline{r}^{SB}\right)}{\partial s}, \tag{1.68}$$

$$\underline{\varepsilon}^{SB}\left(\underline{r}^{SB}\right) \equiv -\frac{(3+2\lambda)\,\eta\alpha D\,(s)\,\sigma s}{2\,(1+\lambda)\left(1+\underline{r}^{SB}\right)m\,(s)}\frac{\partial \underline{V}^{*}\left(s,\underline{r}^{SB}\right)}{\partial s}. \tag{1.69}$$

Then, we have Proposition 1.1.

Proposition 1.1.

1. *$\overline{r}^{SB}$ increases with s if and only if $\varepsilon_{gs}-\varepsilon_{ms} \geqslant \overline{\varepsilon}^{SB}\left(\overline{r}^{SB}\right)$.*
2. *$\underline{r}^{SB}$ increases with s if and only if $\varepsilon_{gs}-\varepsilon_{ms} \geqslant \underline{\varepsilon}^{SB}\left(\underline{r}^{SB}\right)$.*

Proof. We first consider the effect of s on $\overline{r}^{SB}$. Taking the derivative of (1.49) with respect to s, we can obtain:

$$\begin{aligned}\frac{\partial\overline{\Omega}}{\partial s} &= \left\{2\,(3+2\lambda)\,\eta\alpha^{2}D\,(s)\,\sigma^{2}\ln\left(1+\overline{r}^{SB}\right)-(3+2\lambda)\,\eta\alpha^{2}D\,(s)\,\sigma\right.\\ &\quad\left.+(3+2\lambda)\,\eta\left[\xi-\tau+a\,(s)-z-\alpha D\,(s)\right]\alpha\sigma+4\,(1+\lambda)\,\overline{\theta}f\sigma\right\}\\ &\quad\cdot 2\left[\sum_{k=1}^{+\infty}\frac{\left(\frac{\beta T}{f}\right)^{k}e^{-\left(\frac{\beta T}{f}\right)}}{(k-1)!}+\gamma\beta\frac{f}{T}\right]g'\,(s)\\ &\quad+(3+2\lambda)\,\eta\alpha D\,(s)\,\sigma a'\,(s)-4\,(1+\lambda)\left(1+\overline{r}^{SB}\right)m'\,(s)\,. \end{aligned}\tag{1.70}$$

By using (1.66), (1.67) and (1.49), (1.70) becomes:

$$\begin{aligned}\frac{\partial\overline{\Omega}}{\partial s} &= \frac{1}{s}\left\{4\,(1+\lambda)\left(1+\overline{r}^{SB}\right)m\,(s)\,(\varepsilon_{gs}-\varepsilon_{ms})\right.\\ &\quad\left.+(3+2\lambda)\,\eta\alpha D\,(s)\,\sigma\left\{a'\,(s)\,s-\alpha D\,(s)\left[1-\sigma\ln\left(1+\overline{r}^{SB}\right)\right]\varepsilon_{gs}\right\}\right\}. \end{aligned}\tag{1.71}$$

Next, introducing the direct effect of the improvement of safety standard on passenger utility, that is:

$$\begin{aligned}\frac{\partial\overline{V}^{*}\left(s,\overline{r}^{SB}\right)}{\partial s} &= \frac{1}{2}\left\{a'\,(s)-2\alpha\left[\sum_{k=1}^{+\infty}\frac{\left(\frac{\beta T}{f}\right)^{k}e^{-\left(\frac{\beta T}{f}\right)}}{(k-1)!}+\gamma\beta\frac{f}{T}\right]\left[1-\sigma\ln\left(1+\overline{r}^{SB}\right)\right]g'\,(s)\right\}\\ &= \frac{1}{2s}\left\{a'\,(s)\,s-\alpha D\,(s)\left[1-\sigma\ln\left(1+\overline{r}^{SB}\right)\right]\varepsilon_{gs}\right\}, \end{aligned}\tag{1.72}$$

and plugging (1.72) into (1.71), we can obtain the first point of Proposition 1.1.

Moreover, we can obtain the second point analogously. □

According to Proposition 1.1, the optimal degree of the service increases with safety standard if and only if the difference between the safety elasticity of delays and cost is greater than a threshold, which is a function of the direct effect of the improvement of safety standard on passenger utility. Moreover, if passengers can directly benefit from a higher safety standard, the threshold will be negative.

To illustrate the first point of Proposition 1.1, we should analyze the effects of safety standard in (1.70). Specifically, the improvement of safety standard implies longer delays caused by an exceptional event, a higher passengers' utility gain and a higher marginal cost of the service. The first term in (1.70) shows a direct and an indirect effect of longer delays caused by an exceptional event on the degree. On the one hand, longer delays will increase the marginal benefit of the service to society and thus give the ANSP a direct incentive to increase the degree. ε_{gs} in the first point represents part of this direct effect. On the other hand, longer delays will decrease the air traffic, which implies a lower marginal benefit of the service to society, and thus give the ANSP an indirect incentive to decrease the degree. In the second term, the higher utility gain will increase the air traffic, which implies a higher marginal benefit of the service to society, and thus gives the ANSP an indirect incentive to increase the degree. In addition, in the third term, the higher marginal cost of the service gives the ANSP a direct incentive to decrease the degree. ε_{ms} in the first point represents this direct effect. Finally, the condition in the first point is the synthesis of the effects above, which, more precisely, is about whether or not the effects conducive to the increase of the degree can dominate the others.

Moreover, we can illustrate the second point analogously, except considering another effect in terms of the information rent. Specifically, longer delays caused by an exceptional event will increase the information rent of the airline with $\overline{\theta}$, which is a function of $\underline{r}^{SB}$, and thus give the ANSP an indirect incentive to decrease $\underline{r}^{SB}$.

Next, as we discuss a lot about the effects of the improvement of safety standard, through longer delays caused by an exceptional event, we extend our analysis to study how longer expected delays per flight affect optimal degrees. Let $\overline{\alpha}^{SB}\left(\overline{r}^{SB}\right)$ and $\underline{\alpha}^{SB}\left(\underline{r}^{SB}\right)$

denote two thresholds, in which:

$$\overline{\alpha}^{SB}\left(\overline{r}^{SB}\right) \equiv \frac{2}{D(s)}\sqrt{\frac{(1+\lambda)\, m(s)\left(1+\overline{r}^{SB}\right)}{(3+2\lambda)\,\eta\sigma\left[1-\sigma\ln\left(1+\overline{r}^{SB}\right)\right]}}, \tag{1.73}$$

$$\underline{\alpha}^{SB}\left(\underline{r}^{SB}\right) \equiv \frac{2}{D(s)}\sqrt{\frac{(1+\lambda)\, m(s)\left(1+\underline{r}^{SB}\right)}{(3+2\lambda)\,\eta\sigma\left[1-\sigma\ln\left(1+\underline{r}^{SB}\right)\right]}}. \tag{1.74}$$

Then, we have Proposition 1.2.

Proposition 1.2.

1. *$\overline{r}^{SB}$ increases with $D(s)$ if and only if $\alpha \leqslant \overline{\alpha}^{SB}\left(\overline{r}^{SB}\right)$.*
2. *$\underline{r}^{SB}$ increases with $D(s)$ if and only if $\alpha \leqslant \underline{\alpha}^{SB}\left(\underline{r}^{SB}\right)$.*

Proof. In Section 1.7.2. □

According to Proposition 1.2, the optimal degree increases with expected delays per flight if and only if the passengers' value of time is lower than a threshold.

As shown in the analysis of Proposition 1.1, longer delays affect optimal degrees mainly through a direct effect (higher marginal benefit of the service to society) and an indirect effect (less traffic). Proposition 1.2 tells us that, if the passengers' value of time is relatively low, the direct effect will dominate the indirect one, and then optimal degrees will increase. Otherwise, the outside option will be more valuable for passengers. Thus, the direct effect will be dominated by the indirect one, and then optimal degrees will decrease.

Moreover, as $\overline{\alpha}^{SB}\left(\overline{r}^{SB}\right) \geqslant \underline{\alpha}^{SB}\left(\underline{r}^{SB}\right)$, it is possible that when delays become longer, the ANSP should adjust optimal degrees in opposite directions. Considering also $\overline{\alpha}^{FB}\left(\overline{r}^{FB}\right) \geqslant \underline{\alpha}^{FB}\left(\underline{r}^{FB}\right)$ under complete information, we have Corollary 1.1.

Corollary 1.1. *When $D(s)$ becomes longer, if $\underline{\alpha}^{FB}\left(\underline{r}^{FB}\right) \leqslant \alpha \leqslant \overline{\alpha}^{FB}\left(\overline{r}^{FB}\right)$ (resp. $\underline{\alpha}^{SB}\left(\underline{r}^{SB}\right) \leqslant \alpha \leqslant \overline{\alpha}^{SB}\left(\overline{r}^{SB}\right)$), optimal degrees under complete (resp. incomplete) information will move in opposite directions. Moreover, $\overline{\alpha}^{SB}\left(\overline{r}^{SB}\right) - \underline{\alpha}^{SB}\left(\underline{r}^{SB}\right) \geqslant \overline{\alpha}^{FB}\left(\overline{r}^{FB}\right) - \underline{\alpha}^{FB}\left(\underline{r}^{FB}\right)$ implies that the existence of information rent increases the possibility that optimal degrees move in opposite directions.*

1.4.2 Effect of Flight Frequency

In this model, because of the slot control in Europe, we assume that flight frequency is not a endogenous decision variable of the airline. Here, we study how the change of flight frequency affects optimal degrees.

We first introduce some notations. According to Proposition 1.2, $\frac{\partial \overline{r}^{SB}}{\partial D(s)} \geqslant 0$ if and only if $\Phi \geqslant 0$ and $\frac{\partial \underline{r}^{SB}}{\partial D(s)} \geqslant 0$ if and only if $\Psi \geqslant 0$, in which:

$$\Phi \equiv 4\left(1+\lambda\right) m\left(s\right)\left(1+\overline{r}^{SB}\right) - \left(3+2\lambda\right)\eta\alpha^2 D\left(s\right)^2 \sigma\left[1-\sigma\ln\left(1+\overline{r}^{SB}\right)\right], \tag{1.75}$$

$$\Psi \equiv 4\left(1+\lambda\right) m\left(s\right)\left(1+\underline{r}^{SB}\right) - \left(3+2\lambda\right)\eta\alpha^2 D\left(s\right)^2 \sigma\left[1-\sigma\ln\left(1+\underline{r}^{SB}\right)\right]. \tag{1.76}$$

Moreover, let $\overline{\Gamma}_f$ and $\underline{\Gamma}_f$ denote two thresholds, in which:

$$\overline{\Gamma}_f \equiv \Gamma - 2\left(1+\lambda\right)\overline{\theta} D\left(s\right)^2 \sigma \frac{1}{\Phi}\frac{T}{\beta g\left(s\right)}, \tag{1.77}$$

$$\underline{\Gamma}_f \equiv \Gamma - 2\left[\left(1+\lambda\right)\underline{\theta} - \frac{\mu}{1-\mu}\lambda\Delta\theta\right] D\left(s\right)^2 \sigma \frac{1}{\Psi}\frac{T}{\beta g\left(s\right)}, \tag{1.78}$$

$$\Gamma \equiv \sum_{k=1}^{+\infty} \frac{\left(\frac{\beta T}{f}\right)^{k-1} e^{-\left(\frac{\beta T}{f}\right)}\left(k-\frac{\beta T}{f}\right)\frac{T^2}{f^2}}{(k-1)!}. \tag{1.79}$$

In fact, Γ is a threshold such that $\frac{\partial D(s)}{\partial f} \geqslant 0$ if and only if $\gamma \geqslant \Gamma$. Then, we have Proposition 1.3.

Proposition 1.3.

1. *When* $\Phi \geqslant (resp. \ <) \, 0$, $\overline{r}^{SB}$ *increases with* f *if and only if* $\gamma \geqslant (resp. \ \leqslant) \max\left\{0, \overline{\Gamma}_f\right\}$.
2. *When* $\Psi \geqslant (resp. \ <) \, 0$, $\underline{r}^{SB}$ *increases with* f *if and only if* $\gamma \geqslant (resp. \ \leqslant) \max\left\{0, \underline{\Gamma}_f\right\}$.

Proof. We first consider the effect of f on $\overline{r}^{SB}$. Taking the derivative of (1.49) with respect to f, we can obtain:

$$\frac{\partial\overline{\Omega}}{\partial f} = \left[2\left(3+2\lambda\right)\eta\alpha^2 D\left(s\right)\sigma^2\ln\left(1+\overline{r}^{SB}\right) - \left(3+2\lambda\right)\eta\alpha^2 D\left(s\right)\sigma\right.$$
$$\left. + \left(3+2\lambda\right)\eta\left(\xi-\tau+a\left(s\right)-z-\alpha D\left(s\right)\right)\alpha\sigma + 4\left(1+\lambda\right)\overline{\theta} f\sigma\right]$$

$$\cdot 2\left[-\sum_{k=1}^{+\infty}\frac{\left(\frac{\beta T}{f}\right)^{k}e^{-\left(\frac{\beta T}{f}\right)}\left(k-\frac{\beta T}{f}\right)\frac{1}{f}}{(k-1)!}g\left(s\right)+\gamma\beta\frac{1}{T}g\left(s\right)\right]+4\left(1+\lambda\right)\overline{\theta}D\left(s\right)\sigma. \tag{1.80}$$

By using (1.49), (1.80) becomes:

$$\frac{\partial\overline{\Omega}}{\partial f}=\frac{2}{D\left(s\right)}\left\{\left[-\sum_{k=1}^{+\infty}\frac{\left(\frac{\beta T}{f}\right)^{k}e^{-\left(\frac{\beta T}{f}\right)}\left(k-\frac{\beta T}{f}\right)\frac{1}{f}}{(k-1)!}g\left(s\right)+\gamma\beta\frac{1}{T}g\left(s\right)\right]\Phi+2\left(1+\lambda\right)\overline{\theta}D\left(s\right)^{2}\sigma\right\}. \tag{1.81}$$

Next, using (1.77) and (1.79), we can obtain the first point of Proposition 1.3.

Moreover, we can obtain the second point analogously. □

According to Proposition 1.3, when the optimal degree of the service increases with expected delays per flight, it will increase with flight frequency if and only if the delay externality parameter is greater than a threshold. However, when the optimal degree decreases with expected delays per flight, it will increase with flight frequency if and only if the delay externality parameter is less than a threshold.

To illustrate the first point of Proposition 1.3, we should analyze the effects of flight frequency in (1.80). Specifically, the increase of flight frequency implies shorter delays due to exceptional events in own slot, longer delays induced by other flights and a higher supply side delay reduction benefit of the airline. In (1.80) The first term shows the change of delays and the second term shows the change of supply side delay reduction benefit.

First consider the case $\Phi \geqslant 0$. In this case, we have $\Gamma \geqslant \overline{\Gamma}_f$. If the externality of delays between flights is significant, that is, $\gamma \geqslant \Gamma$, the delays induced by other flights will dominate the delays due to exceptional events in own slot. Thus, expected delays per flight become longer when flight frequency increases. Given $\Phi \geqslant 0$, that is, the direct effect of longer delays (higher marginal benefit of the service to society) dominates the indirect effect of longer delays (less traffic), the net effect shown in the first term in (1.80) is to increase $\overline{r}^{SB}$ when flight frequency increases. Considering also the higher supply side delay reduction benefit of the airline shown in the second term in (1.80), $\overline{r}^{SB}$ will increase with flight frequency.

However, if the externality of delays between flights is not significant, that is, $\gamma < \Gamma$, the delays induced by other flights will be dominated by the delays due to exceptional events in own slot. Thus, expected delays per flight become shorter when flight frequency increases, and then the net effect shown in the first term in (1.80) is to decrease $\overline{r}^{SB}$ when flight frequency increases. Next, if $\max\left\{0, \overline{\Gamma}_f\right\} \leqslant \gamma < \Gamma$, that is, the externality of delays between flights is not very insignificant, the effect from the delays due to exceptional events in own slot will be relatively weak, compared with the higher supply side delay reduction benefit of the airline shown in the second term in (1.80). Then, $\overline{r}^{SB}$ will still increase with flight frequency. Nevertheless, if $\gamma < \max\left\{0, \overline{\Gamma}_f\right\}$, the effect from the delays due to exceptional events in own slot will be strong enough. Then, $\overline{r}^{SB}$ will decrease with flight frequency. To summarize, when $\Phi \geqslant 0$, $\overline{r}^{SB}$ will increase with flight frequency if and only if $\gamma \geqslant \max\left\{0, \overline{\Gamma}_f\right\}$. Moreover, we can analyze the case $\Phi < 0$ analogously.

In fact, the analysis for the second point of Proposition 1.3 is similar as above, except that we should also consider the effect of flight frequency on the information rent of the airline with $\overline{\theta}$.

1.5 Use of Public Funds

For a welfare-maximizing ANSP, when the airline's benefit is higher than the ANSP's cost of providing the service (Scenario 3), the ANSP may not have to use public funds. However, when the airline's benefit is lower than the ANSP's cost of providing the service (Scenario 4), the ANSP has to use public funds. Therefore, in this section, by choosing proper parameter values[9] and function specifications, we use numerical examples to study when a welfare-maximizing ANSP has to use public funds to provide the delay reduction service.

Specifically, in all numerical examples, we use $a(s) = 0.5 * \frac{\ln s}{\ln 2}$, $g(s) = 0.01 * 2^s$ and

[9] Parameter values used in numerical examples can satisfy all second-order conditions. They can also ensure that the social values of the service are non-negative, that is, $[PS^*(s,r) - ps^*(s)] + [\Pi^*(\theta,s,r) - \pi^*(\theta,s)] - C_{ANSP}(s,r) - \lambda[C_{ANSP}(s,r) - t]\mathbb{1}_{t<C(s,r)} \geqslant 0$.

$m(s) = 0.06 + 0.01s$. As we can see, $a(s)$ is a concave function of s; $g(s)$ is a convex function of s; and $m(s)$ is a linear function of s. Moreover, we use $\beta = 0.01$, $\gamma = 120$, $\theta = 1$, $\lambda = 0.04$, $N = 2$, $\xi = 3$, $\zeta = 1$, $s = 2$, $\tau = 0.8$, $T = 1.5$, $f = 1$, and $z = 2$.

Consider first the airline's net benefit from the service:

$$\omega = \Pi^*(\theta, s, r) - \pi^*(\theta, s) - C_{ANSP}(s, r), \tag{1.82}$$

which is the net benefit of the airline when it is asked to pay the total cost of providing the service. By using a large number of numerical examples, we can find that, for each set of parameter values, there exists a threshold $\widehat{r}$ such that $\omega \geqslant 0$ if and only if $r \leqslant \widehat{r}$. Thus, as long as $r > \widehat{r}$, the ANSP has to use public funds to provide the service. For example, using $\alpha = 0.35$ and $\sigma = \frac{1}{\ln 2}$, we can obtain Figure 1.2[10].

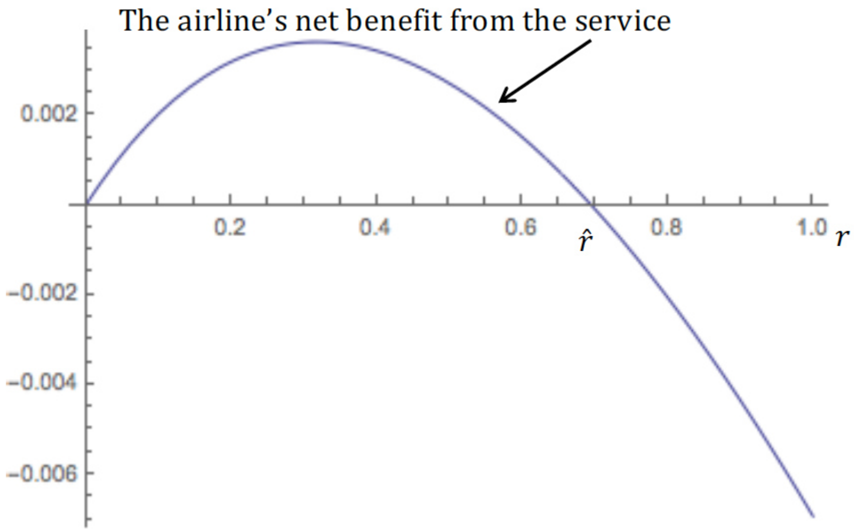

Figure 1.2: The Airline's Net Benefit from the Service and $\widehat{r}$

Denote the marginal benefit of the service to society by MB and the marginal cost of the service to ANSP by MC. If $MB(\widehat{r}) \leqslant MC$, in order to make MB equal to MC, the ANSP will decrease the degree, and then the optimal degree will be r^- with $r^- \leqslant \widehat{r}$. When $r = r^-$, the airline's benefit is higher than the ANSP's cost of providing the service, implying $\omega \geqslant 0$, and thus the ANSP may not have to use public funds.

[10]Note that the curve in Figure 1.2 seems smooth, while it is in fact not. Because that curve only fluctuates in extremely small intervals, we can regard it as a smooth one. Moreover, because of the fluctuation of the curve, some values smaller than $\widehat{r}$ may also make $u = 0$. However, we can omit them because they are extremely close to $\widehat{r}$.

However, if $MB(\widehat{r}) > MC$, in order to make MB equal to MC, the ANSP will increase the degree, and then the optimal degree will be r^+ with $r^+ > \widehat{r}$. When $r = r^+$, the airline's benefit is lower than the ANSP's cost of providing the service, implying $\omega < 0$, and thus the ANSP has to use public funds.

Consider next two important parameters, that is, the passengers' value of time α and the effectiveness of the service σ. Here, we use the following values of α and σ: $\alpha = 0.05\widetilde{\alpha}$ with $\widetilde{\alpha} \in [0, 50] \cap \mathbb{Z}$ and $\sigma = 0.05 * \frac{\widetilde{\sigma}}{\ln 2}$ with $\widetilde{\sigma} \in [9, 20] \cap \mathbb{Z}$. For each set of parameter values, we calculate $\widehat{r}$. Then, we calculate $MB(\widehat{r}) - MC$. If the difference is positive, the ANSP has to use public funds. Otherwise, the ANSP may not have to use public funds. Calculation results are shown in Figure 1.3.

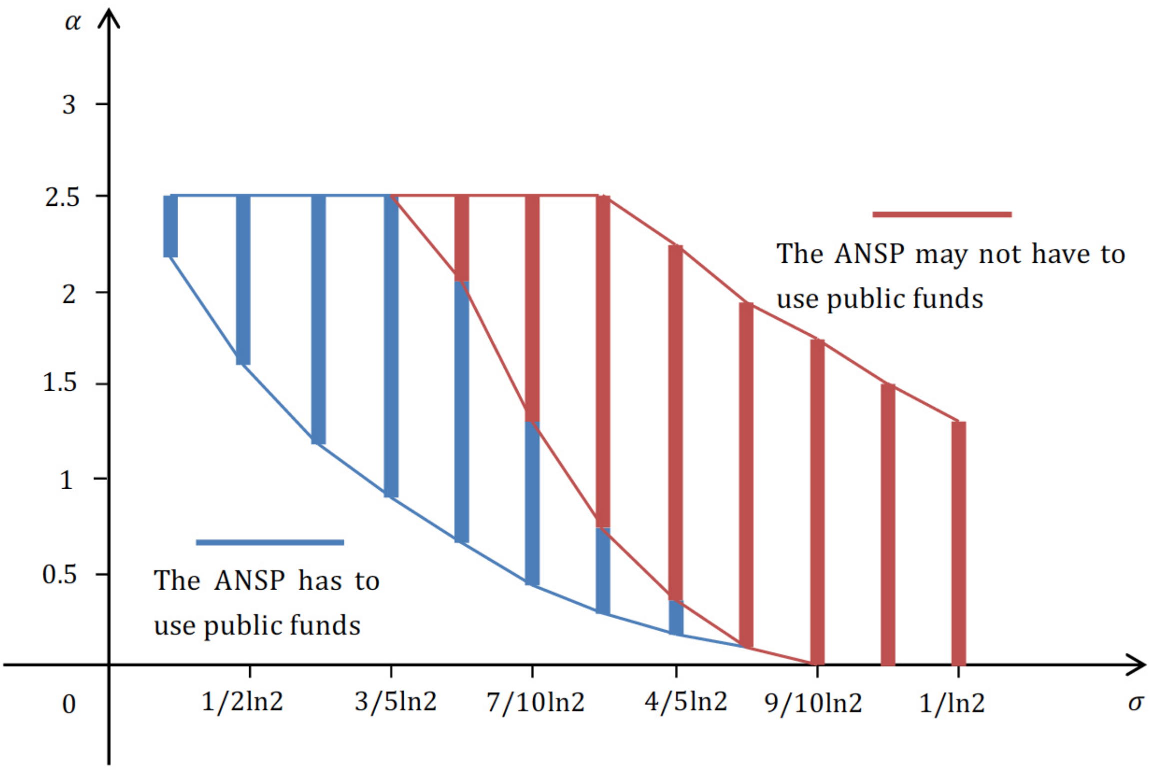

Figure 1.3: α, σ and Use of Public Funds

In Figure 1.3, we find that, if σ is high, that is, the service is very effective, the airline can always obtain a high benefit from the service. Thus, the ANSP may not have to use public funds. If σ is low, that is, the service is very ineffective, the airline can only obtain a very limited benefit from the service. Thus, the ANSP has to use public funds. Finally, if σ falls into an intermediate interval, that is, the effectiveness of the service is intermediate, when the effectiveness decreases, the ANSP may not have to use public funds only when the passengers' value of time α becomes higher. Because the passengers' value of time is positively related to the total benefit of the service to society, a higher α can compensate the loss of benefits resulting from the decrease of the effectiveness of the service.

1.6 Conclusion

In the context of SESAR project, an ANSP can provide a delay reduction service to airlines. Thus, this chapter studies the optimal design of delay reduction contract signed between an ANSP and a monopoly airline. In the contract design, the main issue is to address the adverse selection problem, which comes from airlines' private information about their values of time. Then, we derive optimal contracts analytically considering both the welfare-maximizing and profit-maximizing ANSP, in which we find that, under incomplete information, the optimal degree of the service for the airline with a low value of time may be distorted downwards. Moreover, we conduct comparative-static analysis to study how the changes of safety standard and flight frequency affect optimal contracts. Besides, we use numerical examples to study when a welfare-maximizing ANSP has to use public funds to provide the service.

This chapter focuses on a monopoly airline market structure. Thus, an extension to oligopoly airline market structure and a study for strategic interactions between airlines will be a direction of future research. Moreover, in this chapter, passengers only make a single trip and the passengers' demand is inelastic. Thus, multiple trips and elastic demand are also possible extensions.

1.7 Appendix

1.7.1 Optimal Contracts

1.7.1.1 Second-Order Condition in Complete Information in Scenario 3

In (1.27), the first term of $\overline{\Omega}$ is a logarithmic function of $\overline{r}^{FB}$ and we denote it by $\overline{LOGR}$; and the second term of $\overline{\Omega}$ is a linear function of $\overline{r}^{FB}$ and we denote it by $\overline{LR}$. Analogously, in (1.29), the first term of $\underline{\Omega}$ is a logarithmic function of $\underline{r}^{FB}$ and we denote it by $\underline{LOGR}$; and the second term of $\underline{\Omega}$ is a linear function of $\underline{r}^{FB}$ and we denote it by $\underline{LR}$.

Next, we try to confirm the second-order condition of (1.26) according to the slope of the logarithmic and linear functions and the intercept of the linear function on horizontal axis. Moreover, let I_1 and I_2 denote two expressions indicating the signs of the intercepts, in which:

$$I_1 = 4\left[\overline{\theta} f D\left(s\right)\sigma - m\left(s\right)\right] + 3\eta\left[\xi - \tau + a\left(s\right) - z - \alpha D\left(s\right)\right]\alpha D\left(s\right)\sigma, \tag{1.83}$$

$$I_2 = 4\left[\underline{\theta} f D\left(s\right)\sigma - m\left(s\right)\right] + 3\eta\left[\xi - \tau + a\left(s\right) - z - \alpha D\left(s\right)\right]\alpha D\left(s\right)\sigma. \tag{1.84}$$

Case 1: $0 < \alpha \leqslant \frac{2}{D(s)\sigma}\sqrt{\frac{m(s)}{3\eta}}$ and $I_2 \geqslant 0$

This case is shown in Figure 1.4. $0 < \alpha \leqslant \frac{2}{D(s)\sigma}\sqrt{\frac{m(s)}{3\eta}}$ is equivalent to $0 < 3\eta\alpha^2 D\left(s\right)^2\sigma^2 \leqslant 4m\left(s\right)$ and implies that the slope of $\overline{LR}$ (resp. $\underline{LR}$) is greater than that of $\overline{LOGR}$ (resp. $\underline{LOGR}$) for any r. Moreover, $I_2 \geqslant 0$ implies that the signs of $\overline{LR}$'s and $\underline{LR}$'s intercepts on horizontal axis are positive.

Because $I_2 \geqslant 0$ and $\overline{\theta} > \underline{\theta}$, we can obtain:

$$\begin{aligned}\left.\frac{\partial^2 W}{\partial \overline{r}^2}\right|_{\overline{r}=\overline{r}^{FB}} &= \frac{1}{4\left(1+\overline{r}^{FB}\right)^2}\left\{3\eta\alpha^2 D\left(s\right)^2\sigma^2\left[1-\ln\left(1+\overline{r}^{FB}\right)\right]\right. \\ &\quad \left. -3\eta\left[\xi-\tau+a\left(s\right)-z-\alpha D\left(s\right)\right]\alpha D\left(s\right)\sigma - 4\overline{\theta} f D\left(s\right)\sigma\right\} \\ &< \frac{1}{4\left(1+\overline{r}^{FB}\right)^2}\left\{3\eta\alpha^2 D\left(s\right)^2\sigma^2\left[1-\ln\left(1+\overline{r}^{FB}\right)\right] - 4m\left(s\right)\right\}.\end{aligned} \tag{1.85}$$

Then, according to (1.85) and $0 < 3\eta\alpha^2 D\left(s\right)^2\sigma^2 \leqslant 4m\left(s\right)$, we have $\left.\frac{\partial^2 W}{\partial \overline{r}^2}\right|_{\overline{r}=\overline{r}^{FB}} < 0$.

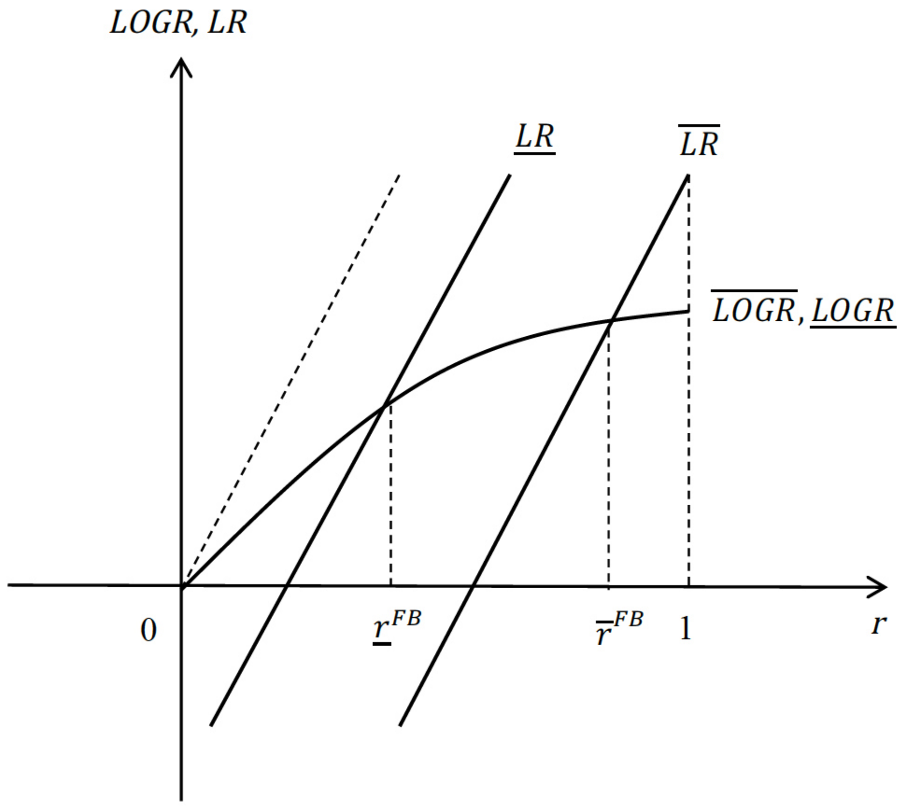

Figure 1.4: Case 1 (Complete Information of Scenario 3)

Moreover, we can show $\left.\frac{\partial^2 W}{\partial \underline{r}^2}\right|_{\underline{r}=\underline{r}^{FB}} < 0$ analogously.

Case 2: $\frac{2}{D(s)\sigma}\sqrt{\frac{m(s)}{3\eta}} < \alpha \leqslant \frac{\xi-\tau+a(s)-z}{D(s)}$ and $I_2 \geqslant 0$

This case is shown in Figure 1.5. In this case, the slope of $\overline{LR}$ (resp. $\underline{LR}$) is first less and then greater than that of $\overline{LOGR}$ (resp. $\underline{LOGR}$). Moreover, the signs of $\overline{LR}$'s and $\underline{LR}$'s intercepts on horizontal axis are positive.

As shown in Figure 1.5, when $r = r_{tan}$:

$$3\eta\alpha^2 D(s)^2 \sigma^2 \frac{1}{1+r_{tan}} = 4m(s)\,. \tag{1.86}$$

Thus, we have:

$$\ln\left[\frac{3\eta\alpha^2 D(s)^2 \sigma^2}{4m(s)}\right] = \ln(1+r_{tan}) < \ln\left(1+\overline{r}^{FB}\right). \tag{1.87}$$

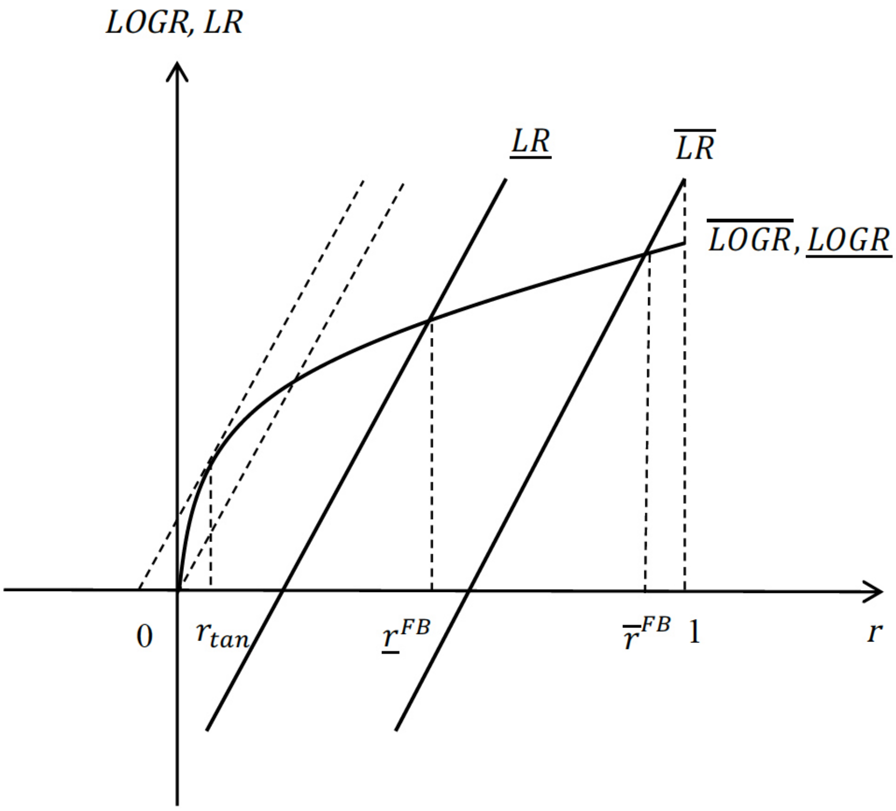

Figure 1.5: Case 2 (Complete Information of Scenario 3)

According to (1.87), we can obtain:

$$\begin{aligned}\left.\frac{\partial^2 W}{\partial \overline{r}^2}\right|_{\overline{r}=\overline{r}^{FB}} &< \frac{1}{4\left(1+\overline{r}^{FB}\right)^2}\left\{3\eta\alpha^2 D\left(s\right)^2\sigma^2\left[1-\ln\left(1+\overline{r}^{FB}\right)\right]-4m\left(s\right)\right\} \\ &< \frac{1}{4\left(1+\overline{r}^{FB}\right)^2}\left\{3\eta\alpha^2 D\left(s\right)^2\sigma^2\left\{1-\ln\left[\frac{3\eta\alpha^2 D\left(s\right)^2\sigma^2}{4m\left(s\right)}\right]\right\}-4m\left(s\right)\right\} \\ &= \frac{m\left(s\right)}{\left(1+\overline{r}^{FB}\right)^2}\left\{\frac{3\eta\alpha^2 D\left(s\right)^2\sigma^2}{4m\left(s\right)}\left\{1-\ln\left[\frac{3\eta\alpha^2 D\left(s\right)^2\sigma^2}{4m\left(s\right)}\right]\right\}-1\right\}. \end{aligned} \tag{1.88}$$

Then, because $[x\left(1-\ln x\right)-1]_{sup}=0$ when $x>1$, we have $\left.\frac{\partial^2 W}{\partial \overline{r}^2}\right|_{\overline{r}=\overline{r}^{FB}}<0$.

Moreover, we can show $\left.\frac{\partial^2 W}{\partial \underline{r}^2}\right|_{\underline{r}=\underline{r}^{FB}}<0$ analogously.

Case 3: $\frac{2}{D(s)\sigma}\sqrt{\frac{m(s)}{3\eta}}<\alpha\leqslant\frac{\xi-\tau+a(s)-z}{D(s)}$, $I_1\geqslant 0$, and $I_2<0$

This case is shown in Figure 1.6. In this case, the slope of $\overline{LR}$ (resp. $\underline{LR}$) is first less and then greater than that of $\overline{LOGR}$ (resp. $\underline{LOGR}$). Moreover, the sign of $\overline{LR}$'s intercept on horizontal axis is positive, while the sign of $\underline{LR}$'s intercept on horizontal axis is negative.

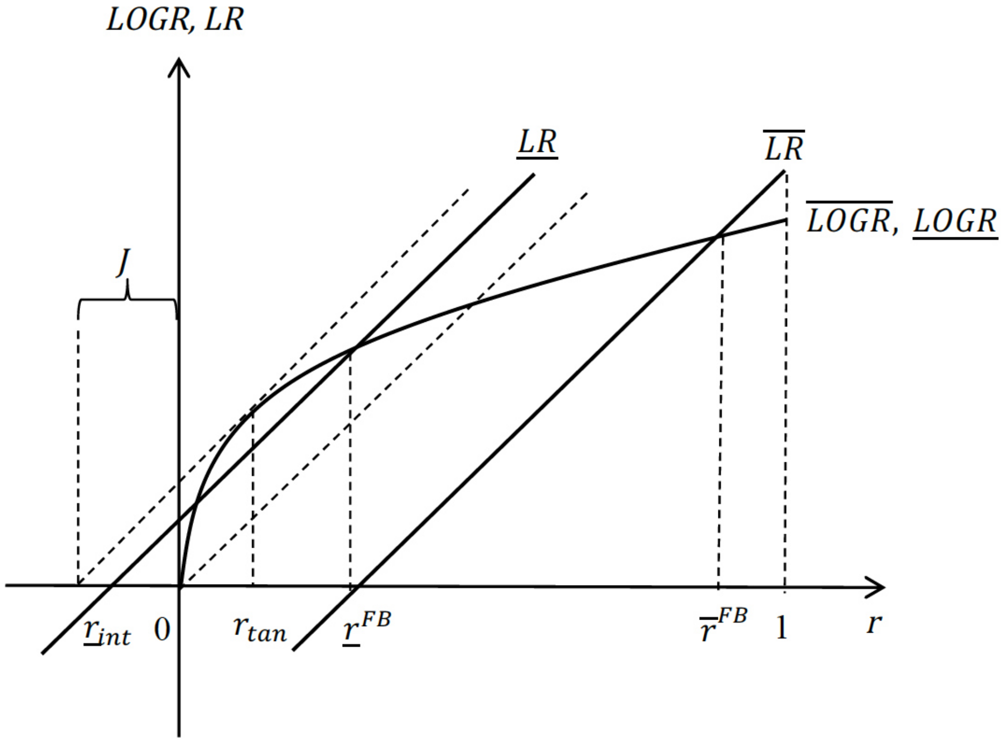

Figure 1.6: Case 3 (Complete Information of Scenario 3)

Following Case 2, we can show $\left.\frac{\partial^2 W}{\partial \overline{r}^2}\right|_{\overline{r}=\overline{r}^{FB}} < 0$ analogously. Next, we consider $\left.\frac{\partial^2 W}{\partial \underline{r}^2}\right|_{\underline{r}=\underline{r}^{FB}}$.

As shown in Figure 1.6, $\underline{LOGR}$ and $\underline{LR}$ have two intersections. For the first intersection $\underline{r} = \widehat{\underline{r}}$, because $I_2 < 0$, we have:

$$\begin{aligned}\left.\frac{\partial^2 W}{\partial \underline{r}^2}\right|_{\underline{r}=\widehat{\underline{r}}} &= \frac{1}{4\left(1+\widehat{\underline{r}}\right)^2}\left\{3\eta\alpha^2 D\left(s\right)^2\sigma^2\left[1-\ln\left(1+\widehat{\underline{r}}\right)\right]\right.\\ &\left.-3\eta\left[\xi-\tau+a\left(s\right)-z-\alpha D\left(s\right)\right]\alpha D\left(s\right)\sigma-4\underline{\theta} f D\left(s\right)\sigma\right\}\\ &> \frac{1}{4\left(1+\widehat{\underline{r}}\right)^2}\left\{3\eta\alpha^2 D\left(s\right)^2\sigma^2\left[1-\ln\left(1+\widehat{\underline{r}}\right)\right]-4m\left(s\right)\right\}. \end{aligned} \tag{1.89}$$

Given the value of $m(s)$, a large value of $3\eta\alpha^2 D(s)^2\sigma^2$ will result in $\left.\frac{\partial^2 W}{\partial \underline{r}^2}\right|_{\underline{r}=\widehat{\underline{r}}} > 0$, violating the second-order condition.

Next, consider the second intersection of $\underline{LOGR}$ and $\underline{LR}$, that is, $\underline{r} = \underline{r}^{FB}$. As shown in Figure 1.6, we have:

$$4m(s) = \frac{3\eta\alpha^2 D(s)^2 \sigma^2 \ln(1 + r_{tan})}{J + r_{tan}}. \tag{1.90}$$

Because the sign of $\underline{LR}$'s intercept on horizontal axis is negative, there exists a solution only when $J > |\underline{r}_{int}|$, that is:

$$\begin{aligned} J &= \frac{3\eta\alpha^2 D(s)^2 \sigma^2 \ln(1 + r_{tan})}{4m(s)} - r_{tan} \\ &> \frac{-3\eta[\xi - \tau + a(s) - z - \alpha D(s)]\alpha D(s)\sigma - 4\underline{\theta} f D(s)\sigma + 4m(s)}{4m(s)} = |\underline{r}_{int}|. \end{aligned} \tag{1.91}$$

Then, according to (1.86), we have:

$$\begin{aligned} &-3\eta[\xi - \tau + a(s) - z - \alpha D(s)]\alpha D(s)\sigma - 4\underline{\theta} f D(s)\sigma \\ &< 3\eta\alpha^2 D(s)^2 \sigma^2 \ln(1 + r_{tan}) - 4m(s)(1 + r_{tan}) \\ &= 3\eta\alpha^2 D(s)^2 \sigma^2 \left\{\ln\left[\frac{3\eta\alpha^2 D(s)^2 \sigma^2}{4m(s)}\right] - 1\right\}. \end{aligned} \tag{1.92}$$

According to (1.92), we can obtain:

$$\begin{aligned} \left.\frac{\partial^2 W}{\partial \underline{r}^2}\right|_{\underline{r}=\underline{r}^{FB}} &= \frac{1}{4(1 + \underline{r}^{FB})^2}\left\{3\eta\alpha^2 D(s)^2 \sigma^2 \left[1 - \ln\left(1 + \underline{r}^{FB}\right)\right]\right. \\ &\quad \left. -3\eta[\xi - \tau + a(s) - z - \alpha D(s)]\alpha D(s)\sigma - 4\underline{\theta} f D(s)\sigma\right\} \\ &< \frac{3\eta\alpha^2 D(s)^2 \sigma^2}{4(1 + \underline{r}^{FB})^2}\left\{\ln\left[\frac{3\eta\alpha^2 D(s)^2 \sigma^2}{4m(s)}\right] - \ln\left(1 + \underline{r}^{FB}\right)\right\} \\ &= \frac{3\eta\alpha^2 D(s)^2 \sigma^2}{4(1 + \underline{r}^{FB})^2} \ln\left[\frac{3\eta\alpha^2 D(s)^2 \sigma^2 \frac{1}{(1+\underline{r}^{FB})}}{4m(s)}\right]. \end{aligned} \tag{1.93}$$

Then, because $3\eta\alpha^2 D(s)^2 \sigma^2 \frac{1}{(1+\underline{r}^{FB})} < 4m(s)$ always holds, according to (1.93), we have $\left.\frac{\partial^2 W}{\partial \underline{r}^2}\right|_{\underline{r}=\underline{r}^{FB}} < 0$.

Case 4: $\frac{2}{D(s)\sigma}\sqrt{\frac{m(s)}{3\eta}} < \alpha \leqslant \frac{\xi-\tau+a(s)-z}{D(s)}$ and $I_1 < 0$

This case is shown in Figure 1.7. In this case, the slope of $\overline{LR}$ ($\underline{LR}$) is first less and then greater than that of $\overline{LOGR}$ ($\underline{LOGR}$). Moreover, the signs of $\overline{LR}$'s and $\underline{LR}$'s intercepts on horizontal axis are negative.

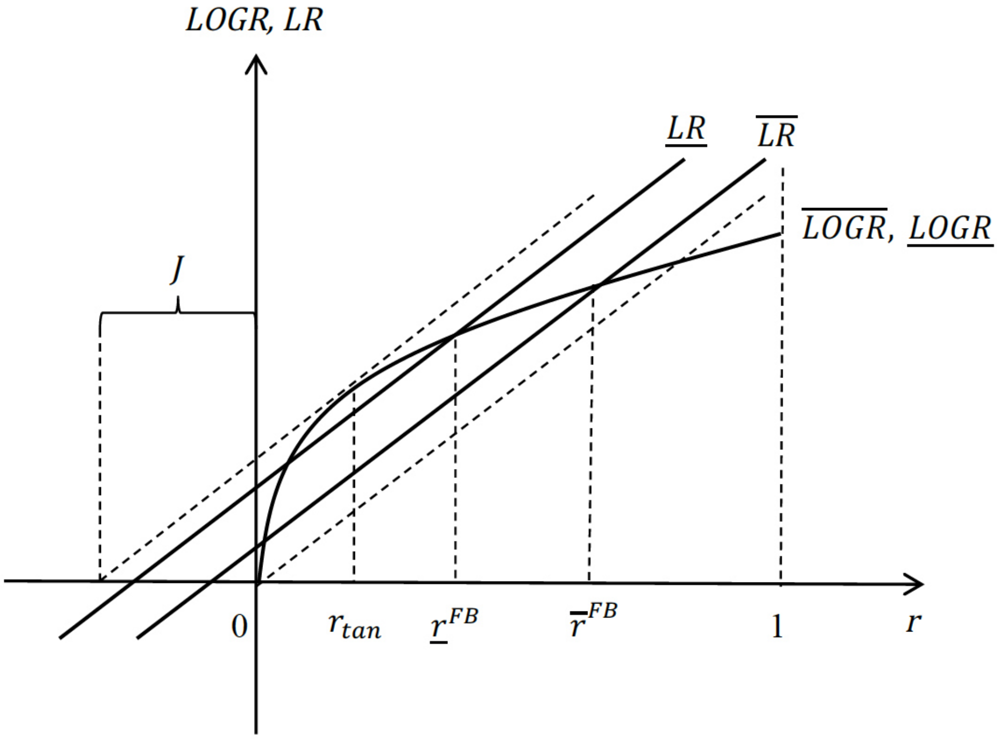

Figure 1.7: Case 4 (Complete Information of Scenario 3)

Following Case 3, we can show $\left.\frac{\partial^2 W}{\partial \overline{r}^2}\right|_{\overline{r}=\overline{r}^{FB}} < 0$ and $\left.\frac{\partial^2 W}{\partial \underline{r}^2}\right|_{\underline{r}=\underline{r}^{FB}} < 0$ analogously.

1.7.1.2 Comparison of Optimal Degrees in Scenario 4

The comparison is shown is Figures 1.8 through 1.12.

1.7.2 Proof of Proposition 1.2

Proof. We first consider the effect of $D(s)$ on $\overline{r}^{SB}$. Taking the derivative of (1.49) with respect to $D(s)$ and using (1.49), we can obtain:

$$\frac{\partial \overline{\Omega}}{\partial D(s)} = \frac{1}{D(s)} \left\{ 4(1+\lambda) m(s) \left(1+\overline{r}^{SB}\right) - (3+2\lambda)\eta\alpha^2 D(s)^2 \sigma \left[1-\sigma \ln\left(1+\overline{r}^{SB}\right)\right] \right\}. \tag{1.94}$$

Rewriting (1.94), we can obtain the first point of Proposition 1.2.

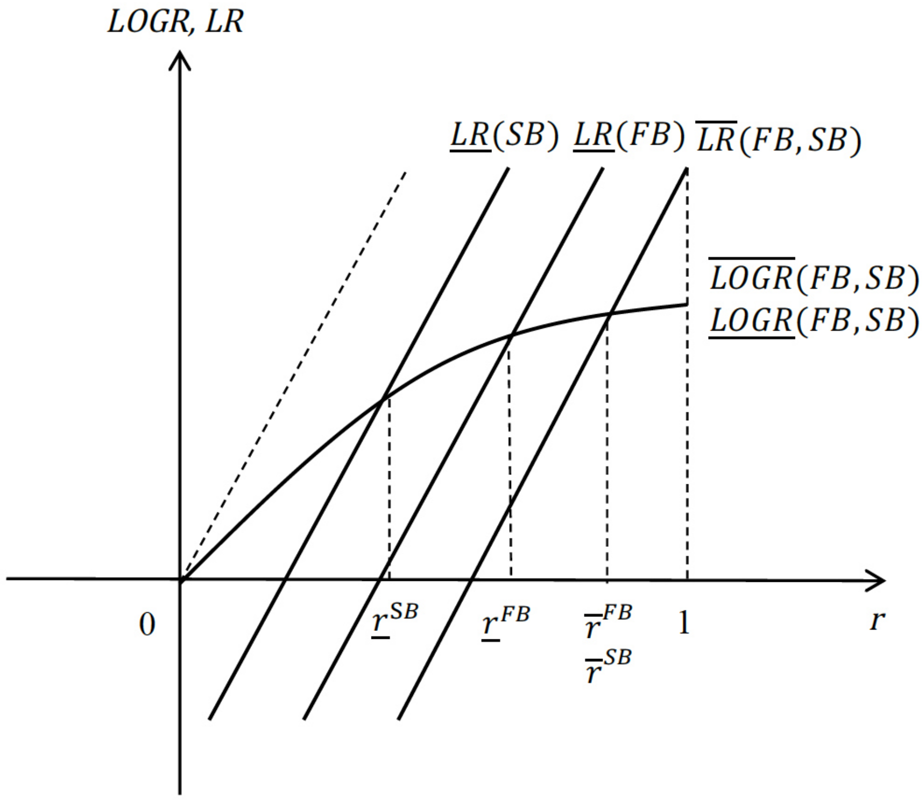

Figure 1.8: Case 1 (Scenario 4)

Moreover, we can obtain the second point analogously. □

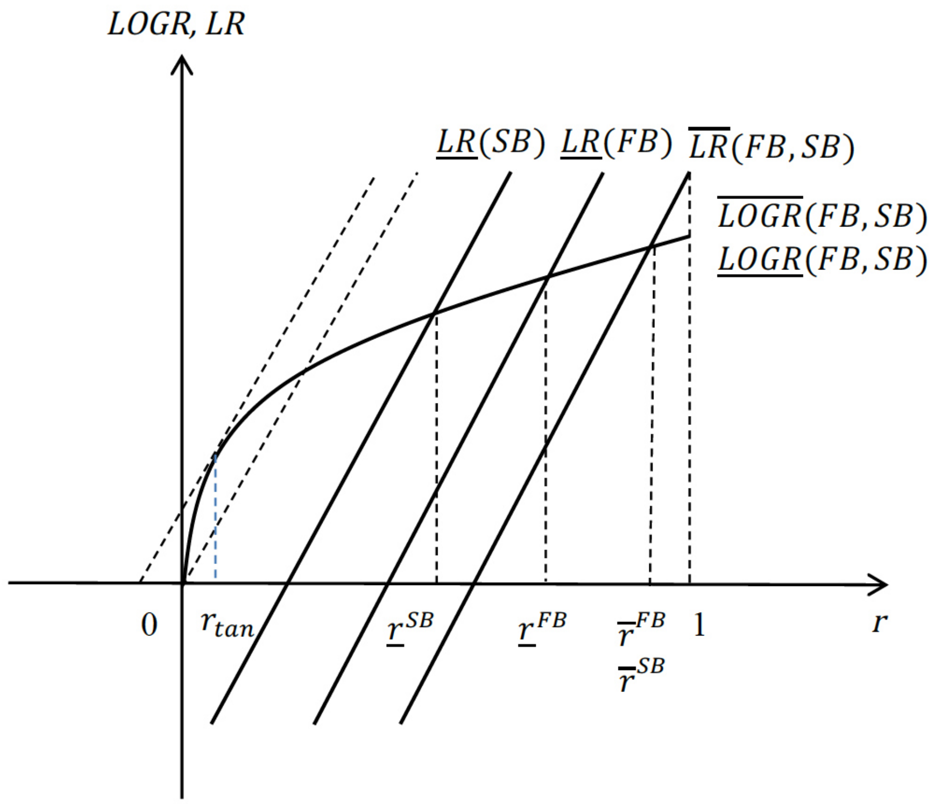

Figure 1.9: Case 2 (Scenario 4)

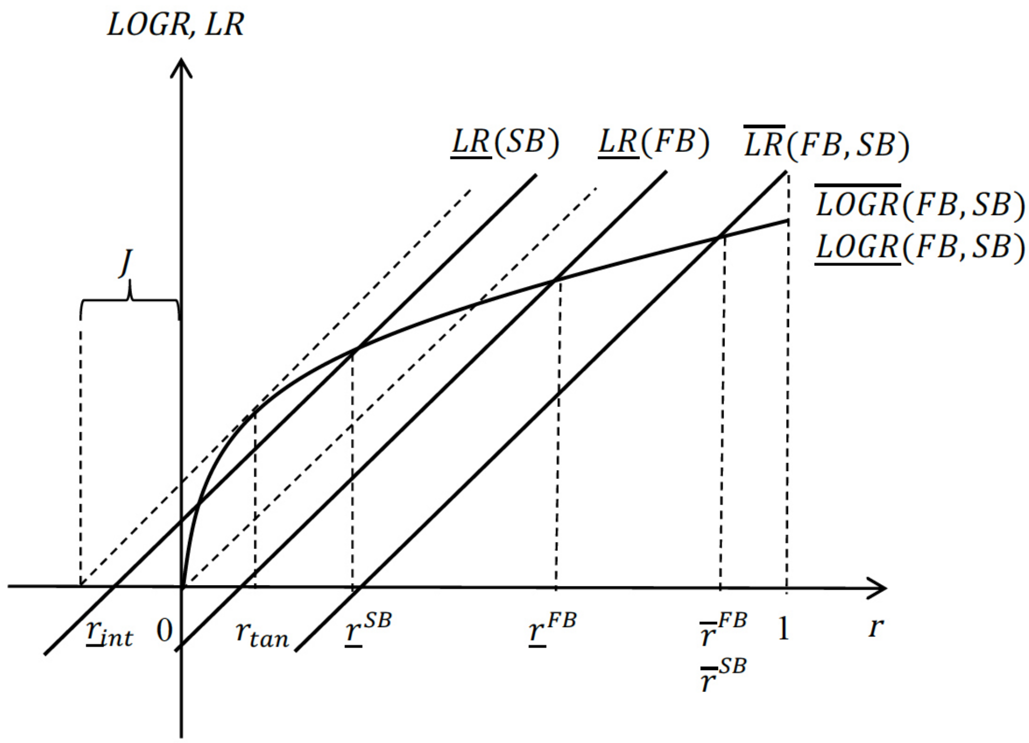

Figure 1.10: Case 3-1 (Scenario 4)

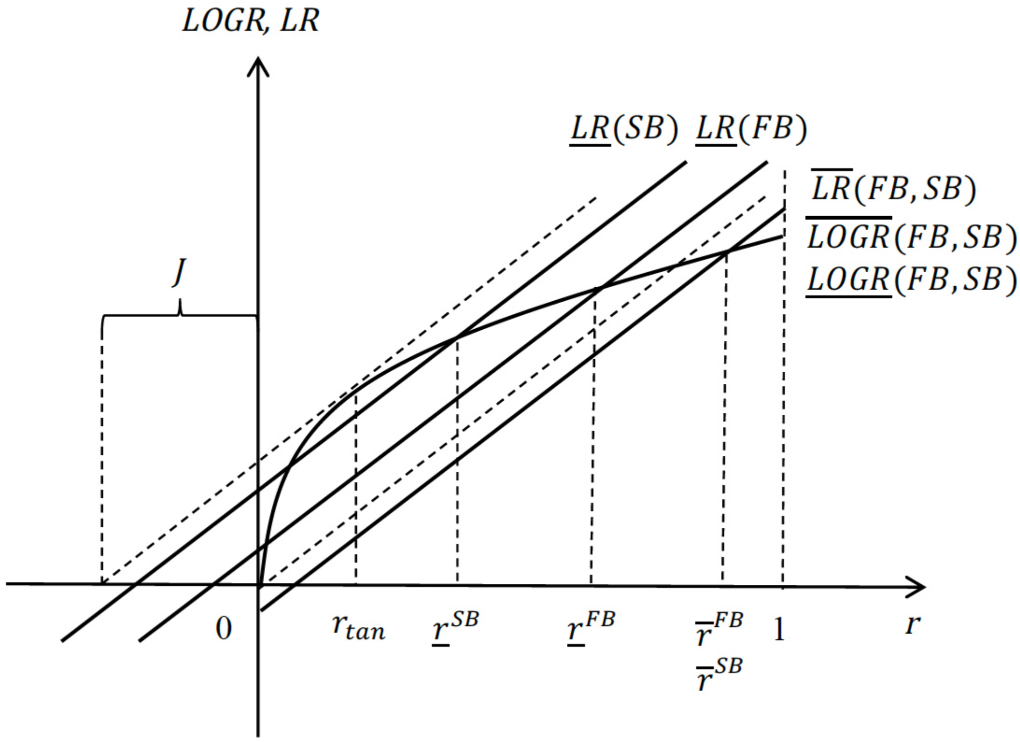

Figure 1.11: Case 3-2 (Scenario 4)

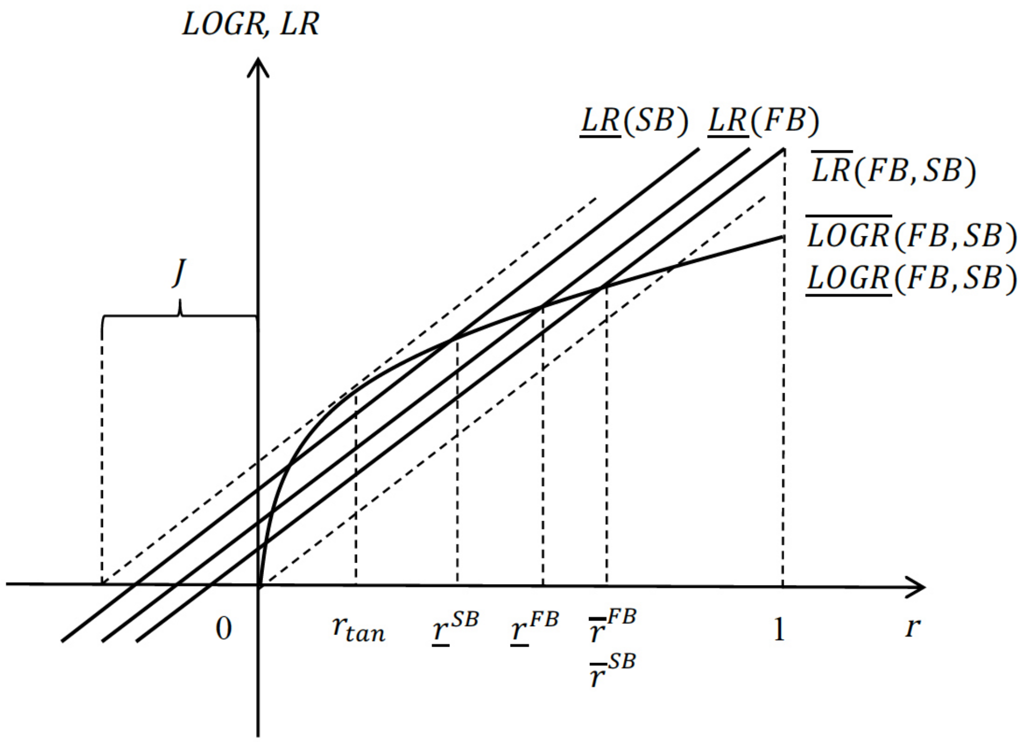

Figure 1.12: Case 4 (Scenario 4)

Chapter 2

Congestion Delays, Horizontal Product Differentiation and Airline Networks

2.1 Introduction

Air traffic delays remain a significant and worldwide reality. In Europe, from 2005 to 2015, the percentages of delayed flights for arrivals are approximately 40%, with an average delay per delayed flight for arrivals of approximately 29 minutes (see Figure 2.1). As a result, and considering also the time values estimated by University of Westminster (2015) and Cook and Tanner (2015), delays are costly to both airlines and passengers. In fact, it has been well established that different airline's network structures may result in varying degrees of delays.[1] However, the means by which airlines respond to these costly delays by adjusting their network structure is, to date, little studied.[2] Therefore, this chapter aims to study how congestion delays shape airline network structure, within which the role of horizontal product differentiation will also be investigated.

In the model, I consider a monopoly airline and passengers in three markets. Passengers maximize their utility, which is a quadratic function of traffic of imperfectly substitutable non-stop and one-stop air transport services. This imperfect substitution

[1]For example, Mayer and Sinai (2003) find that hubbing is the primary economic contributor to air traffic congestion. Moreover, Figure 2.2 also provides an evidence.

[2]In the literature, only Fageda and Flores-Fillol (2015) and Silva et al. (2014) consider congestion delays in airline network choice.

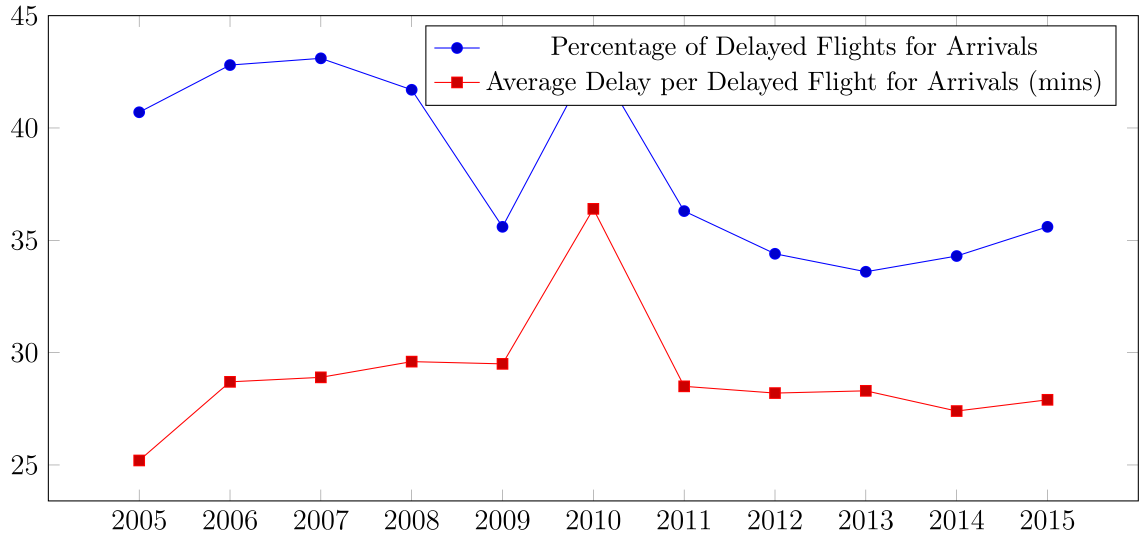

Figure 2.1: Air Traffic Delays for Arrivals in Europe
Source: EUROCONTROL (2010, 2011, 2016)

indicates the horizontal product differentiation of services. Moreover, passengers value flight frequency because they dislike schedule delays. According to Douglas and Miller (1974) and Panzar (1979), schedule delay is the absolute difference between a passenger's most preferred departure time and that of his/her actual departure. The higher the flight frequency is, the shorter schedule delays will be. Thus, considering also the travel time, the difference in flight frequency and travel time indicates the vertical product differentiation of services.

The airline maximizes its profit by choosing a network structure, flight frequencies and passenger traffic. One feature of the model is the coverage of all possible network structures in a three-city network, that is, a hub-and-spoke network (HS), point-to-point network (PP), mixed network (MX), and 2-hub network (2H), as well as a 3-hub network (3H) as an extension. Under HS (e.g., Alitalia), passengers who travel between two spoke airports are required to connect at a hub airport. Differing from HS, under PP (e.g., Ryanair and easyJet), passengers can travel directly from one airport to any other. Moreover, under MX (e.g., Air France), passengers who travel between two spoke airports can choose either one-stop or non-stop services, implying that MX is a combination of HS and PP. Finally, under 2H (e.g., Lufthansa and Air France-KLM group), two hubs are available for connection, while under 3H, each airport works as a hub.

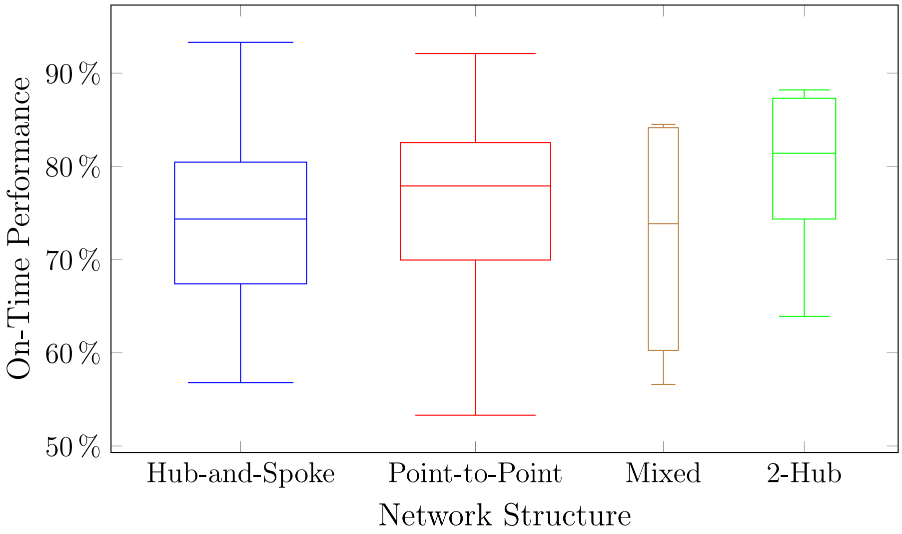

Figure 2.2: Box Plots for Airline Network and On-Time Performance (July 2016)
Source: The on-time performance data is from OAG (2016) and airlines' network structures are from their official websites.
Note: 1. The airline sample includes the top 25 European airlines with their subsidiaries by passenger volume, except Virgin Atlantic, which does not operate flights within Europe; 2. According to OAG (2016), the on-time performance is the percentage of flights that depart or arrive within 15 minutes of schedule.

In addition, I assume that the airline's cost function includes only the expected congestion delay cost. In fact, this cost specification excludes the fixed cost of operating a flight and the variable cost of serving a passenger. According to Doganis (2009), the short-run marginal cost of serving an extra passenger on a flight is close to zero. In addition, Smyth and Pearce (2007) and Pearce (2013) also claim a low marginal cost per passenger. Thus, we can omit this in order to simplify the analysis. Fixed cost, however, indeed accounts for a large share of the total cost. Previous literature models fixed cost to capture the economies of traffic density and then to explain the existence of the hub-and-spoke network. In fact, modeling the fixed cost will not provide further insights but will make the model intractable. More importantly, excluding fixed cost can also help us isolate the impact of the economies of traffic density on airline network choice.

The main trade-offs in the model are between congestion delays and schedule delays and between flight frequency and travel time. For the former, if the flight frequency in one route becomes higher, congestion delays in more than one route will increase. However, passengers' schedule delays in that route will decrease. For the latter, between two spoke airports, non-stop services always have lower flight frequency but shorter travel time, while one-stop services always have higher flight frequency but longer travel time.

When solving the model, I first consider HS, PP and MX alone in order to compare them with Lin (2012). Besides the common results, I find, surprisingly, that the airline may choose PP even when the extra travel time disutility of one-stop services is relatively low. In fact, this result arises from the inclusion of congestion delays, that is, a negative network externality. Specifically, on the one hand, because of the traffic concentration of HS and the convex cost function with respect to flight frequencies, the introduction of congestion delays, or negative network externality on the cost side, creates a cost disadvantage for HS, and thus makes HS less profitable. On the other hand, the introduction of an omitted negative network externality also reduces the region that makes MX feasible. Then, there will emerge an interval of extra travel time disutility, in which the disutility is too high to make HS more profitable and too low to make MX feasible, leaving PP as the airline's optimal network structure. The key insight from this result is that including the omitted negative network externality makes HS and MX less effective than previously understood as in Lin (2012). In fact, in addition to the commonly received advantage of PP, that is, saving the extra travel time of connecting at a hub airport, this result might provide another explanation for why some legacy airlines start to use PP in some local markets.

I then incorporate 2H into the analysis. I find that, without considering the airline's fixed investments of developing a hub airport, 2H will dominate the three other network structures as long as it is feasible. Because non-stop and one-stop services are imperfectly substitutable, passengers can obtain higher utility if both non-stop and one-stop services are available to choose than if only one of them is available. Under 2H, there are two markets, in which both non-stop and one-stop services are available. Under MX, there is only one such market. However, under HS and PP, passengers in any market cannot choose between non-stop and one-stop services. Therefore, because of the exploitation of horizontal product differentiation to a larger extent, 2H can generate higher passenger utility and then a higher airline profit than the three other network structures. To summarize, this result shows the role of horizontal product differentiation in improving passenger utility and airline profit.

In fact, this result can also help us understand the multi-hubbing and de-hubbing phenomena in the airline industry. In reality, we can observe that some airlines develop new hubs and then use 2H (multi-hubbing), while some others change from 2H to a single hub network (de-hubbing). Here comes a question: what is the motivation for airlines to use 2H? One answer is that when an airline's hub airport is congested, the airline can develop another hub to reduce the congestion of the previous hub. However, this answer might not be strong enough because besides 2H, PP and MX can also reduce hub airport congestion. The result above shows the role of horizontal product differentiation and thus provides another explanation for the use of 2H.

Moreover, I conduct comparative-static analysis and find that, for the airline under MX and 2H, there exist some strategic effects when the values of the parameters change, due to the division of local and connecting traffic in one market. For instance, under MX, when the marginal congestion delay cost increases, the change of flight frequency between two spoke airports depends on the trade-off between the direct negative effect of a higher delay cost and the strategic redistribution of traffic among different routes. Furthermore, in welfare analysis, I derive not only the first-best, but also the second-best, socially optimal network structure, and both show the airline's inefficient biases towards PP and 2H. Besides which, I extend the analysis to 3H and again show the role of horizontal product differentiation.

The contributions of this chapter are threefold. First, the majority of the airline network choice literature (see Oum et al., 1995; Berechman and Shy, 1998; Kawasaki, 2008) compare HS and PP alone, with a few others including one more network structure, either MX (see Lin, 2012) or 2H (see Alderighi et al., 2005). In fact, Starr and Stinchcombe (1992) and Hendricks et al. (1995, 1999) use rather general models, allowing the network design to be endogenous. However, their frameworks focus mainly on airline cost but not on passenger demand, which thus leaves little room for the optimality of network structures other than HS and PP. Accordingly, this chapter contributes to the literature by incorporating all possible network structures in a three-city network, each of which has the potential to be an airline's optimal network structure.

Second, most of the previous studies explain airline network choice from the point of view of the economies of traffic density (see Bittlingmayer, 1990; Hendricks et al., 1995, 1999), demand uncertainty (see Barla and Constantatos, 2005) and schedule delays (see Brueckner, 2004), while only Fageda and Flores-Fillol (2015) and Silva et al. (2014) consider congestion delays. Therefore, this chapter contributes to the currently limited literature available on the effects of congestion delays on airline network choice.

The third is regarding the perspective of analyzing 2H. The conventional wisdom concerning the function of 2H (see Bilotkach et al., 2013) is to divert passengers from one hub to another and thus reduce hub airport congestion. In fact, a four-city network (see Figure 2.3) is the minimum requirement in order to show the congestion reduction function of 2H; however, this inevitably brings analytical difficulties. To make the analysis tractable, previous literature (see Bilotkach et al., 2013) has had to simplify many important elements.

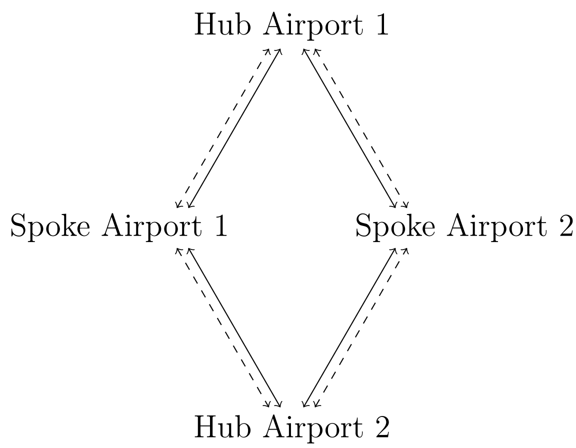

Figure 2.3: Commonly Studied Four-City 2-Hub Network
Note: In all network structure figures, solid (resp. dotted) lines represent the existence of non-stop (resp. one-stop) services.

Nevertheless, in this chapter, I focus on the exploitation of horizontal product differentiation in 2H, which requires a three-city network only. In reality, the horizontal product differentiation can come from, for instance, different departure time slots (see Encaoua et al., 1996) and brand loyalty (see Brueckner and Whalen, 2000; Brueckner and Flores-Fillol, 2007). To see this, one example is the Lufthansa network (see Figure 2.4), in which Frankfurt and Munich are hub airports and Toulouse is a spoke airport.[3] Specifically,

[3]Concrete examples of other network structures are given in Section 2.7.

passengers traveling between Toulouse and Frankfurt can take either Lufthansa flights to arrive directly, or Lufthansa Cityline (a subsidiary of Lufthansa) flights and Lufthansa flights connecting at Munich. Similarly, passengers traveling between Toulouse and Munich can take either Lufthansa Cityline flights or Air Dolomiti (a subsidiary of Lufthansa) flights to arrive directly, or Lufthansa flights connecting at Frankfurt. For different departure time slots, flight schedules between Toulouse and Frankfurt (Munich) provide evidence (see Section 2.7). For brand loyalty, as shown in Figure 2.4, Lufthansa uses differentiated brands to operate the network.

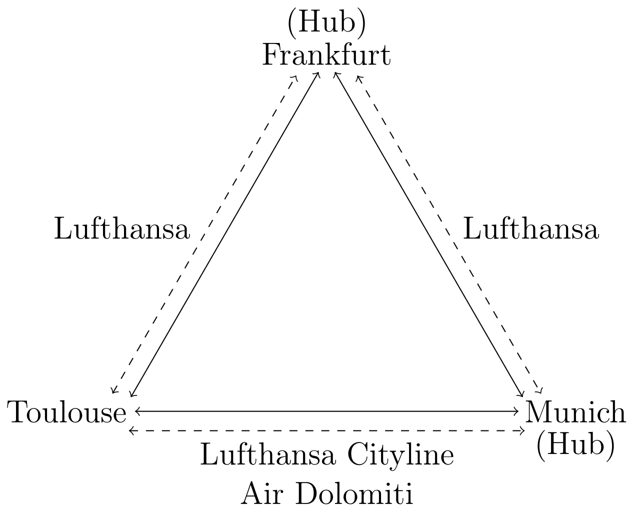

Figure 2.4: An Example of Three-City 2-Hub Network (Lufthansa)

Related Literature

This chapter is closely related to three branches of literature. Since the airline deregulation in the USA in 1978, there is a growing body of literature on airline network choice, in which the first branch is the scope of network structures. Previously, the literature has focused on the choice between HS and PP, but recently there has been a shift towards some "real-life" network structures, that is, MX and 2H. For MX, among others, Dunn (2008) empirically examines an airline's choice of providing non-stop services or not, given that the airline has (not) provided one-stop services. Moreover, Fageda and Flores-Fillol (2012) study hub airlines' incentives to provide non-stop services between spoke airports under two recent innovations, that is, the regional jet technology and low-cost business model. Further, Lin (2012) studies the network choice among HS, PP and MX under both monopoly and duopoly setups. For 2H, Bilotkach et al. (2013) use the utility function à la Mussa-Rosen (see Mussa and Rosen, 1978) to study the function of diverting traffic of 2H. Moreover, Wang (2016) theoretically examines the optimality of 2H in the spirit of Brueckner and Spiller (1991).

The second branch of literature is the theory explaining airline network choice. One theory is that airlines can better exploit the economies of traffic density under HS. According to Hendricks et al. (1995), the economies of traffic density arise when the cost per passenger on a route decreases with the number of passengers flying on that route. As a result, because HS has a higher traffic density than PP, as long as the cost of extra travel time of one-stop services is not high enough, the total cost for a given level of demand may be lower under HS than PP. Some empirical studies have confirmed the economies of traffic density under HS (see Brueckner et al., 1992; Brueckner and Spiller, 1994). Moreover, besides Hendricks, Piccione, and Tan (1995), theoretical studies explaining airline network choice from the point of view of the economies of traffic density include, for example, Bittlingmayer (1990), Oum et al. (1995) and Hendricks et al. (1999) [4].

Another theory concerns demand uncertainty. Barla and Constantatos (2005) show the flexibility of HS under uncertainty. Interestingly, they also find that both airlines may choose PP, because by committing not to enjoy the flexibility, airlines can avoid the spread of competition from one market to others. Furthermore, Hu (2010) also considers demand uncertainty but under a different setup.

Because passengers greatly value flight frequency (see Berry and Jia, 2010), schedule delays may be an important factor affecting airline network choice. Berechman and Shy (1998) and Brueckner and Zhang (2001) first connected airline network structure and scheduling. Then, Brueckner (2004) builds a framework that improves upon the previous two studies and shows that a high disutility of schedule delays would be conducive to HS. In addition, other relevant studies include Kawasaki (2008) and Flores-Fillol (2009).

Furthermore, some studies have introduced congestion delays into their models. In

[4] Hendricks et al. (1999) also consider the nature of competition, that is, airlines compete aggressively or not.

Fageda and Flores-Fillol (2015), they find that, even if there is a higher delay cost, duopoly airlines exhibit a preference for HS, which may be inefficient from the perspective of a welfare-maximizing social planner. Moreover, Silva et al. (2014) also consider congestion delays and show that a higher value of travel time favors PP. However, the frameworks of these two studies do not allow for the analysis of MX and 2H. Besides which, in Fageda and Flores-Fillol (2015), passenger demand is also perfectly inelastic.

The last theory is the horizontal product differentiation. Lin (2012) finds that HS will be the airline's optimal network structure if passengers do not differentiate between non-stop and one-stop services too much, and if the extra travel time disutility of one-stop services is low. However, if the passenger differentiation is substantial, MX (resp. PP) will be the airline's optimal network structure when the extra travel time disutility is low (resp. high).

The third branch of literature is the modeling of vertical and horizontal product differentiation. With respect to the vertical product differentiation, Flores-Fillol (2010) considers that airlines may compete in flight frequencies. Moreover, in the models of Brueckner (2004) and Kawasaki (2008), one-stop services always cost more time than non-stop services. With respect to the horizontal product differentiation, Encaoua et al. (1996) consider the difference of the departure time slot, while Brueckner and Whalen (2000) and Brueckner and Flores-Fillol (2007) consider brand loyalty. In addition, Lin (2012) uses the quadratic utility function in Dixit (1979) to capture the horizontal product differentiation.

The rest of the chapter is organized as follows. Section 2.2 describes the model. Section 2.3 derives the market outcome, which includes comparisons within each network structure and across different network structures, as well as comparative statics. Section 2.4 compares the airline's optimal network structure with the first-best and second-best socially optimal network structures. Section 2.5 extends the analysis to the 3-hub network. Finally, Section 2.6 concludes.

2.2 The Model

In this model, I consider three symmetrically-located airports (cities), K, A and B, in which all of them are capacity-constrained and K and A can work as hub airports. Passengers who wish to travel between city pairs form markets AK, BK and AB, denoted by markets 1, 2 and 3, respectively. The airline can choose from amongst four network structures, that is, a hub-and-spoke network (HS), point-to-point network (PP), mixed network (MX), and a 2-hub network (2H). Moreover, for completeness, I study a 3-hub network (3H), or an all-hub network, as an extension.

2.2.1 Passengers

Assume that there exists one representative passenger in each market. Following Dixit (1979), the passenger utility in market $i\,(\in\{1,2,3\})$ is:

$$U^i\left(Q_0,Q_1\right)=\alpha_0^iQ_0+\alpha_1^iQ_1-\frac{1}{2}\left(\beta_0\left(Q_0\right)^2+2\gamma Q_0Q_1+\beta_1\left(Q_1\right)^2\right). \tag{2.1}$$

In (2.1), Q_0 is the traffic of non-stop services and Q_1 is the traffic of one-stop services. The concavity of passenger utility with respect to Q_0 and Q_1 requires $\beta_0>0$, $\beta_1>0$, and $\gamma^2\leqslant\beta_0\beta_1$. Moreover, the parameter $\gamma>0$ represents the passenger differentiation between non-stop and one-stop services. The greater γ is, the higher the degree of substitution will be between non-stop and one-stop services. To simplify the algebra, I assume $\beta_0=\beta_1=1$ and $\gamma=\frac{1}{2}$.[5] Besides, α_0^i and α_1^i are the marginal utility of local and connecting traffic, respectively, with:

$$\alpha_0^i=v+\lambda F^i, \tag{2.2}$$

$$\alpha_1^i=v+\lambda\frac{\sum_{j\neq i}F^j}{2}-t. \tag{2.3}$$

[5] If we relax the assumptions for β_0, β_1 and γ, the degree of substitutability between non-stop and one-stop services will change, and thus the region that a network structure is optimal for the airline may enlarge or shrink. However, all of the properties and intuitions found in this chapter will not be affected.

In (2.2) and (2.3), $v \in \mathbb{R}_{++}$ is the passenger's travel benefit. F^i is the flight frequency between the city pair of market i, and $t \in \mathbb{R}_{++}$ is the extra travel time disutility, in which the extra travel time of one-stop services includes both the extra in-aircraft travel time compared to non-stop services and the layover time at the hub airport. The parameter $\lambda \in \mathbb{R}_{++}$ is positively related to the marginal benefit of schedule delay reduction and is hereafter used as the proxy of the latter. Terms λF^i and $\lambda \frac{\sum_{j \neq i} F^j}{2}$ are the schedule delay reduction benefits of non-stop and one-stop services, respectively. In particular, following Lin (2012), the schedule delay reduction of one-stop services takes the average of two relevant flight frequencies. Another method is to take the minimum; however this is problematic if we consider the layover time at the hub airport (see Rietveld and Brons, 2001).

Following Singh and Vives (1984), the passenger surplus in market $i\,(\in \{1,2,3\})$ is the difference between passenger utility and total payment:

$$U^i\left(Q_0, Q_1\right) - \sum_{\tau \in \{0,1\}} P^i_\tau Q_\tau. \tag{2.4}$$

In (2.4), $P^i_\tau\,(\tau = 0, 1)$ are the fares of non-stop and one-stop services, respectively.

Next, passengers' demand for each network structure is discussed.

2.2.1.1 Hub-and-Spoke Network

Under HS (Figure 2.5), K is a hub airport and A and B are spoke airports. Then, under HS, passengers in markets AK and BK can choose non-stop services only, and their marginal utility of local traffic is $\alpha^i_0 = v + \lambda f^i_{HS}$, in which f^i_{HS} is the flight frequency between the city pair of market $i\,(\in \{1,2\})$. Moreover, the passenger in market AB can choose one-stop services connecting at K only, and his/her marginal utility of connecting traffic is $\alpha^3_1 = v + \lambda \frac{\sum_{i \in \{1,2\}} f^i_{HS}}{2} - t$.

After observing fares and flight frequencies, the representative passenger in each market chooses traffic in order to maximize passenger surplus. Then, passengers' optimal

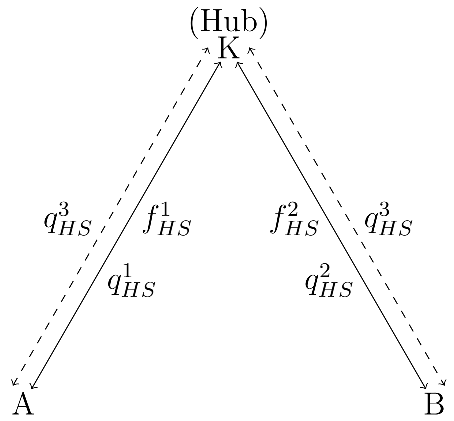

Figure 2.5: Hub-and-Spoke Network

choices give the following inverse demand functions:

$$p^i_{HS} = v + \lambda f^i_{HS} - q^i_{HS} \, (i = 1, 2), \tag{2.5}$$

$$p^3_{HS} = v + \lambda \frac{\sum_{i \in \{1,2\}} f^i_{HS}}{2} - t - q^3_{HS}. \tag{2.6}$$

In (2.5) and (2.6), p^i_{HS} and q^i_{HS} ($i \in \{1, 2, 3\}$) are the fares and traffic in markets AK, BK and AB, respectively.

2.2.1.2 Point-to-Point Network

Under PP (Figure 2.6), all airports are identical. Then, under PP, all passengers can choose non-stop services only, and their marginal utility of local traffic is $\alpha_0 = v + \lambda f^i_{PP}$, in which f^i_{PP} is the flight frequency between the city pair of market $i \, (\in \{1, 2, 3\})$.

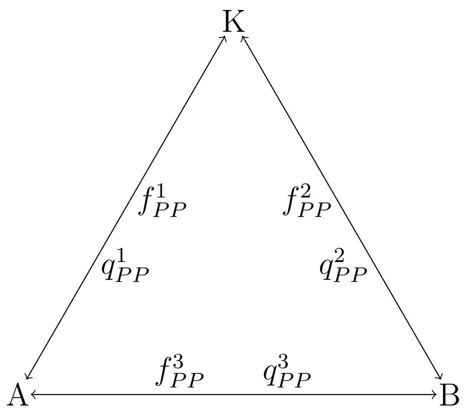

Figure 2.6: Point-to-Point Network

The inverse demand functions will then be:

$$p^i_{PP} = v + \lambda f^i_{PP} - q^i_{PP}. \tag{2.7}$$

In (2.7), p^i_{PP} and q^i_{PP} ($i \in \{1, 2, 3\}$) are the fares and traffic in markets AK, BK and AB, respectively.

2.2.1.3 Mixed Network

Under MX (Figure 2.7), K is a hub airport and A and B are spoke airports. Then, under MX, passengers in markets AK and BK can choose non-stop services only, and their marginal utility of local traffic is $\alpha^i_0 = v + \lambda f^i_{MX}$, in which f^i_{MX} is the flight frequency between the city pair of market $i\,(\in \{1, 2\})$. Moreover, the passenger in market AB can choose not only non-stop services between A and B but also one-stop services connecting at K, with the marginal utility of local traffic $\alpha^3_0 = v + \lambda f^3_{MX}$ and connecting traffic $\alpha^3_1 = v + \lambda \frac{\sum_{i \in \{1,2\}} f^i_{MX}}{2} - t$, in which f^3_{MX} is the flight frequency between the city pair of market AB.[6]

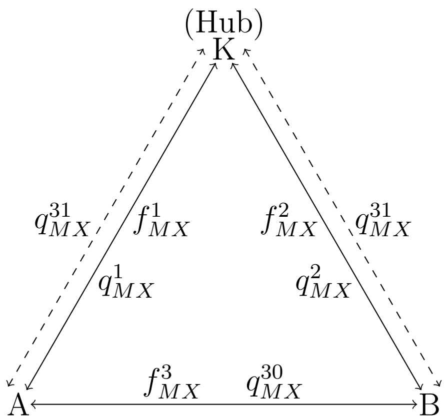

Figure 2.7: Mixed Network

[6] Under PP, theoretically, the passenger in market AB may buy tickets for routes AK and BK and thus has "one-stop services". However, the scheduling of flights in routes AK and BK can be a problem for this possibility. For example, when the passenger arrives at K, flights from K to B may have already departed. Under MX, this problem does not exist because of the coordination of flight schedules. Therefore, the scheduling of flights is a key difference between PP and MX.

The inverse demand functions will then be:

$$p^i_{MX} = v + \lambda f^i_{MX} - q^i_{MX}, \tag{2.8}$$

$$p^{30}_{MX} = v + \lambda f^3_{MX} - q^{30}_{MX} - \frac{1}{2}q^{31}_{MX}, \tag{2.9}$$

$$p^{31}_{MX} = v + \lambda \frac{\sum_{i\in\{1,2\}} f^i_{MX}}{2} - t - \frac{1}{2}q^{30}_{MX} - q^{31}_{MX}. \tag{2.10}$$

In (2.8), p^i_{MX} and q^i_{MX} $(i \in \{1,2\})$ are the fares and traffic in markets AK and BK, respectively. In (2.9) and (2.10), p^{30}_{MX}, q^{30}_{MX}, p^{31}_{MX}, and q^{31}_{MX} are the fares and traffic of local and connecting traffic in market AB, respectively. In particular, the notation $j0$ represents non-stop services and $j1$ represents one-stop services for market j. We will use this style of notation hereafter.

2.2.1.4 2-Hub Network

Under 2H (Figure 2.8), K and A are hub airports and B is a spoke airport. Then, under 2H, the passenger in market AK can choose non-stop services only, and his/her marginal utility of local traffic is $\alpha^1_0 = v + \lambda f^1_{2H}$, in which f^1_{2H} is the flight frequency between the city pair of market AK. Moreover, passengers in markets BK and AB can choose not only non-stop services between B and K and between A and B, respectively, but also one-stop services connecting at A and K, respectively, with the marginal utility of local traffic $\alpha^i_0 = v + \lambda f^i_{2H}$ and connecting traffic $\alpha^i_1 = v + \lambda \frac{\sum_{j\in\{1,2,3\}\setminus\{i\}} f^j_{2H}}{2} - t$, in which f^i_{2H} is the flight frequency between the city pair of market $i \,(\in \{2,3\})$.

The inverse demand functions will then be:

$$p^1_{2H} = v + \lambda f^1_{2H} - q^1_{2H}, \tag{2.11}$$

$$p^{i0}_{2H} = v + \lambda f^i_{2H} - q^{i0}_{2H} - \frac{1}{2}q^{i1}_{2H}, \tag{2.12}$$

$$p^{i1}_{2H} = v + \lambda \frac{\sum_{j\in\{1,2,3\}\setminus\{i\}} f^j_{2H}}{2} - t - \frac{1}{2}q^{i0}_{2H} - q^{i1}_{2H}. \tag{2.13}$$

In (2.11), p^1_{2H} and q^1_{2H} are the fare and traffic in market AK, respectively. In (2.12) and (2.13), p^{i0}_{2H}, q^{i0}_{2H}, p^{i1}_{2H}, and q^{i1}_{2H} $(i \in \{2,3\})$ are the fares and traffic of local and connecting

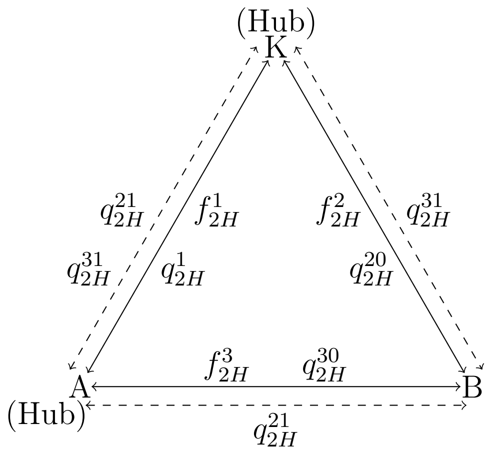

Figure 2.8: 2-Hub Network

traffic in markets BK and AB, respectively.

2.2.2 The Airline

For the airline's cost function, I assume that it includes the expected congestion delay cost only.[7] Specifically, the airline's cost is based on route, and the expected congestion delay cost of a flight in route n is proportional to the total number of aircraft movements at the origin and destination airports in that route, that is:

$$\psi \sum_{m \in \{1,2,3\}} \eta^{mn} F^m, \tag{2.14}$$

in which η^{mn} is a network-specific parameter with $\eta^{mn} \in \mathbb{N}$.[8] In (2.14), the parameter $\psi \in \mathbb{R}_{++}$ is positively related to the marginal congestion delay cost and is hereafter used as the proxy of the latter.

Next, I discuss the airline's cost for each network structure. Under HS, the numbers

[7]In fact, passengers also endure some level of disutility from congestion delays. However, because omitting the impact of congestion delays on passengers does not affect the insights of the model but makes the model more tractable, I only consider the negative effect of congestion delays on the airline.

[8]For the literature on delay function, among others, Brueckner (2002, 2005) models the delay cost as a non-decreasing function of the number of flights during the peak travel period of a day. Moreover, US Federal Aviation Administration (1969) models delays as a convex function of the number of flights. This delay function is estimated from steady-state queuing theory and has been used by Morrison (1987), Zhang and Zhang (1997, 2003, 2006) and Basso (2008). Pels and Verhoef (2004), De Borger and Van Dender (2006), Basso and Zhang (2007), and Yang and Zhang (2011) use a linear delay function. In this chapter, following Brueckner and Van Dender (2008) and Flores-Fillol (2010), I collapse the peak and off-peak travel periods, as in Brueckner (2002, 2005), into a single travel period, in which delays always exist.

of aircraft movements at airports A, B and K are f_{HS}^1, f_{HS}^2 and $f_{HS}^1 + f_{HS}^2$, respectively. Thus, the total number of aircraft movements between the city pair of market $i\,(\in \{1,2\})$ is $2f_{HS}^i + f_{HS}^j$ $(j \in \{1,2\} \setminus \{i\})$. Then, the expected congestion delay cost for each flight is $\psi\left(2f_{HS}^i + f_{HS}^j\right)$.

Under PP, the numbers of aircraft movements at airports A, B and K are $f_{PP}^1 + f_{PP}^3$, $f_{PP}^2 + f_{PP}^3$ and $f_{PP}^1 + f_{PP}^2$, respectively. Thus, the total number of aircraft movements between the city pair of market $i\,(\in \{1,2,3\})$ is $2f_{PP}^i + \sum_{j\in\{1,2,3\}\setminus\{i\}} f_{PP}^j$. Then, the expected congestion delay cost for each flight is $\psi\left(2f_{PP}^i + \sum_{j\in\{1,2,3\}\setminus\{i\}} f_{PP}^j\right)$. Note that the cost structures of MX and 2H are the same as PP.

The airline's objective is to maximize profit by choosing a network structure, flight frequencies, and traffic. For convenience of exposition, I solve the optimization problem sequentially as for a dynamic model (see the timeline in Figure 2.9). That is, given a chosen network structure, I derive the airline's optimal decision on flight frequencies and traffic. Then, I consider the airline's optimal network structure.

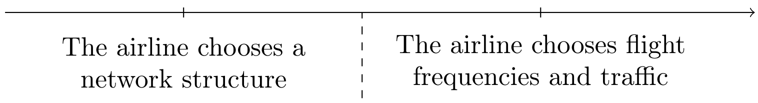

Figure 2.9: Timeline

Finally, following Brueckner (2004), I assume that the load factor of each flight equals 100 percent. In fact, the analysis will be the same if we fix a load factor realistically less than 100 percent. In this way, the aircraft size:

$$s = \frac{q}{f}, \tag{2.15}$$

is implicitly determined if the airline selects the optimal flight frequency f and traffic q.

2.3 Market Outcome

2.3.1 Flight Frequencies and Traffic

I first consider the airline's optimal decision on flight frequencies and traffic, given a chosen network structure.

2.3.1.1 Hub-and-Spoke Network

Let $\pi_{HS} : \mathbb{R}^5_+ \to \mathbb{R}^1$ be a $\mathbb{C}^2$ function of $\Omega_{HS} = (f^1_{HS}, f^2_{HS}, q^1_{HS}, q^2_{HS}, q^3_{HS})$. The airline maximizes profit subject to a non-arbitrage condition[9] and five non-negativity constraints, that is:

$$\begin{aligned} \max_{f^1_{HS}, f^2_{HS}, q^1_{HS}, q^2_{HS}, q^3_{HS}} \pi_{HS} &= \sum_{i \in \{1,2,3\}} p^i_{HS} q^i_{HS} - \sum_{i \in \{1,2\}, j \in \{1,2\} \setminus \{i\}} \psi f^i_{HS} \left(2 f^i_{HS} + f^j_{HS}\right), \\ \text{subject to} \quad & p^1_{HS} + p^2_{HS} \geqslant p^3_{HS}, \\ \text{and} \quad & f^1_{HS}, f^2_{HS}, q^1_{HS}, q^2_{HS}, q^3_{HS} \geqslant 0. \end{aligned} \tag{2.16}$$

The process of solving (2.16) is to first omit the non-arbitrage condition, then derive the Kuhn-Tucker conditions, and finally to examine the satisfaction of the omitted condition. In fact, I show in the appendix that the only candidate for the optimum is the interior solution. For the second-order condition, making the Hessian $D^2 \pi_{HS}$ negative definite for all $\Omega_{HS} \in \mathbb{R}^5_{++}$ requires $\psi \in (\psi^{soc}_{HS}, +\infty)$, in which $\psi^{soc}_{HS} = \frac{\lambda^2}{4}$. Moreover, I also show that the optimal solution is a unique global maximizer of π_{HS} on $\mathbb{R}^5_{++}$. All details of derivation and proof are in Section 2.8.1.1.

In addition, under $\psi \in (\psi^{soc}_{HS}, +\infty)$, the constraint of extra travel time disutility t to make all variables positive is $t \in \left(0, \overline{t_{HS}}\right)$, in which $\overline{t_{HS}} = \frac{12\psi v}{12\psi - \lambda^2}$, coming from $q^{3*}_{HS} > 0$. In fact, if the extra travel time disutility becomes too high, the passenger in market AB cannot obtain positive utility from choosing one-stop services, and thus the airline has no interest in providing the service, implying $q^{3*}_{HS} \to 0$.

[9] There is in fact another type of possible arbitrage, that is, passengers buy a connecting ticket but use only one portion. According to Hendricks et al. (1997) and Barla and Constantatos (2005), airlines can easily stop this practice. Thus, this type of arbitrage is not considered here.

I also compare variables within HS and find that all variables in market AK are equal to those in market BK, respectively. This result comes from the symmetry of these two markets. Moreover, I also find $q_{HS}^{i*} > q_{HS}^{3*}$ and $p_{HS}^{i*} > p_{HS}^{3*}$ ($i \in \{1,2\}$). Indeed, we can explain these results by the following first-order condition:

$$\frac{\partial \pi_{HS}}{\partial q_{HS}^i} = v + \lambda f_{HS}^i - 2q_{HS}^i = 0, \tag{2.17}$$

$$\frac{\partial \pi_{HS}}{\partial q_{HS}^3} = v + \lambda \frac{\sum_{i\in\{1,2\}} f_{HS}^i}{2} - t - 2q_{HS}^3 = 0. \tag{2.18}$$

In (2.17) and (2.18), as all passengers enjoy the same flight frequency $f_{HS}^{1*} = f_{HS}^{2*}$, $q_{HS}^{i*} > q_{HS}^{3*}$ comes directly from the fact that the passenger in market AB has to endure the extra travel time disutility and thus has a lower willingness-to-pay. Then, for the passenger in market AB, the downward shift of demand curve from the extra travel time disutility dominates the upward movement along the curve due to lower traffic. Thus, we can obtain $p_{HS}^{i*} > p_{HS}^{3*}$.

2.3.1.2 Point-to-Point Network

Let $\pi_{PP} : \mathbb{R}_+^6 \to \mathbb{R}^1$ be a $\mathbb{C}^2$ function of $\Omega_{PP} = (f_{PP}^i, q_{PP}^i)$ ($i \in \{1,2,3\}$). The airline maximizes profit subject to six non-negativity constraints, that is:

$$\begin{aligned} \max_{f_{PP}^1, f_{PP}^2, f_{PP}^3, q_{PP}^1, q_{PP}^2, q_{PP}^3} \pi_{PP} &= \sum_{i\in\{1,2,3\}} \left[p_{PP}^i q_{PP}^i - \psi f_{PP}^i \left(2f_{PP}^i + \sum_{j\in\{1,2,3\}\setminus\{i\}} f_{PP}^j \right) \right], \\ \text{subject to} \quad & f_{PP}^i, q_{PP}^i \geqslant 0 \quad (i \in \{1,2,3\}). \end{aligned} \tag{2.19}$$

The second-order condition for solving (2.19) requires $\psi \in (\psi_{PP}^{soc}, +\infty)$, in which $\psi_{PP}^{soc} = \frac{\lambda^2}{4}$. All details of derivation and proof are in Section 2.8.1.2. Moreover, under $\psi \in (\psi_{PP}^{soc}, +\infty)$, all variables are positive and symmetric across three markets.

2.3.1.3 Mixed Network

Then, let π_{MX}: $\mathbb{R}_+^7 \to \mathbb{R}^1$ be a $\mathbb{C}^2$ function of $\Omega_{MX} = (f_{MX}^1, f_{MX}^2, f_{MX}^3, q_{MX}^1, q_{MX}^2, q_{MX}^{30}, q_{MX}^{31})$. The airline maximizes profit subject to a non-arbitrage condition and seven non-negativity constraints, that is:

$$
\begin{aligned}
\max_{f^1_{MX},f^2_{MX},f^3_{MX},q^1_{MX},q^2_{MX},q^{30}_{MX},q^{31}_{MX}} \pi_{MX} &= \sum_{i\in\{1,2\}} p^i_{MX}q^i_{MX} + \sum_{i\in\{0,1\}} p^{3i}_{MX}q^{3i}_{MX} \\
&\quad - \sum_{i\in\{1,2,3\}} \psi f^i_{MX}\left(2f^i_{MX} + \sum_{j\in\{1,2,3\}\setminus\{i\}} f^j_{MX}\right), \\
\text{subject to} &\quad p^1_{MX} + p^2_{MX} \geqslant p^{31}_{MX}, \qquad (2.20) \\
\text{and} &\quad f^1_{MX}, f^2_{MX}, f^3_{MX}, q^1_{MX}, q^2_{MX}, q^{30}_{MX}, q^{31}_{MX} \geqslant 0.
\end{aligned}
$$

The second-order condition for solving (2.20) requires $\psi \in (\psi^{soc}_{MX}, +\infty)$, in which $\psi^{soc}_{MX} = \frac{(13+\sqrt{97})\lambda^2}{48}$. All details of derivation and proof are in Section 2.8.1.3.

In addition, under $\psi \in (\psi^{soc}_{MX}, +\infty)$, the constraint of extra travel time disutility t to make all variables positive is $t \in \left(\underline{t_{MX}}, \overline{t_{MX}}\right)$, in which $\underline{t_{MX}} = \frac{(4\psi+3\lambda^2)v}{20\psi-\lambda^2}$ and $\overline{t_{MX}} = \frac{8\psi v}{16\psi-\lambda^2}$, coming from $f^{3*}_{MX} > 0$ and $q^{31*}_{MX} > 0$, respectively. In fact, if the extra travel time disutility becomes too low, the disadvantage of one-stop services will become less significant, implying $f^{3*}_{MX} \to 0$. However, if the extra travel time disutility becomes too high, the passenger in market AB cannot obtain positive utility from choosing one-stop services, and thus the airline has no interest in providing the service, implying $q^{31*}_{MX} \to 0$.

I also compare variables within MX and find that all variables in market AK are equal to those in market BK, respectively. This result comes from the symmetry of these two markets. Moreover, I also find $f^{i*}_{MX} > f^{3*}_{MX}$, $q^{i*}_{MX} > q^{30*}_{MX} > q^{31*}_{MX}$ and $p^{i*}_{MX} > p^{30*}_{MX} > p^{31*}_{MX}$, as well as the comparison of aircraft sizes $s^{i*}_{MX} = \frac{q^{i*}_{MX}+q^{31*}_{MX}}{f^{i*}_{MX}} < \frac{q^{30*}_{MX}}{f^{3*}_{MX}} = s^{3*}_{MX}$ $(i \in \{1,2\})$.

We can in fact explain $f^{i*}_{MX} > f^{3*}_{MX}$ by the first-order condition:

$$
\frac{\partial \pi_{MX}}{\partial f^i_{MX}} = \lambda\left(q^i_{MX} + \frac{q^{31}_{MX}}{2}\right) - 2\psi\left(2f^i_{MX} + \sum_{j\in\{1,2,3\}\setminus\{i\}} f^j_{MX}\right) = 0, \qquad (2.21)
$$

$$
\frac{\partial \pi_{MX}}{\partial f^3_{MX}} = \lambda q^{30}_{MX} - 2\psi\left(2f^3_{MX} + \sum_{i\in\{1,2\}} f^i_{MX}\right) = 0. \qquad (2.22)
$$

According to (2.21) and (2.22),

$$f_{MX}^{i*} > f_{MX}^{3*} \Leftrightarrow q_{MX}^{i*} + \frac{q_{MX}^{31*}}{2} > q_{MX}^{30*}, \tag{2.23}$$

in which the latter holds (unsurprisingly, given the nature of traffic). That is, the dispersion of traffic in market AB makes the local traffic in market AB lower than the sum of the total traffic in market i and half of the connecting traffic in market AB.[10] Intuitively, the increase in flight frequency can raise the willingness-to-pay of passengers. Then, if the traffic effectively paying the increased fare between the city pair of market i is higher than that between the city pair of market AB, the marginal benefit of f_{MX}^{i} will be greater than that of f_{MX}^{3}, that is, $\lambda\left(q_{MX}^{i*} + \frac{q_{MX}^{31*}}{2}\right) > \lambda q_{MX}^{30*}$. Therefore, f_{MX}^{i*} should be higher than f_{MX}^{3*} to balance the marginal benefit and cost.

Analogously, $q_{MX}^{i*} > q_{MX}^{30*}$ also comes from the dispersion of traffic in market AB. This dispersion pushes the flight frequency of non-stop services for market AB down and thus reduces the willingness-to-pay of local traffic in market AB. In addition, the substitution of connecting traffic in market AB can also reduce q_{MX}^{30*}.

For $q_{MX}^{30*} > q_{MX}^{31*}$, the effect of extra travel time disutility endured by the connecting traffic dominates the more significant schedule delay reduction. Thus, the willingness-to-pay of local traffic in market AB is higher than that of connecting traffic, and then q_{MX}^{30*} will be set higher than q_{MX}^{31*}.

Moreover, the comparison of fares comes from the relationship between the shift of demand curve, for various reasons discussed above, and the movement along the demand curve due to higher or lower traffic. Finally, the comparison of aircraft sizes shows that the airline optimally uses smaller aircraft between A (B) and K in order to significantly increase flight frequency and thus reduce passengers' schedule delays.

[10] Half of the connecting traffic in market AB comes from the fact that flights between the city pair of markets AK and BK carry the connecting traffic in market AB together.

2.3.1.4 2-Hub Network

Then, let π_{2H}: $\mathbb{R}^8_+ \to \mathbb{R}^1$ be a $\mathbb{C}^2$ function of $\Omega_{2H} = (f^1_{2H}, f^2_{2H}, f^3_{2H}, q^1_{2H}, q^{20}_{2H}, q^{21}_{2H}, q^{30}_{2H}, q^{31}_{2H})$. The airline maximizes profit subject to two non-arbitrage conditions and eight non-negativity constraints, that is:

$$\begin{aligned}
\max_{f^1_{2H}, f^2_{2H}, f^3_{2H}, q^1_{2H}, q^{20}_{2H}, q^{21}_{2H}, q^{30}_{2H}, q^{31}_{2H}} \pi_{2H} &= p^1_{2H} q^1_{2H} + \sum_{j\in\{0,1\}} \sum_{i\in\{2,3\}} p^{ij}_{2H} q^{ij}_{2H} \\
&\quad - \sum_{i\in\{1,2,3\}} \psi f^i_{2H} \left(2 f^i_{2H} + \sum_{j\in\{1,2,3\}\setminus\{i\}} f^j_{2H} \right), \qquad (2.24) \\
\text{subject to} \quad & p^1_{2H} + p^{i0}_{2H} \geqslant p^{j1}_{2H} \quad (i \in \{2,3\}, j \in \{2,3\} \setminus \{i\}), \\
\text{and} \quad & f^1_{2H}, f^2_{2H}, f^3_{2H}, q^1_{2H}, q^{20}_{2H}, q^{21}_{2H}, q^{30}_{2H}, q^{31}_{2H} \geqslant 0.
\end{aligned}$$

The second-order condition for solving (2.24) requires $\psi \in (\psi^{soc}_{2H}, +\infty)$, in which $\psi^{soc}_{2H} = \frac{7\lambda^2}{12}$. All details of derivation and proof are in Section 2.8.1.4.

In addition, under $\psi \in (\psi^{soc}_{2H}, +\infty)$, the constraint of extra travel time disutility t to make all variables positive is $t \in \left(\max\left\{0, \underline{t_{2H}}\right\}, \overline{t_{2H}}\right)$, in which $\underline{t_{2H}} = -\frac{\left(4\psi - 5\lambda^2\right)v}{16\psi}$ and $\overline{t_{2H}} = \frac{8\psi v}{16\psi - \lambda^2} \left(= \overline{t_{MX}}\right)$, coming from $f^{i*}_{2H} > 0$ and $q^{i1*}_{2H} > 0$ ($i \in \{2,3\}$), respectively.

We first consider $t > \max\left\{0, \underline{t_{2H}}\right\}$. Between airports B and K (A), there exist both local and connecting traffic. Furthermore, the local traffic involves one flight, while the connecting traffic involves two flights. Thus, when the marginal congestion delay cost is high enough, that is, $\psi \in \left[\frac{5\lambda^2}{4}, +\infty\right)$, the airline will prefer the local traffic. As a result, for any extra travel time disutility, flight frequencies between airports B and K (A) are positive. When the marginal congestion delay cost is not too high, that is, $\psi \in \left(\psi^{soc}_{2H}, \frac{5\lambda^2}{4}\right)$, if the extra travel time disutility becomes lower, the local traffic will decrease because the disadvantage of one-stop services becomes less significant. However, the connecting traffic is uncertain because on the one hand, the lower extra travel time disutility will make one-stop services more attractive, while on the other hand, due to the existence of the demand side network effect, the decrease in local traffic will weaken the schedule delay reduction and subsequently make one-stop services less attractive. In this model, the effect of the decrease in local traffic dominates that of lower extra travel time disutility. Hence, if the extra travel time disutility becomes lower, the connecting traffic will also decrease. Therefore, both the decrease in local and connecting traffic resulting from the lower t will impel $f_{2H}^{i*} \to 0$.

Then, for $t < \overline{t_{2H}}$, if the extra travel time disutility is too high, passengers in markets BK and AB cannot obtain positive utility from choosing one-stop services, and thus the airline has no interest in providing the service, implying $q_{2H}^{i1*} \to 0$.

I also compare variables within 2H and find that all variables in market BK are equal to those in market AB, respectively. This result comes from the symmetry of these two markets. Moreover, I also find $f_{2H}^{1*} > f_{2H}^{i*}$, $q_{2H}^{1*} > \max\{q_{2H}^{i0*}, q_{2H}^{i1*}\}$, $p_{2H}^{1*} > \max\{p_{2H}^{i0*}, p_{2H}^{i1*}\}$, and $q_{2H}^{i0*} > q_{2H}^{i1*}$ and $p_{2H}^{i0*} > p_{2H}^{i1*}$ if and only if $t > \frac{16\lambda^2\psi v}{192\psi^2 - 52\lambda^2\psi + 3\lambda^4}$, as well as the comparison of aircraft sizes $s_{2H}^{1*} = \frac{q_{2H}^{1*} + \sum_{i\in\{2,3\}} q_{2H}^{i1*}}{f_{2H}^{1*}} < \frac{q_{2H}^{i0*} + q_{2H}^{j1*}}{f_{2H}^{i*}} = s_{2H}^{i*}$ $(i \in \{2,3\}$ and $j \in \{2,3\} \setminus \{i\})$. In fact, explanations for these comparisons are essentially the same as those in MX, except that the effect of extra travel time disutility no longer always dominates.

2.3.1.5 Comparative Statics

Table 2.1 presents the comparative statics of HS and PP. For HS, the increase in passenger's travel benefit v can raise the passenger's willingness-to-pay, and thus raise the marginal benefit of serving one more unit of traffic. The traffic will then increase with v. Moreover, higher traffic implies a higher marginal benefit of flight frequency. Then, flight frequencies will increase with v. For fares, the upward shift of demand curve from a higher travel benefit and flight frequency dominates the downward movement along the curve due to higher traffic. Then, fares will increase with v. Furthermore, as the airline can extract more surplus if passengers can enjoy more benefits from travel and passenger surplus is in fact an increasing function of traffic, the airline's profit, passenger surplus and social welfare will all increase with v. Otherwise, according to Table 2.1, under HS, flight frequency is less volatile than traffic when v changes. Thus, the aircraft size will become larger when v increases. We can then analyze the effect of extra travel time disutility t analogously.

The same logic can also apply to the effect of the marginal benefit of schedule delay reduction λ. Thus, all variables except aircraft size increase with λ. In particular, for the aircraft size under HS, a higher λ will provide the airline with a direct motivation to increase flight frequency. Thus, when λ increases, the extent to which flight frequency increases is larger than traffic and then the aircraft size will decrease.

In addition, the effect of marginal congestion delay cost ψ comes from the fact that it can raise the airline's cost. As a result, all variables except aircraft size decrease with ψ. In particular, for the aircraft size under HS, a higher ψ will provide the airline with a direct motivation to decrease flight frequency. Thus, when ψ increases, the extent that flight frequency decreases is larger than traffic and then the aircraft size will increase.

For PP, the only difference compared to HS is that the aircraft size is independent of v. This result comes from the fact that PP does not relate to t, and thus both traffic and flight frequency are proportional to v.

HS	v	t	λ	ψ	PP	v	λ	ψ
f_{HS}^{1*}	+	-	+	-	f_{PP}^{1*}	+	+	-
q_{HS}^{1*}	+	-	+	-	q_{PP}^{1*}	+	+	-
q_{HS}^{3*}	+	-	+	-	p_{PP}^{1*}	+	+	-
p_{HS}^{1*}	+	-	+	-	π_{PP}^{1*}	+	+	-
p_{HS}^{3*}	+	-	+	-	ps_{PP}^{1*}	+	+	-
π_{HS}^{*}	+	-	+	-	$3ps_{PP}^{1*}$	+	+	-
ps_{HS}^{1*}	+	-	+	-	sw_{PP}^{1*}	+	+	-
ps_{HS}^{3*}	+	-	+	-	s_{PP}^{1*}	0	-	+
$2ps_{HS}^{1*} + ps_{HS}^{3*}$	+	-	+	-				
sw_{HS}^{*}	+	-	+	-				
s_{HS}^{1*}	+	-	-	+				

Table 2.1: Comparative Statics of HS and PP

Table 2.2 presents the comparative statics of MX. All critical values in Table 2.2, as well as Table 2.3, are given in Table 2.15 in Section 2.10. For MX, compared to HS and PP, the effect of parameters is reversed on some variables. In fact, this new pattern is essentially due to the fact that MX is a network structure that involves the division of local and connecting traffic in one market.

Let us use the effect of ψ on f^3_{MX} as an example to illustrate this pattern. When ψ increases, operating flights for local traffic becomes costly and then f^3_{MX} will become lower. This mechanism shows the *direct effect* of ψ. However, in market AB, the local traffic involves one flight, while the connecting traffic involves two flights. Thus, when ψ increases, allocating more traffic to non-stop services is more cost efficient and then f^3_{MX} will become higher. This mechanism shows the *strategic effect* of ψ. Therefore, when ψ increases, the eventual change of f^3_{MX} depends on the trade-off between the direct and strategic effects. Then, if t is relatively high, that is, $t \geqslant t^5_{MX}$, the connecting traffic in market AB is already low and thus the strategic effect will be too weak to dominate the direct effect, implying that f^3_{MX} will become lower when ψ increases. However, if t is relatively low, that is, $t < t^5_{MX}$, the connecting traffic in market AB is significant and thus the strategic effect will be strong enough to dominate the direct effect, implying that f^3_{MX} will become higher when ψ increases. From the above analysis, we see that the division of traffic gives the airline more flexibility in the decision-making process.

MX	v	t	λ	ψ
f^{1*}_{MX}	+	-	+	-
f^{3*}_{MX}	-	+	+ iff $t > t^3_{MX}$	+ iff $t < t^5_{MX}$
q^{1*}_{MX}	+	-	+	-
q^{30*}_{MX}	+ iff $\psi > \frac{3\lambda^2}{4}$	+	- iff $t < t^4_{MX}$	+ iff $t < t^4_{MX}$
q^{31*}_{MX}	+	-	+	-
p^{1*}_{MX}	+	-	+	-
p^{30*}_{MX}	+ iff $\psi > \frac{7\lambda^2}{12}$	+	- iff $t < t^5_{MX}$	+ iff $t < t^5_{MX}$
p^{31*}_{MX}	+	-	+	-
π^*_{MX}	+	-	+	-
ps^{1*}_{MX}	+	-	+	-
ps^{3*}_{MX}	+ iff $t < t^1_{MX}$	+ iff $t > t^2_{MX}$	- iff $t < t^6_{MX}$	+ iff $t < t^6_{MX}$
$2ps^{1*}_{MX} + ps^{3*}_{MX}$	+	-	+	-
sw^*_{MX}	+	-	+	-
s^{1*}_{MX}	-	+	-	+
s^{3*}_{MX}	+	-	+ iff $t < t^7_{MX}$	+ iff $t > t^8_{MX}$

Table 2.2: Comparative Statics of MX

Table 2.3 presents the comparative statics of 2H and shows similar patterns to MX.

2H	v	t	λ	ψ
f_{2H}^{1*}	+	-	+	-
f_{2H}^{2*}	+ iff $\psi > \frac{5\lambda^2}{4}$	+	+ iff $t > t_{2H}^2$	+ iff $t < t_{2H}^4$
q_{2H}^{1*}	+	-	+	-
q_{2H}^{20*}	+	+	- iff $t < t_{2H}^3$	+ iff $t < t_{2H}^3$
q_{2H}^{21*}	+	-	+	-
p_{2H}^{1*}	+	-	+	-
p_{2H}^{20*}	+	+	- iff $t < t_{2H}^4$	+ iff $t < t_{2H}^4$
p_{2H}^{21*}	+	-	+	-
π_{2H}^{*}	+	-	+	-
ps_{2H}^{1*}	+	-	+	-
ps_{2H}^{2*}	+	+ iff $t > t_{2H}^1$	+	-
$ps_{2H}^{1*} + 2ps_{2H}^{2*}$	+	-	+	-
sw_{2H}^{*}	+	-	+	-
s_{2H}^{1*}	-	+	-	+
s_{2H}^{2*}	+	-	+ iff $t < t_{2H}^5$	+ iff $t > t_{2H}^6$

Table 2.3: Comparative Statics of 2H

2.3.2 Network Structure

I next consider which network structure will bring the airline the highest profit. To compare with the literature, I first focus on HS, PP and MX. I then incorporate 2H into the analysis to examine the airline's optimal network structure under four network structures. Note that in market outcome analysis, the parameter space is $\psi \in (\psi_{HS}^{soc}, +\infty)$ for comparing HS and PP, $\psi \in (\psi_{MX}^{soc}, +\infty)$ for comparing HS, PP and MX, and $\psi \in (\psi_{2H}^{soc}, +\infty)$ for comparing HS, PP, MX, and 2H.

2.3.2.1 Comparison Between HS and PP

First, I compare the important variables between HS and PP, some of which are crucial in order to understand the comparison of profits, and some, for example, the total flight frequency and aircraft size, are indeed important issues in the literature. For all of the following propositions, I will only provide the proofs that are essential to the derivation of the main conclusions in the appendix.[11] Nonetheless, I provide all critical values in Table 2.16 in Section 2.10. Letting $s_{HS}^{1*} = \frac{q_{HS}^{1*}+q_{HS}^{3*}}{f_{HS}^{1*}}$, $s_{PP}^{1*} = \frac{q_{PP}^{1*}}{f_{PP}^{1*}}$, $\Theta_{HS}^{*} = 6\psi\left(f_{HS}^{1*}\right)^2$, and $\Theta_{PP}^{*} = 12\psi\left(f_{PP}^{1*}\right)^2$, we have Proposition 2.1.

Proposition 2.1.

1. *The flight frequency under HS is higher than PP, that is,* $f_{HS}^{1*} > f_{PP}^{1*}$.

2. *The total flight frequency and total cost under HS are higher than PP, respectively, that is,* $2f_{HS}^{1*} > 3f_{PP}^{1*}$ *and* $\Theta_{HS}^{*} > \Theta_{PP}^{*}$ *if* $t \in \left(0, \frac{3v}{4}\right]$.

3. *The traffic and fare in markets AK and BK under HS are higher than PP, respectively, that is,* $q_{HS}^{1*} > q_{PP}^{1*}$ *and* $p_{HS}^{1*} > p_{PP}^{1*}$.

4. *The traffic and fare in market AB under HS are higher than PP, respectively, that is,* $q_{HS}^{3*} > q_{PP}^{1*}$ *and* $p_{HS}^{3*} > p_{PP}^{1*}$*, if and only if* $t \in (0, t_{HS\sim PP}^{1})$.

5. *The aircraft size under HS is smaller than PP, that is,* $s_{HS}^{1*} < s_{PP}^{1*}$.

6. *The cost per passenger under HS is higher than PP, that is,* $\frac{\Theta_{HS}^{*}}{2q_{HS}^{1*}+q_{HS}^{3*}} > \frac{\Theta_{PP}^{*}}{3q_{PP}^{1*}}$.

According to Proposition 2.1, generally speaking, passengers in most of the markets face higher flight frequencies and fares, while the airline faces a higher cost under HS than PP. Then, first, $f_{HS}^{1*} > f_{PP}^{1*}$ is a classic result in the literature (see Brueckner, 2004) but the interpretation is different in this model because of the existence of congestion delays. Specifically, we can also explain it by relevant first-order conditions after imposing symmetry:

$$\left.\begin{aligned} \frac{\partial \pi_{HS}}{\partial f_{HS}^{1}} &= 0 \\ f_{HS}^{1} &= f_{HS}^{2} \end{aligned}\right\} \Rightarrow \lambda\left(q_{HS}^{1*} + \frac{q_{HS}^{3*}}{2}\right) - 6\psi f_{HS}^{1*} = 0, \tag{2.25}$$

$$\left.\begin{aligned} \frac{\partial \pi_{PP}}{\partial f_{PP}^{1}} &= 0 \\ f_{PP}^{1} = f_{PP}^{2} &= f_{PP}^{3} \end{aligned}\right\} \Rightarrow \lambda q_{PP}^{1*} - 8\psi f_{PP}^{1*} = 0. \tag{2.26}$$

In (2.25) and (2.26), intuitively, the increase in flight frequency can raise the willingness-

[11] Other proofs are available upon request.

to-pay of passengers. Then, because of the pooling of traffic under HS, the traffic effectively paying the increased fare between the city pair of markets AK or BK under HS is higher than that between the city pair of any market under PP. Thus, the marginal benefit of f_{HS}^1 will be greater than that of f_{PP}^1, that is, $\lambda\left(q_{HS}^{1*}+\frac{q_{HS}^{3*}}{2}\right)>\lambda q_{PP}^{1*}$. Moreover, the existence of congestion delays makes the marginal cost of flight frequency less sensitive to flight frequency under HS than PP. Therefore, f_{HS}^{1*} should be higher than f_{PP}^{1*} in order to balance the higher marginal benefit and the lower sensitivity of marginal cost under HS.

For the second point, I compare only under the extra travel time disutility $t\in\left(0,\frac{3v}{4}\right]$, which in fact includes all relevant cases. Specifically, the higher total flight frequency under HS stands in stark contrast to one of the core assumptions in Brueckner (2004) and Fageda and Flores-Fillol (2015), in which they justify their position by arguing that under HS, the airline operates fewer routes and thus naturally provides a lower total flight frequency. In this model, because of the existence of congestion delays and the consequent lower sensitivity of marginal cost under HS, the total flight frequency of two routes under HS can be so high that it exceeds the total flight frequency of three routes under PP. Moreover, $6\psi\left(f_{HS}^{1*}\right)^2>12\psi\left(f_{PP}^{1*}\right)^2$ results directly from $2f_{HS}^{1*}>3f_{PP}^{1*}$.

Then, we can also use first-order conditions to explain the comparison of traffic:

$$\frac{\partial\pi_{HS}}{\partial q_{HS}^1}=v+\lambda f_{HS}^1-2q_{HS}^1=0, \tag{2.27}$$

$$\frac{\partial\pi_{PP}}{\partial q_{PP}^1}=v+\lambda f_{PP}^1-2q_{PP}^1=0. \tag{2.28}$$

In (2.27) and (2.28), $q_{HS}^{1*}>q_{PP}^{1*}$ comes directly from the more significant schedule delay reduction under HS. For $p_{HS}^{1*}>p_{PP}^{1*}$, compared to PP, under HS, the upward shift of demand curve from a higher flight frequency dominates the downward movement along the curve due to higher traffic. In any case, we can interpret the comparison between q_{HS}^{3*} and q_{PP}^{1*} and between p_{HS}^{3*} and p_{PP}^{1*} analogously, except that the extra travel time disutility can significantly reduce the willingness-to-pay of the passenger in market AB.

In particular, what is also surprising is $s_{HS}^{1*}<s_{PP}^{1*}$, which is totally contrary to the

previous literature (see Brueckner (2004)). Conventional wisdom suggests that HS can efficiently employ the economies of traffic density by using larger aircraft in order to save cost. However, results in this model show that HS may also be more profitable by significantly increasing flight frequency and thus reducing the schedule delays of passengers. In this way, using smaller aircraft can be the airline's optimal strategy. Finally, another consequence of the significantly higher flight frequency under HS is the higher cost per passenger shown in the last point, which is also contrary to the literature.

Then, the comparison between π^*_{HS} and π^*_{PP} gives Proposition 2.2.

Proposition 2.2. *Under $\psi \in (\psi^{soc}_{HS}, +\infty)$, for the airline, HS is more profitable than PP if and only if $t \in (0, t^{\pi}_{HS\sim PP}]$.*

Proof. In Section 2.8.2.1. □

Proposition 2.2 is shown in Figure 2.10. To interpret this, we decompose the difference of π^*_{HS} and π^*_{PP} as the difference of schedule delay reduction, extra travel time disutility, the rest of margin, and total cost, which are positive, positive, negative, and positive, respectively, if the extra travel time disutility $t \in \left(0, \frac{3v}{4}\right]$:

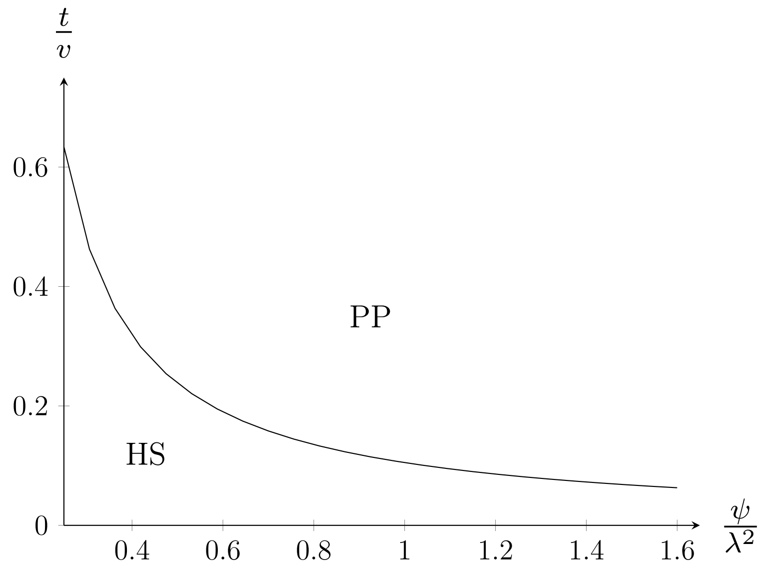

Figure 2.10: Comparison of Profits Between HS and PP

$$\pi_{HS}^{*} - \pi_{PP}^{*} = \underbrace{\kappa_{HS}^{SDR} - \kappa_{PP}^{SDR}}_{\text{Difference of Schedule Delay Reduction}} - \underbrace{tq_{HS}^{2*}}_{\text{Difference of Extra Travel Time Disutility}} + \underbrace{\kappa_{HS}^{RM} - \kappa_{PP}^{RM}}_{\text{Difference of the Rest of Margin}} - \underbrace{\left(\kappa_{HS}^{TC} - \kappa_{PP}^{TC}\right)}_{\text{Difference of Total Cost}}. \tag{2.29}$$

in which:

$$\kappa_{HS}^{SDR} = \lambda \sum_{i\in\{1,2,3\}} f_{HS}^{1*} q_{HS}^{i*}, \quad \kappa_{HS}^{RM} = \sum_{i\in\{1,2,3\}} \left(v - q_{HS}^{i*}\right) q_{HS}^{i*}, \quad \kappa_{HS}^{TC} = 6\psi \left(f_{HS}^{1*}\right)^{2},$$
$$\kappa_{PP}^{SDR} = \lambda \sum_{i\in\{1,2,3\}} f_{PP}^{1*} q_{PP}^{i*}, \quad \kappa_{PP}^{RM} = \sum_{i\in\{1,2,3\}} \left(v - q_{PP}^{i*}\right) q_{PP}^{i*}, \quad \kappa_{PP}^{TC} = 12\psi \left(f_{PP}^{1*}\right)^{2}.$$

Then, for given values of the passenger's travel benefit v and the marginal benefit of schedule delay reduction λ, I will discuss the comparison between π_{HS}^{*} and π_{PP}^{*} by considering the marginal congestion delay cost ψ and the extra travel time disutility t. Let us first fix ψ and explain the choice of network structure if t increases. According to Proposition 2.2, given a marginal congestion delay cost, HS is more profitable if the extra travel time disutility is low and PP is more profitable otherwise. Under HS, if the extra travel time disutility is low, the passenger in market AB does not have to endure a considerable loss when connecting at the hub, and thus there will be high traffic of one-stop services. Meanwhile, according to (2.25), the high traffic in market AB will increase the marginal benefit of flight frequency and consequently motivate the airline to set a high flight frequency. Furthermore, as in (2.27), a high flight frequency will enable passengers in markets AK and BK to enjoy a more significant schedule delay reduction, and as a result there will also be high traffic of non-stop services. In fact, this process shows the mechanism of the demand side network effect.

Therefore, if the extra travel time disutility t is low, even though HS endures the loss of passengers, which is due first to the existence of extra travel time disutility that does not relate to PP, second, as it incurs a higher total cost than PP, and third, as it obtains a lower value of the rest of margin, as in (2.29), it can still exert its advantage of reducing schedule delays sufficiently and will thus be more profitable than PP. However, if the extra travel time disutility becomes high, connecting at the hub will greatly reduce the utility of the passenger in market AB, and the benefit from reducing schedule delays will also become very weak. In this way, HS cannot be more profitable.

Let us then fix t and explain the choice of network structure if ψ increases. Obviously, if the extra travel time disutility t is high enough, PP is always more profitable than HS, as discussed above. Nevertheless, given the relatively low extra travel time disutility, as the marginal congestion delay cost ψ increases, first, HS is more profitable and then PP is more profitable. In fact, a relatively low marginal congestion delay cost will motivate the airline under HS to set high flight frequency in order to sufficiently enjoy the advantage of schedule delay reduction. The advantage of schedule delay reduction of HS will dominate the three disadvantages, and thus HS will be more profitable. However, when the marginal congestion delay cost becomes high, maintaining high flight frequency becomes more costly. This change will motivate the airline under HS to reduce flight frequency, and the advantage of schedule delay reduction of HS will be dominated by the three disadvantages. Hence, PP will be more profitable.

2.3.2.2 Comparison Between HS (PP) and MX

I also compare the important variables of HS and PP with those of MX. Letting $\Theta_{MX}^{*} = 2\psi\left[3\left(f_{MX}^{1*}\right)^{2} + 2f_{MX}^{1*}f_{MX}^{3*} + \left(f_{MX}^{3*}\right)^{2}\right]$, we have Proposition 2.3.

Proposition 2.3. *Table 2.4 summarizes the comparison of variables between HS (PP) and MX.*

According to Proposition 2.3, compared to HS and PP, passengers in markets AK and BK under MX face an intermediate flight frequency and fare, while the airline under MX faces an intermediate cost. This result arises due to the fact that MX is a combination of HS and PP. Thus, variables under MX are not as high as HS and not as low as PP.

Specifically, results in Proposition 2.3 can essentially be explained by the same approach as in Proposition 2.1. Thus, only the comparison of aircraft sizes between HS and MX will be discussed. For MX, a central problem is to allocate the traffic of market AB

	HS *vs* MX		PP *vs* MX	
Markets	AK/BK	AB	AK/BK	AB
Flight Frequency	$f_{HS}^{1*} > f_{MX}^{1*}$	$f_{HS}^{1*} > f_{MX}^{1*} > f_{MX}^{3*}$	$f_{PP}^{1*} < f_{MX}^{1*}$	$f_{MX}^{3*} < f_{PP}^{1*} < f_{MX}^{1*}$
Traffic	$q_{HS}^{1*} > q_{MX}^{1*}$	$q_{MX}^{31*} < q_{HS}^{3*} < q_{MX}^{30*}$	$q_{PP}^{1*} < q_{MX}^{1*}$	$q_{PP}^{1*} > q_{MX}^{30*} > q_{MX}^{31*}$
Fare	$p_{HS}^{1*} > p_{MX}^{1*}$	$p_{MX}^{31*} < p_{HS}^{3*} < p_{MX}^{30*}$	$p_{PP}^{1*} < p_{MX}^{1*}$	$p_{PP}^{1*} > p_{MX}^{30*} > p_{MX}^{31*}$
Total Flight Frequency	$2f_{HS}^{1*} < 2f_{MX}^{1*} + f_{MX}^{3*}$		$3f_{PP}^{1*} < 2f_{MX}^{1*} + f_{MX}^{3*}$	
Total Cost	$\Theta_{HS}^{*} > \Theta_{MX}^{*}$		$\Theta_{PP}^{*} < \Theta_{MX}^{*}$	
Cost per Passenger	$\frac{\Theta_{HS}^{*}}{2q_{HS}^{1*}+q_{HS}^{3*}} > \frac{\Theta_{MX}^{*}}{2q_{MX}^{1*}+q_{MX}^{30*}+q_{MX}^{31*}}$		$\frac{\Theta_{PP}^{*}}{3q_{PP}^{1*}} < \frac{\Theta_{MX}^{*}}{2q_{MX}^{1*}+q_{MX}^{30*}+q_{MX}^{31*}}$	
Aircraft Size	$s_{HS}^{1*} < s_{MX}^{1*}$ iff $t > t_{HS\sim MX}^{1}, s_{HS}^{1*} < s_{MX}^{3*}$		$s_{MX}^{3*} > s_{PP}^{1*} > s_{MX}^{1*}$	

Table 2.4: Comparison of Variables Between HS (PP) and MX

between non-stop and one-stop services. Specifically, if the extra travel time disutility t is low, the airline has a stronger motivation to attract passengers to choose one-stop services. As a result, between city pairs of markets AK and BK, using small aircraft to increase flight frequency and then to reduce passengers' schedule delays is the airline's optimal strategy. We then obtain $s_{HS}^{1*} \geqslant s_{MX}^{1*}$. Otherwise, such motivation is weak and thus $s_{HS}^{1*} < s_{MX}^{1*}$ is possible. Moreover, because of the limited scale of non-stop services of market AB under MX, using larger aircraft to decrease flight frequency and reduce congestion delay cost is the airline's optimal strategy. We then obtain $s_{HS}^{1*} < s_{MX}^{3*}$.

The comparison between π_{HS}^{*} (π_{PP}^{*}) and π_{MX}^{*} gives Proposition 2.4.

Proposition 2.4. *Under $\psi \in (\psi_{MX}^{soc}, +\infty)$, for the airline:*

1. HS is less (resp. more) profitable than MX if $t \in \left(\underline{t_{MX}}, \overline{t_{MX}}\right)$ (resp. $t \in \left(0, \underline{t_{MX}}\right] \bigcup \left[\overline{t_{MX}}, \overline{t_{HS}}\right)$);

2. PP is less profitable than MX if and only if $t \in \left(\underline{t_{MX}}, \overline{t_{MX}}\right)$.

Proof. In Section 2.8.2.2. □

Figures of Proposition 2.4 are in Section 2.8.2.2. According to Proposition 2.4, without considering the airline's fixed investments of developing a hub airport, MX will dominate HS and PP as long as it is feasible.[12] Because non-stop and one-stop services are imperfectly substitutable, passengers can obtain higher utility if both non-stop and one-stop

[12]The reason I do not model the airline's fixed investments of developing a hub airport is the concern of tractability. However, in spite of this limitation, we can gain a better understanding of the role of

services are available to choose than if only one of them is available. Under MX, there is one market, in which both non-stop and one-stop services are available. However, under HS and PP, passengers in any market cannot choose between non-stop and one-stop services. Therefore, because of the exploitation of horizontal product differentiation, MX can generate higher passenger utility and then a higher airline profit than HS and PP as long as it is feasible. Moreover, when MX is no longer feasible, HS and PP are the only available choices. To summarize, this result shows the role of horizontal product differentiation in improving passenger utility and airline profit.

2.3.2.3 The Airline's Optimal Network Structure (HS, PP and MX)

I then compare π^*_{HS}, π^*_{PP}, and π^*_{MX} with each other and provide the airline's optimal network structure from amongst HS, PP and MX in Proposition 2.5.

Proposition 2.5. *Considering only HS, PP and MX, under* $\psi \in (\psi^{soc}_{MX}, +\infty)$*, the airline's optimal network structure is HS if* $t \in (0, t^{\pi}_{HS\sim PP}]$*, PP if* $t \in \left(t^{\pi}_{HS\sim PP}, \underline{t_{MX}}\right] \bigcup \left[\overline{t_{MX}}, +\infty\right)$ *and MX if* $t \in \left(\underline{t_{MX}}, \overline{t_{MX}}\right)$*.*

Proof. In Section 2.8.2.3. □

Proposition 2.5 is shown in Figure 2.11. According to the intuition from Proposition 2.2 and 2.4, if the extra travel time disutility t is low, focusing on one-stop services in market AB can sufficiently use the schedule delay reduction without losing a lot of traffic, and thus HS will be the airline's optimal network structure. However, if the extra travel time disutility is high, focusing on non-stop services in market AB can efficiently keep the traffic, and thus PP will be the airline's optimal network structure.

For the rest of the region, MX (resp. PP) will be the airline's optimal network structure if the extra travel time disutility t is relatively high (resp. low). This result is somewhat contrary to the conventional wisdom that an airline prefers PP only if the extra travel time disutility is high (see Lin (2012)). In fact, the region we are discussing is the

horizontal product differentiation in airline network choice from this chapter. In fact, if we model the fixed investments, the network structures with fewer hubs will be favored more by the airline, while all of the properties and intuitions found in this chapter will not be affected.

intermediate extra travel time disutility, which is not high enough to greatly reduce the utility of the passenger in market AB who connects at the hub. Intuitively, if the extra travel time disutility is relatively high, MX is feasible. Then, because of the exploitation of horizontal product differentiation, MX will be the airline's optimal network structure. However, if the extra travel time disutility is relatively low, the flight frequency between the city pair of market AB under MX cannot be positive, while HS cannot also be more profitable than PP. Thus, PP will be the airline's optimal network structure. I will then explain why, in this model, PP can be the airline's optimal network structure even when the extra travel time disutility is relatively low, while it cannot be in Lin (2012).

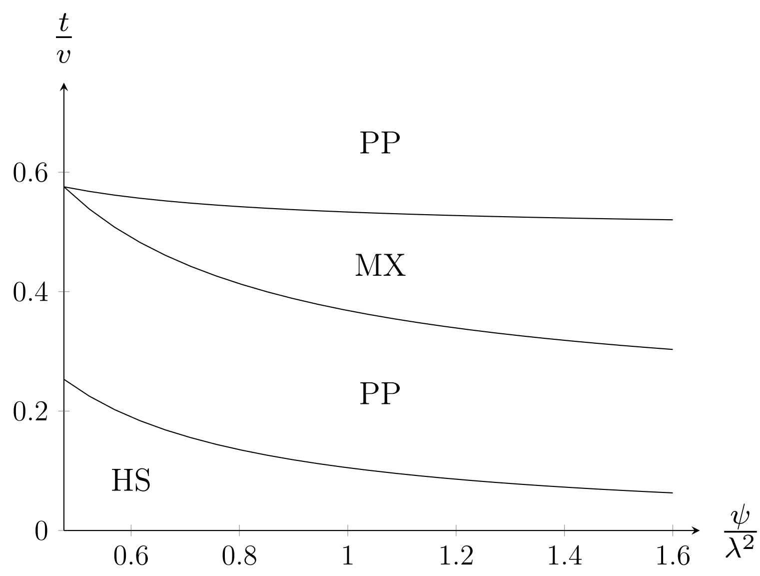

Figure 2.11: The Airline's Optimal Network Structure (HS, PP and MX)

In Lin (2012), letting $\widehat{t}^{\pi}_{HS\sim PP}$ denote the indifference curve of HS and PP, $\overline{\widehat{t}_{HS}}$ denote the upper bound of extra travel time disutility $\widehat{t}$ of HS and $\underline{\widehat{t}_{MX}}$ and $\overline{\widehat{t}_{MX}}$ denote the lower and upper bounds of $\widehat{t}$ of MX, respectively, HS is more profitable than PP if and only if $\widehat{t} \in \left(0, \widehat{t}^{\pi}_{HS\sim PP}\right]$. HS is less (resp. more) profitable than MX if $\widehat{t} \in \left(\underline{\widehat{t}_{MX}}, \overline{\widehat{t}_{MX}}\right)$ (resp. $\widehat{t} \in \left(0, \underline{\widehat{t}_{MX}}\right] \bigcup \left[\overline{\widehat{t}_{MX}}, \overline{\widehat{t}_{HS}}\right)$). Moreover, PP is less profitable than MX if and only if $\widehat{t} \in \left(\underline{\widehat{t}_{MX}}, \overline{\widehat{t}_{MX}}\right)$. Then, given the order of critical values $0 < \underline{\widehat{t}_{MX}} < \widehat{t}^{\pi}_{HS\sim PP} < \overline{\widehat{t}_{MX}} < \overline{\widehat{t}_{HS}}$, the airline's optimal network structure is HS if $t \in \left(0, \underline{\widehat{t}_{MX}}\right]$, PP if $t \in \left[\overline{\widehat{t}_{MX}}, +\infty\right)$ and MX if $t \in \left(\underline{\widehat{t}_{MX}}, \overline{\widehat{t}_{MX}}\right)$.

Therefore, as shown in Proposition 2.5, the new result in this model comes from the fact that the indifference curve of HS and PP is below the lower bound of MX, instead of between the bounds of MX. In fact, the different relative positions of curves and bounds result essentially from different cost functions. Without considering congestion delays, as in Lin (2012), the airline's cost for one route depends on the flight frequency of this route only. However, in this model, in considering congestion delays, that is, a negative network externality, the airline's cost for one route depends on not only the flight frequency of this route but also those of other routes. Next, I analyze how different cost functions affect the indifference curve of HS and PP and the bounds of MX.

Let us first rewrite the airline's cost function under HS and PP as:

$$\psi\left[\sum_{i\in\{1,2\}}\left(f^i_{HS}\right)^2+\delta_{HS}\left(\sum_{i\in\{1,2\}}f^i_{HS}\right)^2\right], \tag{2.30}$$

$$\psi\left[\sum_{i\in\{1,2,3\}}\left(f^i_{PP}\right)^2+\delta_{PP}\left(\sum_{i\in\{1,2,3\}}f^i_{PP}\right)^2\right], \tag{2.31}$$

respectively. In this way, $\delta_{HS}=\delta_{PP}=0$ is the case of Lin (2012) and $\delta_{HS}=\delta_{PP}=1$ is the case of this model. Given (2.30) and (2.31), we can obtain the general form of indifference curve of HS and PP:

$$\begin{aligned}\widetilde{t}^{\pi}_{HS\sim PP}=&\frac{v}{4\psi\left(1+2\delta_{HS}\right)-\lambda^2}\left\{4\psi\left(1+2\delta_{HS}\right)\right.\\&\left.-\sqrt{\frac{2\psi\left[8\psi\left(1+2\delta_{HS}\right)-3\lambda^2\right]\left[4\psi\left(1+2\delta_{HS}\right)\left(1+3\delta_{PP}\right)-\lambda^2\left(1-4\delta_{HS}+9\delta_{PP}\right)\right]}{4\psi\left(1+3\delta_{PP}\right)-\lambda^2}}\right\}.\end{aligned} \tag{2.32}$$

Then, we find that:

$$\frac{\partial\widetilde{t}^{\pi}_{HS\sim PP}}{\partial\delta_{HS}}<0, \tag{2.33}$$

$$\frac{\partial\widetilde{t}^{\pi}_{HS\sim PP}}{\partial\delta_{PP}}>0, \tag{2.34}$$

hold for the appropriate parameter space. Specifically, for (2.33), a higher δ_{HS} will disadvantage HS in cost and thus make the region in which HS is more profitable smaller, that is, $\widetilde{t}^{\pi}_{HS\sim PP}$ moves downwards. However, for (2.34), a higher δ_{PP} will disadvantage PP in cost, or equivalently advantage HS in cost. Thus, the region in which HS is more profitable will become larger, that is, $\widetilde{t}^{\pi}_{HS\sim PP}$ moves upwards. Then, the change of $\widetilde{t}^{\pi}_{HS\sim PP}$ depends crucially on the comparison of the effects of δ_{HS} and δ_{PP}. In fact, under the general framework of (2.30), (2.31) and (2.32), when $(\delta_{HS}, \delta_{PP})$ changes from $(0, 0)$ to $(1, 1)$, the effect of δ_{HS} on $\widetilde{t}^{\pi}_{HS\sim PP}$ dominates that of δ_{PP}, that is, the indifference curve moves downwards. To gain a more concrete understanding of these partial derivatives, we look at a numerical example.

Example 2.1. Setting $\frac{\psi}{\lambda^2} = \frac{7}{10}$ and changing δ_{HS} and δ_{PP} from 0 through 1, as shown in Figure 2.12, given δ_{PP}, we can observe that $\frac{\widetilde{t}^{\pi}_{HS\sim PP}}{v}$ decreases with δ_{HS}, and given δ_{HS}, we can observe that $\frac{\widetilde{t}^{\pi}_{HS\sim PP}}{v}$ increases with δ_{PP}. Moreover, according to (2.35):

$$\frac{\widetilde{t}^{\pi}_{HS\sim PP}}{v}\Big|_{(\delta_{HS},\delta_{PP})=(0,0)} > \frac{\widetilde{t}^{\pi}_{HS\sim PP}}{v}\Big|_{(\delta_{HS},\delta_{PP})=(1,1)}, \tag{2.35}$$

when $(\delta_{HS}, \delta_{PP})$ changes from $(0, 0)$ to $(1, 1)$, the indifference curve moves downwards.

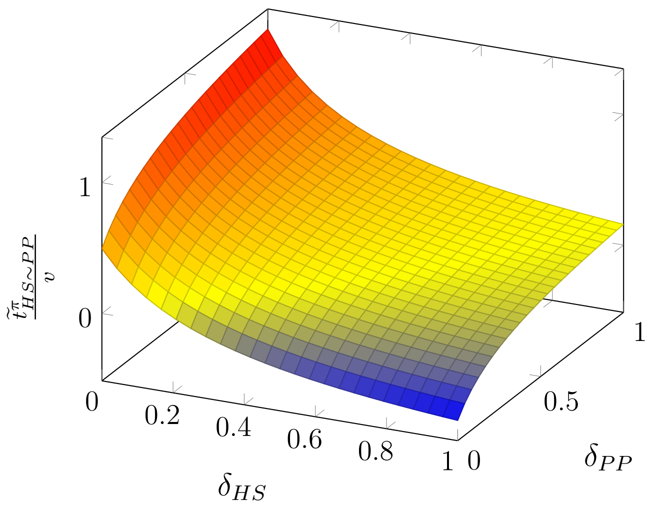

Figure 2.12: $\frac{\widetilde{t}^{\pi}_{HS\sim PP}}{v}$ when $\frac{\psi}{\lambda^2} = \frac{7}{10}$

Let us then consider the bounds of MX. The total cost with and without considering congestion delays is:

$$\psi \sum_{i\in\{1,2,3\}} f^i_{MX} \left(2f^i_{MX} + \sum_{j\in\{1,2,3\}\setminus\{i\}} f^j_{MX} \right), \tag{2.36}$$

$$\psi \sum_{i \in \{1,2,3\}} \left(f_{MX}^{i}\right)^2, \tag{2.37}$$

respectively. Changing from (2.37) to (2.36), the effect of flight frequency on the total cost is amplified. Consequently, compared to Lin (2012), f_{MX}^{3*} and q_{MX}^{31*} in this model are lower.[13] Then, because the flight frequency between the city pair of market AB under MX increases with the extra travel time disutility t, f_{MX}^{3*} in this model will first become zero if t decreases. Analogously, because the connecting traffic in market AB under MX decreases with the extra travel time disutility t, q_{MX}^{31*} in this model will first become zero if t increases. Therefore, the bounds of MX in this model are tighter.

To summarize, as shown in Figure 2.13, given the tighter bounds of MX and the lower indifference curve of HS and PP in this model, PP could be the airline's optimal network structure even when the extra travel time disutility is relatively low. In fact, the key insight from Proposition 2.5 is that including the omitted negative network externality makes HS and MX less effective than previously understood as in Lin (2012).

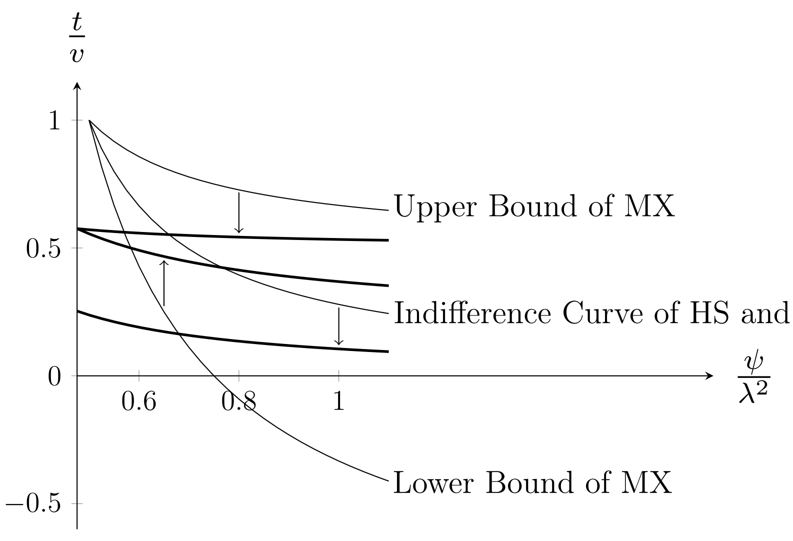

Figure 2.13: Moves of Curve and Bounds after Introducing Congestion Delays

[13]Compared to Lin (2012), f_{MX}^{3*} in this model is in fact higher when $\psi < \frac{7\lambda^2}{12}$, which, however, does not affect the result.

2.3.2.4 Comparison Between HS (PP, MX) and 2H

Next, I incorporate 2H into the analysis. As before, I first compare the important variables of these network structures. Letting $\Theta^*_{2H} = 2\psi\left[\left(f^{1*}_{2H}\right)^2 + 2f^{1*}_{2H}f^{2*}_{2H} + 3\left(f^{2*}_{2H}\right)^2\right]$, we have Proposition 2.6.

Proposition 2.6. *Table 2.5 (resp. Table 2.6, Table 2.7) summarizes the comparison of variables between HS (resp. PP, MX) and 2H.*

HS *vs* 2H			
Markets	AK (iff $t > t^1_{HS\sim 2H}$)	BK	AB
Flight Frequency	$f^{1*}_{HS} > f^{1*}_{2H}$	$f^{1*}_{HS} > \frac{f^{1*}_{2H}+f^{2*}_{2H}}{2} > f^{2*}_{2H}$	$f^{1*}_{HS} > \frac{f^{1*}_{2H}+f^{2*}_{2H}}{2} > f^{2*}_{2H}$
Traffic	$q^{1*}_{HS} > q^{1*}_{2H}$	$q^{1*}_{HS} > q^{20*}_{2H}, q^{1*}_{HS} > q^{21*}_{2H}$	$q^{3*}_{HS} < q^{20*}_{2H}$ iff $t > t^2_{HS\sim 2H}, q^{3*}_{HS} > q^{21*}_{2H}$
Fare	$p^{1*}_{HS} > p^{1*}_{2H}$	$p^{1*}_{HS} > p^{20*}_{2H}, p^{1*}_{HS} > p^{21*}_{2H}$	$p^{3*}_{HS} < p^{20*}_{2H}$ iff $t > t^3_{HS\sim 2H}, p^{3*}_{HS} > p^{21*}_{2H}$
Total Flight Frequency	$2f^{1*}_{HS} > f^{1*}_{2H} + 2f^{2*}_{2H}$	Total Cost	$\Theta^*_{HS} > \Theta^*_{2H}$ iff $t > t^4_{HS\sim 2H}$
Cost per Passenger	$\frac{\Theta^*_{HS}}{2q^{1*}_{HS}+q^{3*}_{HS}} > \frac{\Theta^*_{2H}}{q^{1*}_{2H}+2q^{20*}_{2H}+2q^{21*}_{2H}}$	Aircraft Size	$s^{1*}_{HS} < s^{1*}_{2H}$ iff $t > t^5_{HS\sim 2H}, s^{1*}_{HS} < s^{2*}_{2H}$

Table 2.5: Comparison of Variables Between HS and 2H

PP *vs* 2H			
Markets	AK	BK/AB	
Flight Frequency	$f^{1*}_{PP} < f^{1*}_{2H}$	$\frac{f^{1*}_{2H}+f^{2*}_{2H}}{2} > f^{1*}_{PP} > f^{2*}_{2H}$	
Traffic	$q^{1*}_{PP} < q^{1*}_{2H}$	$q^{1*}_{PP} > q^{20*}_{2H}, q^{1*}_{PP} > q^{21*}_{2H}$	
Fare	$p^{1*}_{PP} < p^{1*}_{2H}$	$p^{1*}_{PP} > p^{20*}_{2H}, p^{1*}_{PP} > p^{21*}_{2H}$ iff $t > t^1_{PP\sim 2H}$	
Total Flight Frequency	$3f^{1*}_{PP} < f^{1*}_{2H} + 2f^{2*}_{2H}$	Total Cost	$\Theta^*_{PP} < \Theta^*_{2H}$
Cost per Passenger	$\frac{\Theta^*_{PP}}{3q^{1*}_{PP}} < \frac{\Theta^*_{2H}}{q^{1*}_{2H}+2q^{20*}_{2H}+2q^{21*}_{2H}}$	Aircraft Size	$s^{2*}_{2H} > s^{1*}_{PP} > s^{1*}_{2H}$

Table 2.6: Comparison of Variables Between PP and 2H

According to Proposition 2.6, generally speaking, compared to the three other network structures, the passenger in market AK under 2H faces a relatively high flight frequency and fare, while the airline under 2H faces a relatively high cost. This result comes mainly from the fact that the airline gathers high traffic between the city pair of market AK under 2H. Moreover, specific intuitions are the same as those from previous analysis.

Then, the comparison between $\pi^*_{HS}\,(\pi^*_{PP}, \pi^*_{MX})$ and π^*_{2H} gives Proposition 2.7.

MX *vs* 2H			
Markets	AK	BK	AB
Flight Frequency	$f_{MX}^{1*} < f_{2H}^{1*}$	$f_{MX}^{1*} > \frac{f_{2H}^{1*}+f_{2H}^{2*}}{2} > f_{2H}^{2*}$	$f_{MX}^{1*} > \frac{f_{2H}^{1*}+f_{2H}^{2*}}{2} > f_{2H}^{2*} > f_{MX}^{3*}$
Traffic	$q_{MX}^{1*} < q_{2H}^{1*}$	$q_{MX}^{1*} > q_{2H}^{20*} > q_{2H}^{21*}$	$q_{2H}^{20*} > q_{MX}^{30*} > q_{MX}^{31*} > q_{2H}^{21*}$
Fare	$p_{MX}^{1*} < p_{2H}^{1*}$	$p_{MX}^{1*} > p_{2H}^{20*} > p_{2H}^{21*}$	$p_{2H}^{20*} > p_{MX}^{30*} > p_{MX}^{31*} > p_{2H}^{21*}$
Total Flight Frequency	$2f_{MX}^{1*} + f_{MX}^{3*} < f_{2H}^{1*} + 2f_{2H}^{2*}$	Total Cost	$\Theta_{MX}^{*} < \Theta_{2H}^{*}$ iff $t > t_{MX\sim 2H}^{1}$
Cost per Passenger	$\frac{\Theta_{MX}^{*}}{2q_{MX}^{1*}+q_{MX}^{30*}+q_{MX}^{31*}} < \frac{\Theta_{2H}^{*}}{q_{2H}^{1*}+2q_{2H}^{20*}+2q_{2H}^{21*}}$ iff $t > t_{MX\sim 2H}^{2}$		
Aircraft Size	$s_{MX}^{3*} > s_{2H}^{2*} > s_{MX}^{1*} > s_{2H}^{1*}$		

Table 2.7: Comparison of Variables Between MX and 2H

Proposition 2.7. *Under $\psi \in (\psi_{2H}^{soc}, +\infty)$, for the airline:*

1. *HS is less (resp. more) profitable than 2H if $t \in \left(\max\left\{0, \underline{t_{2H}}\right\}, \overline{t_{2H}}\right)$ (resp. $t \in \left(0, \underline{t_{2H}}\right] \bigcup \left[\overline{t_{2H}}, \overline{t_{HS}}\right)$ when $\psi \in \left(\psi_{2H}^{soc}, \frac{5\lambda^2}{4}\right)$ and $t \in \left[\overline{t_{2H}}, \overline{t_{HS}}\right)$ when $\psi \in \left[\frac{5\lambda^2}{4}, +\infty\right)$);*
2. *PP is less profitable than 2H if and only if $t \in \left(\max\left\{0, \underline{t_{2H}}\right\}, \overline{t_{2H}}\right)$;*
3. *MX is always less profitable than 2H.*

Proof. In Section 2.8.2.4. □

Figures of Proposition 2.7 are in Section 2.8.2.4. First of all, without considering the airline's fixed investments of developing a hub airport, 2H dominates the three other network structures within its feasible region $t \in \left(\max\left\{0, \underline{t_{2H}}\right\}, \overline{t_{2H}}\right)$. This result essentially arises from the exploitation of horizontal product differentiation to a larger extent in 2H. That is, compared to HS, PP and MX, 2H allows passengers in more markets to choose between imperfectly substitutable non-stop and one-stop services. Moreover, HS and PP will be more profitable than 2H when the latter is infeasible. However, MX can never be more profitable than 2H, as $\underline{t_{MX}} > \underline{t_{2H}}$ and $\overline{t_{MX}} = \overline{t_{2H}}$.

To explain $\underline{t_{MX}} > \underline{t_{2H}}$, first, we know that $\underline{t_{MX}}$ and $\underline{t_{2H}}$ come from $f_{MX}^{3*} = 0$ and $f_{2H}^{i*} = 0\,(i \in \{2, 3\})$, respectively. Then, because f_{MX}^{3*} only carries local traffic between A and B but f_{2H}^{i*} carries both the local traffic in market i and the connecting traffic in market j $(i \in \{2, 3\}, j \in \{2, 3\} \setminus \{i\})$, f_{MX}^{3*} will naturally be lower than f_{2H}^{i*}.

Moreover, we can show that both f_{MX}^{3*} and f_{2H}^{i*} increase with the extra travel time disutility t. Intuitively, under MX, the higher extra travel time disutility can make the airline allocate more traffic to non-stop services in market AB, and thus, a higher marginal benefit of f_{MX}^{3*} will motivate the airline to increase f_{MX}^{3*}. However, for f_{2H}^{i*}, because the marginal benefit of f_{2H}^{i*} is an increasing function of q_{2H}^{i0} and q_{2H}^{j1}, there are two opposite effects if the extra travel time disutility becomes higher. On the one hand, the higher extra travel time disutility can make the airline allocate more traffic to non-stop services in markets i and j. Then, a higher marginal benefit of f_{2H}^{i*} will motivate the airline to increase f_{2H}^{i*}. On the other hand, the higher extra travel time disutility can also make the airline allocate less traffic to one-stop services in markets i and j. Then, a lower marginal benefit of f_{2H}^{i*} will motivate the airline to decrease f_{2H}^{i*}. Furthermore, as only f_{2H}^{i*} carries q_{2H}^{i0}, the whole q_{2H}^{i0} affects the marginal benefit of f_{2H}^{i*}. Nonetheless, as f_{2H}^{1*} and f_{2H}^{i*} carry q_{2H}^{j1} together, only half of q_{2H}^{j1} affects the marginal benefit of f_{2H}^{i*}. Then, the first effect dominates the second, and thus f_{2H}^{i*} also increases with the extra travel time disutility.

To summarize, as f_{MX}^{3*} is lower than f_{2H}^{i*}, and both f_{MX}^{3*} and f_{2H}^{i*} increase with the extra travel time disutility t, compared to f_{2H}^{i*}, f_{MX}^{3*} will first become zero if t decreases. We can then obtain $\underline{t_{MX}} > \underline{t_{2H}}$.

Furthermore, in order to obtain that MX can never be more profitable than 2H, we must also have $\overline{t_{MX}} \leqslant \overline{t_{2H}}$. In fact, we can show it by contradiction. Suppose there exists a marginal congestion delay cost $\widetilde{\psi}$ such that $\underline{t_{MX}}\left(\widetilde{\psi}\right) < \overline{t_{2H}}\left(\widetilde{\psi}\right) < \overline{t_{MX}}\left(\widetilde{\psi}\right)$. Thus, we have $\pi_{MX}^{*}\left(\overline{t_{2H}}\left(\widetilde{\psi}\right), \widetilde{\psi}\right) > \pi_{PP}^{*}\left(\widetilde{\psi}\right) = \pi_{2H}^{*}\left(\overline{t_{2H}}\left(\widetilde{\psi}\right), \widetilde{\psi}\right)$, which contradicts the fact that the airline's optimal profit under MX is always lower than that under 2H when both MX and 2H are feasible. Then, we can obtain that $\overline{t_{MX}} \leqslant \overline{t_{2H}}$ holds for any ψ.

2.3.2.5 The Airline's Optimal Network Structure (HS, PP, MX, and 2H)

I then compare π_{HS}^{*}, π_{PP}^{*}, π_{MX}^{*}, and π_{2H}^{*} with each other and provide the airline's optimal network structure from amongst HS, PP, MX, and 2H in Proposition 2.8.

Proposition 2.8. *Considering HS, PP, MX, and 2H, under* $\psi \in (\psi_{2H}^{soc}, +\infty)$, *the airline's optimal network structure is:*

1. *HS if* $t \in (0, t_{HS\sim PP}^{\pi}]$, *PP if* $t \in \left(t_{HS\sim PP}^{\pi}, \underline{t_{2H}}\right] \bigcup \left[\overline{t_{2H}}, +\infty\right)$ *and 2H if* $t \in \left(\underline{t_{2H}}, \overline{t_{2H}}\right)$, *when* $\psi \in \left(\psi_{2H}^{soc}, -\frac{15\lambda^2}{\xi_1 - 111}\right)$;

2. *HS if* $t \in \left(0, \underline{t_{2H}}\right]$, *2H if* $t \in \left(\underline{t_{2H}}, \overline{t_{2H}}\right)$ *and PP if* $t \in \left[\overline{t_{2H}}, +\infty\right)$, *when* $\psi \in \left[-\frac{15\lambda^2}{\xi_1 - 111}, \frac{5\lambda^2}{4}\right)$;

3. *2H if* $t \in \left(0, \overline{t_{2H}}\right)$ *and PP if* $t \in \left[\overline{t_{2H}}, +\infty\right)$, *when* $\psi \in \left[\frac{5\lambda^2}{4}, +\infty\right)$.

Proof. In Section 2.8.2.5. □

Proposition 2.8 is shown in Figure 2.14. First of all, PP could still be the airline's optimal network structure when the extra travel time disutility is relatively low, even though the region that PP is optimal for the airline shrinks due to the domination of 2H. Moreover, the result in Proposition 2.8 that 2H dominates the three other network structures when it is feasible can also help us understand the multi-hubbing and de-hubbing phenomena in the airline industry. In reality, we can observe that some airlines develop new hubs and then use 2H (multi-hubbing), while some others change from 2H to a single hub network (de-hubbing). Here comes a question: what is the motivation for airlines to use 2H? One answer is that when an airline's hub airport is congested, the airline can develop another hub to reduce the congestion of the previous hub. However, this answer might not be strong enough because besides 2H, PP and MX can also reduce hub airport congestion. The result in Proposition 2.8 shows the role of horizontal product differentiation and thus provides another explanation for the use of 2H. Other intuitions from Proposition 2.8 are the same as those from Proposition 2.5.

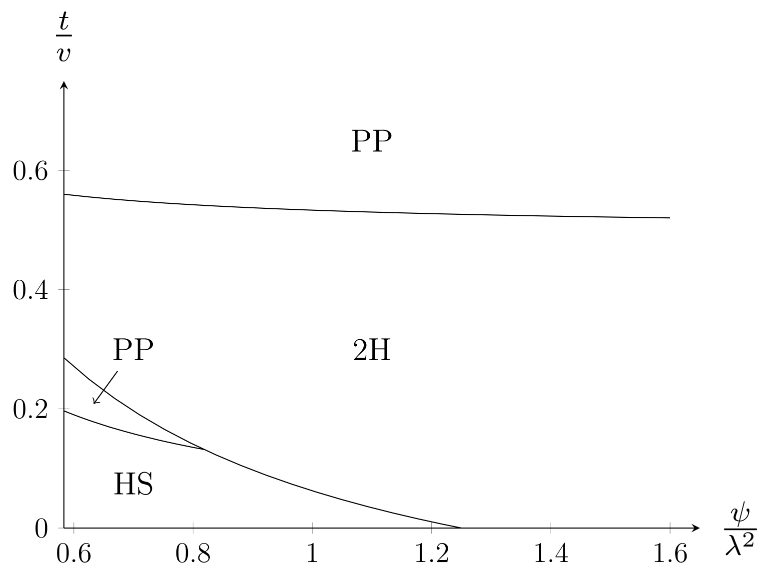

Figure 2.14: The Airline's Optimal Network Structure (HS, PP, MX, and 2H)

2.4 Welfare Analysis

In this section, I assume that a social planner maximizes social welfare. In the second-best socially optimal network structure analysis, the social planner chooses only a network structure, while the airline chooses flight frequencies and traffic. However, in the first-best socially optimal network structure analysis, the social planner decides all. The parameter space of the first-best socially optimal network structure analysis is $\psi \in \left(\frac{7\lambda^2}{6}, +\infty\right)$, under which I cannot infer whether or not the airline's optimal network structure is efficient when $\psi \in \left(\frac{7\lambda^2}{12}, \frac{7\lambda^2}{6}\right]$. Therefore, considering both the first-best and second-best socially optimal network structures will make the conclusion more robust. All critical values of this section are in Table 2.17 in Section 2.10.

2.4.1 Second-Best Socially Optimal Network Structure

I define social welfare as the sum of airline profit and passenger surplus, in which the passenger surplus is in fact a quadratic form of the traffic in different markets. That is:

$$sw_{HS}^* = \pi_{HS}^* + \left(q_{HS}^{1*}\right)^2 + \frac{1}{2}\left(q_{HS}^{3*}\right)^2, \tag{2.38}$$

$$sw_{PP}^* = \pi_{PP}^* + \frac{3}{2}\left(q_{PP}^{1*}\right)^2, \tag{2.39}$$

$$sw_{MX}^* = \pi_{MX}^* + \left(q_{MX}^{1*}\right)^2 + \frac{1}{2}\left[\left(q_{MX}^{30*}\right)^2 + q_{MX}^{30*}q_{MX}^{31*} + \left(q_{MX}^{31*}\right)^2\right], \tag{2.40}$$

$$sw_{2H}^* = \pi_{2H}^* + \frac{1}{2}\left(q_{2H}^{1*}\right)^2 + \left[\left(q_{2H}^{20*}\right)^2 + q_{2H}^{20*}q_{2H}^{21*} + \left(q_{2H}^{21*}\right)^2\right]. \tag{2.41}$$

According to (2.38) through (2.41), the network structure with higher traffic will make itself more socially desirable.

First, I compare sw_{HS}^*, sw_{PP}^*, sw_{MX}^*, and sw_{2H}^* with each other and provide the second-best socially optimal network structure from amongst HS, PP, MX, and 2H in Lemma 2.1.

Lemma 2.1. *Considering HS, PP, MX, and 2H, under $\psi \in (\psi_{2H}^{soc}, +\infty)$, the second-best socially optimal network structure is:*

1. *HS if $t \in (0, t_{HS\sim PP}^{sw}]$, PP if $t \in \left(t_{HS\sim PP}^{sw}, \underline{t_{2H}}\right] \bigcup \left[\overline{t_{2H}}, +\infty\right)$, MX if $t \in \left(\underline{t_{MX}}, t_{MX\sim 2H}^{sw}\right]$,*

and 2H if $t \in (\underline{t_{2H}}, \underline{t_{MX}}] \bigcup (t^{sw}_{MX\sim 2H}, \overline{t_{2H}})$, *when* $\psi \in \left(\psi^{soc}_{2H}, \frac{\lambda^2}{\xi_3}\right)$;

2. *HS if* $t \in (0, \underline{t_{2H}}]$, *2H if* $t \in (\underline{t_{2H}}, \underline{t_{MX}}] \bigcup (t^{sw}_{MX\sim 2H}, \overline{t_{2H}})$, *MX if* $t \in (\underline{t_{MX}}, t^{sw}_{MX\sim 2H}]$, *and PP if* $t \in [\overline{t_{2H}}, +\infty)$, *when* $\psi \in \left[\frac{\lambda^2}{\xi_3}, \frac{\lambda^2}{\xi_2}\right)$;

3. *HS if* $t \in (0, \underline{t_{2H}}]$, *2H if* $t \in (\underline{t_{2H}}, \overline{t_{2H}})$ *and PP if* $t \in [\overline{t_{2H}}, +\infty)$, *when* $\psi \in \left[\frac{\lambda^2}{\xi_2}, \frac{5\lambda^2}{4}\right)$;

4. *2H if* $t \in (0, \overline{t_{2H}})$ *and PP if* $t \in [\overline{t_{2H}}, +\infty)$, *when* $\psi \in \left[\frac{5\lambda^2}{4}, +\infty\right)$.

Proof. In Section 2.9.1.1. □

A general observation from Lemma 2.1 tells us that MX can be socially desirable when the marginal congestion delay cost ψ is low, which is different from the fact that MX can never be the airline's optimal network structure as long as 2H is under consideration.

Next, I compare Proposition 2.8 and Lemma 2.1 to find the difference between the second-best socially optimal network structure and the airline's optimal network structure (HS, PP, MX, and 2H). Proposition 2.9 summarizes this comparison.

Proposition 2.9. *Considering HS, PP, MX, and 2H, under* $\psi \in (\psi^{soc}_{2H}, +\infty)$:

1. *The second-best socially optimal network structure is HS, while the airline's optimal network structure is PP, if* $t \in \left(t^{\pi}_{HS\sim PP}, \min\left\{t^{sw}_{HS\sim PP}, \underline{t_{2H}}\right\}\right]$, *when* $\psi \in \left(\psi^{soc}_{2H}, -\frac{15\lambda^2}{\xi_1 - 111}\right)$;

2. *The second-best socially optimal network structure is MX, while the airline's optimal network structure is 2H, if* $t \in (\underline{t_{MX}}, t^{sw}_{MX\sim 2H}]$, *when* $\psi \in \left(\psi^{soc}_{2H}, \frac{\lambda^2}{\xi_2}\right)$.

Proof. In Section 2.9.1.2. □

Proposition 2.9 is shown in Figure 2.15. Obviously, the inclusion of passenger surplus does not help HS and PP to erode 2H.

For the first point in Proposition 2.9, according to the comparison of variables in Proposition 2.1, the traffic in all markets under HS is higher than PP if the extra travel time disutility t is low. Thus, given the higher traffic under HS, the social planner will favor HS more than the airline.

For the second point, among all flight frequencies of MX and 2H, f^{i*}_{MX} $(i \in \{1, 2\})$ is relatively high and carries all traffic in market AK (BK) and the connecting traffic in

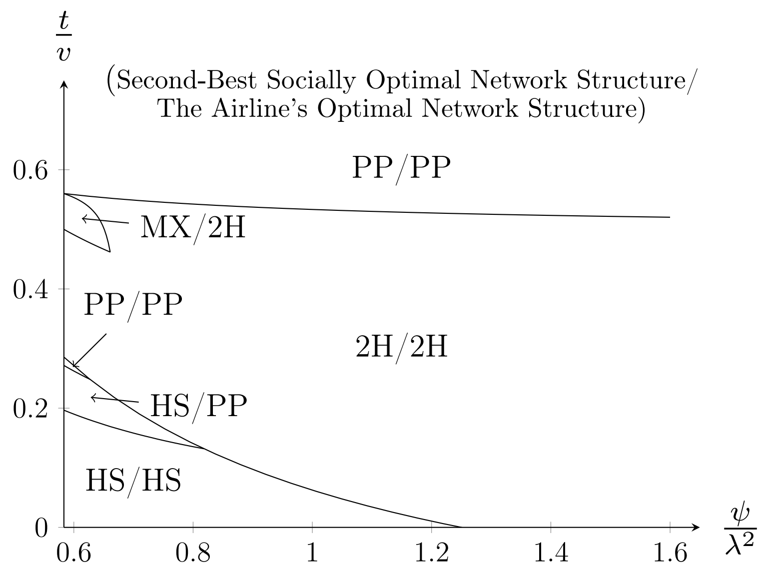

Figure 2.15: Second-Best Socially Optimal Network Structure *vs* The Airline's Optimal Network Structure (HS, PP, MX, and 2H)

market AB. When the marginal congestion delay cost ψ is low, the airline is motivated to set a high f_{MX}^{i*}, which in fact provides relevant passengers with a significant schedule delay reduction and thus is conducive to achieving the high traffic of MX. Besides, if the extra travel time disutility t is close to $\underline{t_{MX}}$, f_{MX}^{3*} will tend to zero and then MX is almost de facto HS. According to Propositions 2.1, 2.3 and 2.6, HS has relatively high traffic among all network structures. This fact thus explains why the traffic of MX is higher only if t is relatively low, given a low ψ. Then, given the higher traffic under MX in some region, the social planner will favor MX more than the airline. To have a more concrete understanding, we look at a numerical example.

Example 2.2. Setting $v = 1$, in Figure 2.16, we find that the passenger surplus of MX is higher than 2H only when both $\frac{\psi}{\lambda^2}$ and t are low.

2.4.2 First-Best Socially Optimal Network Structure

I define social welfare as the difference between passenger utility and the cost of operating flights. In fact, the social welfare function that the social planner maximizes under HS,

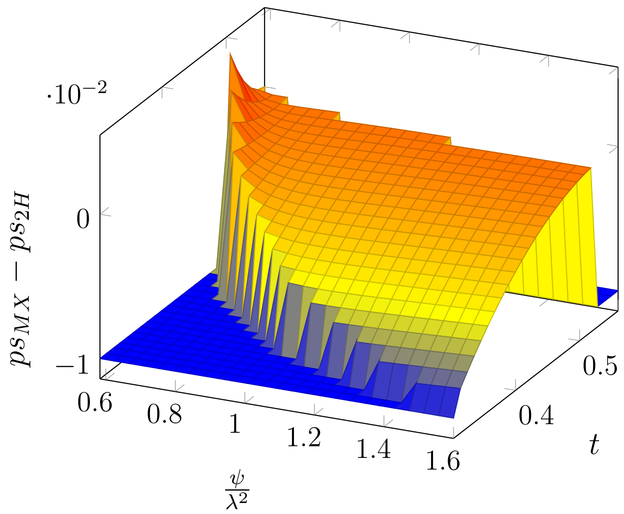

Figure 2.16: Passenger Surplus Difference Between MX and 2H ($v = 1$)

PP, MX, and 2H is equivalent to:

$$SW_{HS} = \pi_{HS} + \frac{1}{2} \sum_{i \in \{1,2,3\}} \left(q_{HS}^i\right)^2, \tag{2.42}$$

$$SW_{PP} = \pi_{PP} + \frac{1}{2} \sum_{i \in \{1,2,3\}} \left(q_{PP}^i\right)^2, \tag{2.43}$$

$$SW_{MX} = \pi_{MX} + \frac{1}{2} \left[\sum_{i \in \{1,2\}} \left(q_{MX}^i\right)^2 + \left(q_{MX}^{30}\right)^2 + q_{MX}^{30} q_{MX}^{31} + \left(q_{MX}^{31}\right)^2 \right], \tag{2.44}$$

$$SW_{2H} = \pi_{2H} + \frac{1}{2} \left\{ \left(q_{2H}^1\right)^2 + \sum_{i \in \{2,3\}} \left[\left(q_{2H}^{i0}\right)^2 + q_{2H}^{i0} q_{2H}^{i1} + \left(q_{2H}^{i1}\right)^2 \right] \right\}, \tag{2.45}$$

respectively. Compared to the airline, as in (2.42) through (2.45), the social planner takes the passenger surplus into account when making the decision.

Moreover, the parameter space is $\psi \in (\Psi_{2H}^{soc}, +\infty)$ for comparing HS, PP, MX, and 2H in the first-best socially optimal network structure analysis, in which $\Psi_{2H}^{soc} = \frac{7\lambda^2}{6}$. When we compare the first-best socially optimal and the airline's optimal network structures (HS, PP, MX, and 2H), the parameter space is also $\psi \in (\Psi_{2H}^{soc}, +\infty)$.

Except that the non-arbitrage condition does not exist any more, the process of solving (2.42) through (2.45) is similar to that of solving (2.16), (2.19), (2.20), and (2.24). These solutions are provided in Section 2.9.2. In the first-best socially optimal network structure analysis, the constraint of extra travel time disutility t of HS is $t \in \left(0, \overline{T_{HS}}\right)$, in which

$\overline{T_{HS}} = \frac{6\psi v}{6\psi - \lambda^2}$. Moreover, the constraint of t of MX is $t \in \left(\underline{T_{MX}}, \overline{T_{MX}}\right)$, in which $\underline{T_{MX}} = \frac{\left(2\psi+3\lambda^2\right)v}{10\psi-\lambda^2}$ and $\overline{T_{MX}} = \frac{4\psi v}{8\psi-\lambda^2}$, and the constraint of t of 2H is $t \in \left(\max\left\{0, \underline{T_{2H}}\right\}, \overline{T_{2H}}\right)$, in which $\underline{T_{2H}} = -\frac{\left(2\psi-5\lambda^2\right)v}{8\psi}$ and $\overline{T_{2H}} = \frac{4\psi v}{8\psi-\lambda^2}\left(= \overline{T_{MX}}\right)$. For the following lemma and propositions, I also omit proofs, which do not provide further information compared to previous ones.[14]

First, I provide the first-best socially optimal network structure from amongst HS, PP, MX, and 2H in Lemma 2.2.

Lemma 2.2. *Considering HS, PP, MX, and 2H, under $\psi \in (\Psi_{2H}^{soc}, +\infty)$, the first-best socially optimal network structure is:*

1. HS if $t \in \left(0, T_{HS\sim PP}^{SW}\right]$, PP if $t \in \left(T_{HS\sim PP}^{SW}, \underline{T_{2H}}\right] \bigcup \left[\overline{T_{2H}}, +\infty\right)$ and 2H if $t \in \left(\underline{T_{2H}}, \overline{T_{2H}}\right)$, when $\psi \in \left(\Psi_{2H}^{soc}, -\frac{30\lambda^2}{\xi_1-111}\right)$;

2. HS if $t \in \left(0, \underline{T_{2H}}\right]$, 2H if $t \in \left(\underline{T_{2H}}, \overline{T_{2H}}\right)$ and PP if $t \in \left[\overline{T_{2H}}, +\infty\right)$, when $\psi \in \left[-\frac{30\lambda^2}{\xi_1-111}, \frac{5\lambda^2}{2}\right)$;

3. 2H if $t \in \left(0, \overline{T_{2H}}\right)$ and PP if $t \in \left[\overline{T_{2H}}, +\infty\right)$, when $\psi \in \left[\frac{5\lambda^2}{2}, +\infty\right)$.

In fact, because the difference between the airline's and the social planner's objective function is only the passenger surplus, we can obtain the first-best socially optimal network structure simply by shifting the airline's optimal network structure horizontally to the right.

Next, I compare Proposition 2.8 and Lemma 2.2 to find the difference between the first-best socially optimal network structure and the airline's optimal network structure (HS, PP, MX, and 2H). Proposition 2.10 summarizes this comparison.

Proposition 2.10. *Considering HS, PP, MX, and 2H, under $\psi \in (\Psi_{2H}^{soc}, +\infty)$:*

1. The first-best socially optimal network structure is HS, while the airline's optimal network structure is 2H, if $t \in \left(\max\left\{0, \underline{t_{2H}}\right\}, \min\left\{T_{HS\sim PP}^{SW}, \underline{T_{2H}}\right\}\right]$, when $\psi \in \left(\Psi_{2H}^{soc}, \frac{5\lambda^2}{2}\right)$;

2. The first-best socially optimal network structure is PP, while the airline's optimal network structure is 2H, if $t \in \left(T_{HS\sim PP}^{SW}, \underline{T_{2H}}\right]$, when $\psi \in \left(\Psi_{2H}^{soc}, -\frac{30\lambda^2}{\xi_1-111}\right)$;

[14] All omitted proofs are available upon request.

3. The first-best socially optimal network structure is 2H, while the airline's optimal network structure is PP, if $t \in \left[\overline{t_{2H}}, \overline{T_{2H}}\right)$*, when* $\psi \in (\Psi_{2H}^{soc}, +\infty)$*.*

Proposition 2.10 is shown in Figure 2.17. According to Proposition 2.10, the airline's optimal network structure exhibits an inefficient bias towards 2H if the extra travel time disutility t is relatively low and PP if t is relatively high.

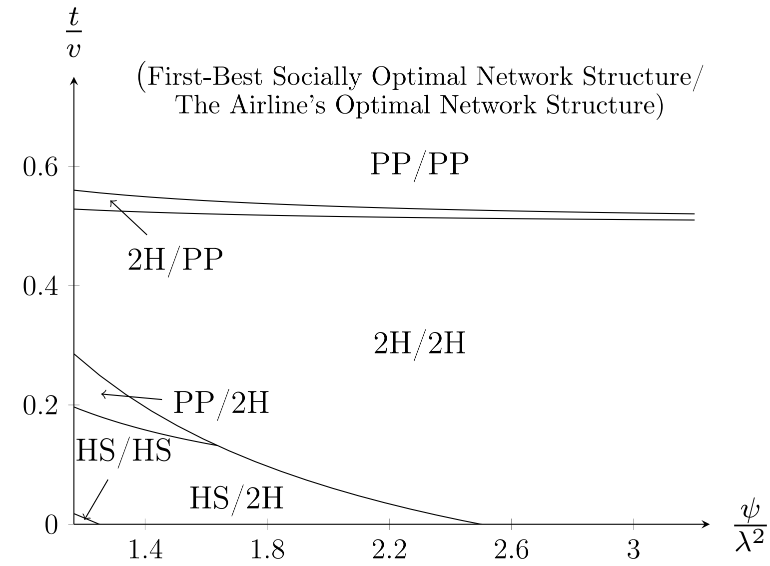

Figure 2.17: First-Best Socially Optimal Network Structure *vs* The Airline's Optimal Network Structure (HS, PP, MX, and 2H)

Finally, according to the analysis of both first-best and second-best socially optimal network structures, we can obtain Proposition 2.11.

Proposition 2.11. *From the social welfare point of view, there exist inefficient biases towards PP and 2H in the airline's optimal network structure.*

2.5 Extension: 3-Hub Network

For completeness, I extend this work to incorporate a 3-hub network (3H), or an all-hub network. This type of network structure does not exist in reality. One reason may be due to the airline's substantial fixed investments of developing all airports as hubs. However, the analysis for such network structure can help us better understand the role of horizontal product differentiation.

Under 3H (Figure 2.18), all airports are hubs. Then, under 3H, the passenger in each market can choose not only non-stop but also one-stop services, with the marginal utility of local traffic $\alpha_0^i = v + \lambda f_{3H}^i$ and connecting traffic $\alpha_1^i = v + \lambda \frac{\sum_{j\in\{1,2,3\}\setminus\{i\}} f_{3H}^j}{2} - t$, in which f_{3H}^i is the flight frequency between the city pair of market $i\,(\in \{1,2,3\})$.

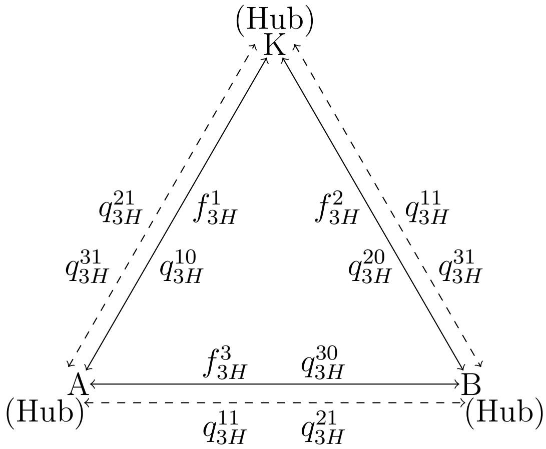

Figure 2.18: 3-Hub Network

The inverse demand functions will then be:

$$p_{3H}^{i0} = v + \lambda f_{3H}^i - q_{3H}^{i0} - \frac{1}{2} q_{3H}^{i1}, \tag{2.46}$$

$$p_{3H}^{i1} = v + \lambda \frac{\sum_{j\in\{1,2,3\}\setminus\{i\}} f_{3H}^j}{2} - t - \frac{1}{2} q_{3H}^{i0} - q_{3H}^{i1}. \tag{2.47}$$

In (2.46) and (2.47), p_{3H}^{i0}, q_{3H}^{i0}, p_{3H}^{i1}, and q_{3H}^{i1} $(i \in \{1,2,3\})$ are the fares and traffic of local and connecting traffic in each market.

Moreover, the cost structure of 3H is the same as that of PP.

Therefore, the airline's problem is:

$$\begin{aligned} \max_{f_{3H}^1, f_{3H}^2, f_{3H}^3, q_{3H}^{10}, q_{3H}^{11}, q_{3H}^{20}, q_{3H}^{21}, q_{3H}^{30}, q_{3H}^{31}} \pi_{3H} \quad = & \sum_{j\in\{0,1\}} \sum_{i\in\{1,2,3\}} p_{3H}^{ij} q_{3H}^{ij} \\ & - \sum_{i\in\{1,2,3\}} \psi f_{3H}^i \left(2 f_{3H}^i + \sum_{j\in\{1,2,3\}\setminus\{i\}} f_{3H}^j \right), \\ \text{subject to} \quad & \sum_{j\in\{1,2,3\}\setminus\{i\}} p_{3H}^{j0} \geqslant p_{3H}^{i1} \quad (i \in \{1,2,3\}), \\ \text{and} \quad & f_{3H}^i, q_{3H}^{i0}, q_{3H}^{i1} \geqslant 0 \quad (i \in \{1,2,3\}). \end{aligned} \tag{2.48}$$

The second-order condition for solving (2.48) requires $\psi \in (\psi_{3H}^{soc}, +\infty)$, in which $\psi_{3H}^{soc} = \frac{7\lambda^2}{12}$. The solution is found in Section 2.10.

In addition, under $\psi \in (\psi_{3H}^{soc}, +\infty)$, the constraint of extra travel time disutility t to make all variables positive is $t \in \left(0, \overline{t_{3H}}\right)$, in which $\overline{t_{3H}} = \frac{8\psi v}{16\psi - \lambda^2} \left(= \overline{t_{MX}} = \overline{t_{2H}}\right)$, coming from $q_{3H}^{i1*} > 0$ $(i \in \{1, 2, 3\})$. For $t < \overline{t_{3H}}$, if the extra travel time disutility is too high, the passenger in each market cannot obtain positive utility from choosing one-stop services, and thus the airline has no interest in providing the service, implying $q_{3H}^{i1*} \to 0$.

I then compare π_{HS}^*, π_{PP}^*, π_{MX}^*, π_{2H}^*, and π_{3H}^* with each other and provide the airline's optimal network structure from amongst HS, PP, MX, 2H, and 3H in Proposition 2.12. The parameter space is $\psi \in (\psi_{3H}^{soc}, +\infty)$ for this comparison.

Proposition 2.12. *Considering HS, PP, MX, 2H, and 3H, under $\psi \in (\psi_{3H}^{soc}, +\infty)$, the airline's optimal network structure is 3H if $t \in \left(0, \overline{t_{3H}}\right)$ and PP if $t \in \left[\overline{t_{3H}}, +\infty\right)$.*

Proposition 2.12 is shown in Figure 2.19 and tells us that, without considering the airline's fixed investments of developing a hub airport, 3H is optimal for the airline as long as the extra travel time disutility t is not too high. In fact, this result comes from the presence of horizontal product differentiation in each market under 3H. Therefore, Proposition 2.12 further shows the role of horizontal product differentiation in improving passenger utility and airline profit.

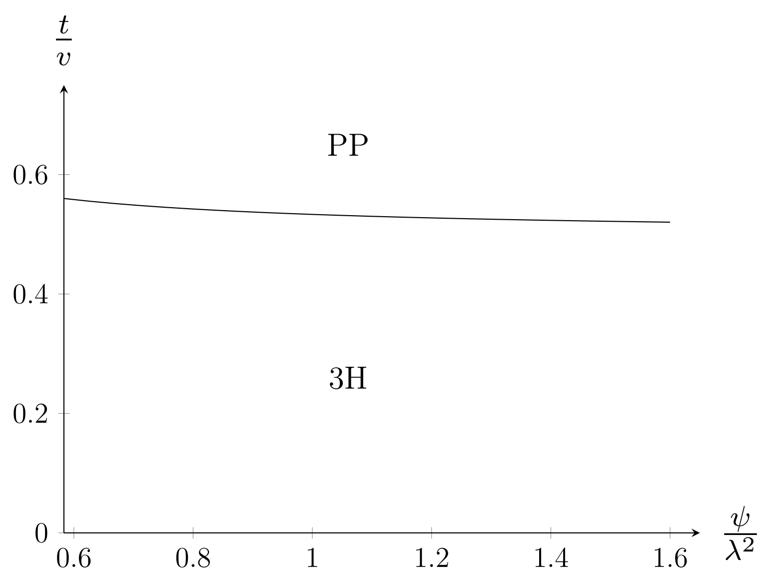

Figure 2.19: The Airline's Optimal Network Structure (HS, PP, MX, 2H, and 3H)

2.6 Conclusion

This chapter investigates how congestion delays and horizontal product differentiation shape airline network structures. The novelty of this work includes the complete coverage of network structures, an increased understanding of the role of congestion delays in airline network choice, and a new perspective of studying the 2-hub network (2H), that is, the horizontal product differentiation.

The key results regarding the profitability of different network structures show the impact of omitting negative network externality and the important role of horizontal product differentiation in improving passenger utility and airline profit. Moreover, comparative statics show that the division of local and connecting traffic in one market can provide airlines with some level of flexibility in the decision-making process.

In addition, welfare analysis also delivers a policy implication. When an airline proposes merging with another in order to form a 2H, the antitrust authority should scrutinize the case for the following reasons. First, when the network structure changes to 2H, fares in many markets may increase but the social welfare may not. Second, with the sharp decrease of oil price and the improvement of air traffic management systems, both the marginal congestion delay cost and extra travel time disutility may decrease, and thus the claimed efficiency of 2H could be in doubt.

Next, some caveats should be noted. One important assumption in this chapter is the symmetry of demand among different markets. If the demand from a spoke airport to hub airport is too limited, the airline's benefit of horizontal product differentiation from multiple hubs may be too weak compared to its fixed investments of developing a hub airport. In this case, the network structures with fewer hubs will be favored more by the airline. Moreover, according to the results of this chapter, one may try to hypothesize that for some region, the more hubs there are, the higher the airline's profit will be. However, this hypothesis holds only when the airline's fixed investments of developing a hub airport are low enough. Furthermore, even though the focus of this chapter is congestion delays and horizontal product differentiation, this does not mean that they are the most important factors in airline network choice. Essentially, this chapter abstracts from other factors, for instance, the economies of traffic density and demand uncertainty, and identifies how congestion delays and horizontal product differentiation themselves affect airline network choice.

Finally, a direction of future research is to introduce airline competition, in which airlines can compete in one or more markets. Moreover, extending the analysis to arbitrary number of cities in a network is also valuable.

2.7 Appendix A: Examples of Network Structures and Schedules

First, an example of the hub-and-spoke network (HS) is the Alitalia network shown in Figure 2.20. In this network, passengers traveling between Naples (Venice) and Rome can take Alitalia flights to arrive directly, while passengers traveling between Naples and Venice have to connect at Rome.

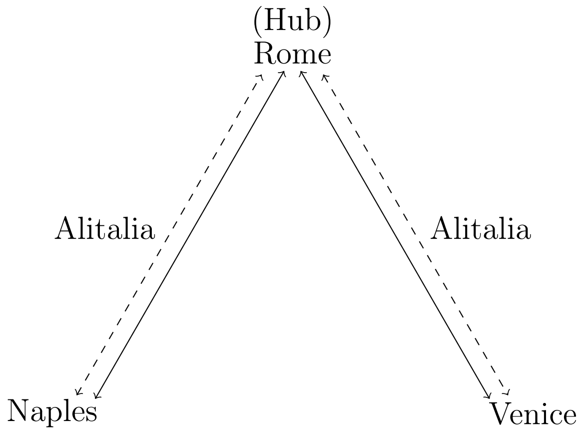

Figure 2.20: An Example of Hub-and-Spoke Network (Alitalia)

Second, an example of the point-to-point network (PP) is the Ryanair network shown in Figure 2.21. In this network, passengers traveling between any two airports from amongst Edinburgh, Brussels-CRL and Bordeaux can take Ryanair flights to arrive directly.

Third, an example of the mixed network (MX) is the Air France network shown in Figure 2.22. In this network, passengers traveling between Bordeaux (Nice) and Paris-

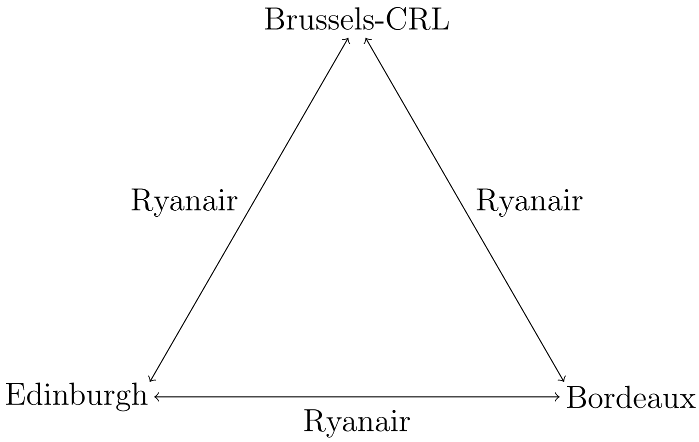

Figure 2.21: An Example of Point-to-Point Network (Ryanair)

Orly can take Air France flights to arrive directly. However, for passengers traveling between Bordeaux and Nice, they can take either HOP! flights to arrive directly or Air France flights connecting at Paris-Orly.

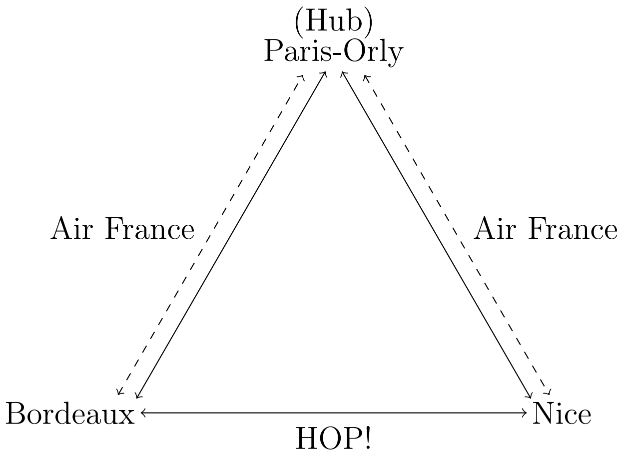

Figure 2.22: An Example of Mixed Network (Air France)

Finally, as shown in Tables 2.8 through 2.11, flight schedules between Toulouse and Frankfurt (Munich) indicate different flight departure time slots.

No.	Departure	Arrival	Connection Airport	Operated by	
				First Part	Second Part
1	07:05	09:00		Lufthansa	
2	12:45	14:35		Lufthansa	
3	18:55	20:45		Lufthansa	
4	06:05	09:40	Munich	Lufthansa Cityline	Lufthansa
5	06:05	10:10	Munich	Lufthansa Cityline	Lufthansa
6	08:35	12:10	Munich	Lufthansa Cityline	Lufthansa
7	08:35	13:05	Munich	Lufthansa Cityline	Lufthansa
8	13:00	17:05	Munich	Lufthansa Cityline	Lufthansa

Table 2.8: From Toulouse to Frankfurt on November 1, 2016
Source: Lufthansa's reservation website

No.	Departure	Arrival	Connection Airport	Operated by	
				First Part	Second Part
1	10:10	11:50		Lufthansa	
2	16:35	18:15		Lufthansa	
3	21:00	22:40		Lufthansa	
4	07:15	12:25	Munich	Lufthansa	Lufthansa Cityline
5	08:15	12:25	Munich	Lufthansa	Lufthansa Cityline
6	13:15	17:35	Munich	Lufthansa	Air Dolomiti
7	16:15	21:00	Munich	Lufthansa	Lufthansa Cityline
8	17:15	21:00	Munich	Lufthansa	Lufthansa Cityline

Table 2.9: From Frankfurt to Toulouse on November 10, 2016
Source: Lufthansa's reservation website

2.8 Appendix B: Derivations and Proofs of Market Outcome

2.8.1 Derivations and Proofs of Flight Frequencies and Traffic

2.8.1.1 Hub-and-Spoke Network

I first do not put the non-arbitrage condition as a constraint of the maximization problem. Then, the Kuhn-Tucker conditions are:

$$\frac{\partial \pi_{HS}}{\partial f^i_{HS}} = \lambda \left(q^i_{HS} + \frac{q^3_{HS}}{2} \right) - 2\psi \left(2f^i_{HS} + f^j_{HS} \right) \leqslant 0, \quad f^i_{HS} \geqslant 0, \quad f^i_{HS} \frac{\partial \pi_{HS}}{\partial f^i_{HS}} = 0; \quad (2.49)$$

$$\frac{\partial \pi_{HS}}{\partial q^i_{HS}} = v + \lambda f^i_{HS} - 2q^i_{HS} \leqslant 0, \quad q^i_{HS} \geqslant 0, \quad q^i_{HS} \frac{\partial \pi_{HS}}{\partial q^i_{HS}} = 0; \quad (2.50)$$

No.	Departure	Arrival	Connection Airport	Operated by	
				First Part	Second Part
1	06:05	07:50		Lufthansa Cityline	
2	08:35	10:20		Lufthansa Cityline	
3	13:00	14:45		Lufthansa Cityline	
4	18:50	20:35		Air Dolomiti	
5	07:05	11:15	Frankfurt	Lufthansa	Lufthansa
6	07:05	13:15	Frankfurt	Lufthansa	Lufthansa
7	12:45	17:15	Frankfurt	Lufthansa	Lufthansa
8	12:45	18:15	Frankfurt	Lufthansa	Lufthansa

Table 2.10: From Toulouse to Munich on November 1, 2016
Source: Lufthansa's reservation website

No.	Departure	Arrival	Connection Airport	Operated by	
				First Part	Second Part
1	06:10	08:00		Lufthansa Cityline	
2	10:35	12:25		Lufthansa Cityline	
3	15:45	17:35		Air Dolomiti	
4	19:10	21:00		Lufthansa Cityline	
5	08:00	11:50	Frankfurt	Lufthansa	Lufthansa
6	14:00	18:15	Frankfurt	Lufthansa	Lufthansa
7	18:00	22:40	Frankfurt	Lufthansa	Lufthansa
8	19:00	22:40	Frankfurt	Lufthansa	Lufthansa

Table 2.11: From Munich to Toulouse on November 10, 2016
Source: Lufthansa's reservation website

$$\frac{\partial \pi_{HS}}{\partial q_{HS}^3} = v + \lambda \frac{\sum_{i\in\{1,2\}} f_{HS}^i}{2} - t - 2q_{HS}^3 \leqslant 0, \qquad q_{HS}^3 \geqslant 0, \quad q_{HS}^3 \frac{\partial \pi_{HS}}{\partial q_{HS}^3} = 0, \quad (2.51)$$

in which $i \in \{1,2\}$ and $j \in \{1,2\} \setminus \{i\}$.

Next, according to (2.50), we have $q_{HS}^i > 0$ because otherwise $v > 0$ and $f_{HS}^i \geqslant 0$ imply $\frac{\partial \pi_{HS}}{\partial q_{HS}^i} > 0$, contradicting $\frac{\partial \pi_{HS}}{\partial q_{HS}^i} \leqslant 0$. Then, we have $f_{HS}^i > 0$ because otherwise $q_{HS}^i > 0$ cannot hold. Finally, we have $q_{HS}^3 > 0$ because otherwise PP will dominate HS. Thus, only the interior solution matters, implying:

$$f_{HS}^{1*} = f_{HS}^{2*} = \lambda \left(3v - t\right) \left(3\phi_{HS}\right)^{-1},$$

$$q_{HS}^{1*} = q_{HS}^{2*} = \left(24\psi v - \lambda^2 t\right) \left(6\phi_{HS}\right)^{-1},$$

$$q_{HS}^{2*} = \left[12\psi \left(v - t\right) + \lambda^2 t\right] \left(3\phi_{HS}\right)^{-1},$$

in which $\phi_{HS} = 8\psi - \lambda^2$.

Then, the exclusion of all corner solutions implies that the domain of π_{HS} becomes $\mathbb{R}^5_{++}$, which is a convex open set. Moreover, when $\psi \in (\psi^{soc}_{HS}, +\infty)$, the Hessian $D^2\pi_{HS}$ is negative definite for all $\Omega_{HS} \in \mathbb{R}^5_{++}$, implying that π_{HS} is a strictly concave function on $\mathbb{R}^5_{++}$. Consequently, the optimal solution is a unique global maximizer of π_{HS} on $\mathbb{R}^5_{++}$.

Proof of Non-Arbitrage Condition

Proof. The non-arbitrage condition is satisfied because of:

$$\begin{aligned} p^{1*}_{HS} + p^{2*}_{HS} - p^{3*}_{HS} &= \left[\left(6\psi - \lambda^2\right) t + 6\psi v\right] \left(3\phi_{HS}\right)^{-1} \\ \underbrace{>}_{\text{from } \psi \in \left(\psi^{soc}_{HS}, +\infty\right)} & \quad \lambda^2 \left(3v + t\right) \left(6\phi_{HS}\right)^{-1} > 0. \end{aligned}$$

□

2.8.1.2 Point-to-Point Network

The Kuhn-Tucker conditions are:

$$\frac{\partial \pi_{PP}}{\partial f^i_{PP}} = \lambda q^i_{PP} - 2\psi \left(2f^i_{PP} + \sum_{j\in\{1,2,3\}\setminus\{i\}} f^j_{PP}\right) \leqslant 0, \quad f^i_{PP} \geqslant 0, \quad f^i_{PP}\frac{\partial \pi_{PP}}{\partial f^i_{PP}} = 0; \quad (2.52)$$

$$\frac{\partial \pi_{PP}}{\partial q^i_{PP}} = v + \lambda f^i_{PP} - 2q^i_{PP} \leqslant 0, \quad q^i_{PP} \geqslant 0, \quad q^i_{PP}\frac{\partial \pi_{PP}}{\partial q^i_{PP}} = 0, \quad (2.53)$$

in which $i \in \{1, 2, 3\}$.

Next, according to (2.53), we have $q^i_{PP} > 0$ because otherwise $v > 0$ and $f^i_{PP} \geqslant 0$ imply $\frac{\partial \pi_{PP}}{\partial q^i_{PP}} > 0$, contradicting $\frac{\partial \pi_{PP}}{\partial q^i_{PP}} \leqslant 0$. Then, we have $f^i_{PP} > 0$ because otherwise $q^i_{PP} > 0$ cannot hold. Thus, only the interior solution matters, implying:

$$f^{i*}_{PP} = \lambda v \phi^{-1}_{PP},$$

$$q^{i*}_{PP} = 8\psi v \phi^{-1}_{PP},$$

in which $i \in \{1, 2, 3\}$ and $\phi_{PP} = 16\psi - \lambda^2$.

Then, following the reasoning in Section 2.8.1.1, the optimal solution under PP is a unique global maximizer of π_{PP} on $\mathbb{R}^6_{++}$.

2.8.1.3 Mixed Network

The Kuhn-Tucker conditions are:

$$\begin{aligned}
\frac{\partial \pi_{MX}}{\partial f^i_{MX}} &= \lambda\left(q^i_{MX} + \frac{q^{31}_{MX}}{2}\right) \\
&\quad - 2\psi\left(2f^i_{MX} + \sum_{j\in\{1,2,3\}\setminus\{i\}} f^j_{MX}\right) \leqslant 0, \quad f^i_{MX} \geqslant 0, \quad f^i_{MX}\frac{\partial \pi_{MX}}{\partial f^i_{MX}} = 0; \quad (2.54)\\
\frac{\partial \pi_{MX}}{\partial f^3_{MX}} &= \lambda q^{30}_{MX} - 2\psi\left(2f^3_{MX} + \sum_{i\in\{1,2\}} f^i_{MX}\right) \leqslant 0, \quad f^3_{MX} \geqslant 0, \quad f^3_{MX}\frac{\partial \pi_{MX}}{\partial f^3_{MX}} = 0; \quad (2.55)\\
\frac{\partial \pi_{MX}}{\partial q^i_{MX}} &= v + \lambda f^i_{MX} - 2q^i_{MX} \leqslant 0, \quad q^i_{MX} \geqslant 0, \quad q^i_{MX}\frac{\partial \pi_{MX}}{\partial q^i_{MX}} = 0; \quad (2.56)\\
\frac{\partial \pi_{MX}}{\partial q^{30}_{MX}} &= v + \lambda f^3_{MX} - 2q^{30}_{MX} - q^{31}_{MX} \leqslant 0, \quad q^{30}_{MX} \geqslant 0, \quad q^{30}_{MX}\frac{\partial \pi_{MX}}{\partial q^{30}_{MX}} = 0; \quad (2.57)\\
\frac{\partial \pi_{MX}}{\partial q^{31}_{MX}} &= v + \lambda\frac{\sum_{i\in\{1,2\}} f^i_{MX}}{2} - t \\
&\quad - q^{30}_{MX} - 2q^{31}_{MX} \leqslant 0, \quad q^{31}_{MX} \geqslant 0, \quad q^{31}_{MX}\frac{\partial \pi_{MX}}{\partial q^{31}_{MX}} = 0, \quad (2.58)
\end{aligned}$$

in which $i \in \{1,2\}$.

Next, (I) according to (2.56), we have $q^i_{MX} > 0$ because otherwise $v > 0$ and $f^i_{MX} \geqslant 0$ imply $\frac{\partial \pi_{MX}}{\partial q^i_{MX}} > 0$, contradicting $\frac{\partial \pi_{MX}}{\partial q^i_{MX}} \leqslant 0$. (II) We have $f^i_{MX} > 0$ because otherwise $q^i_{MX} > 0$ cannot hold. (III) Analogously, $f^3_{MX} = 0$ and $q^{30}_{MX} > 0$ also cannot hold simultaneously. (IV) $f^3_{MX} > 0$ and $q^{30}_{MX} = 0$ cannot hold simultaneously because it is not optimal for the airline to operate empty flights. (V) If all variables except f^3_{MX} and q^{30}_{MX} (resp. q^{31}_{MX}) are positive, MX is de facto HS (resp. PP). (VI) If all variables except f^3_{MX}, q^{30}_{MX} and q^{31}_{MX} are positive, PP will dominate MX. Thus, only the interior solution matters, implying:

$$\begin{aligned}
f^{1*}_{MX} &= f^{2*}_{MX} = \lambda\left[12\psi\left(v-t\right) - \lambda^2\left(3v-t\right)\right]\phi^{-1}_{MX},\\
f^{3*}_{MX} &= \lambda\left[4\psi\left(-v+5t\right) - \lambda^2\left(3v+t\right)\right]\phi^{-1}_{MX},
\end{aligned}$$

$$q_{MX}^{1*} = q_{MX}^{2*} = \left[96\psi^2 v - 4\lambda^2\psi\left(10v + 3t\right) + \lambda^4 t\right]\left(2\phi_{MX}\right)^{-1},$$

$$q_{MX}^{30*} = 8\psi\left[4\psi\left(v + t\right) - 3\lambda^2 v\right]\phi_{MX}^{-1},$$

$$q_{MX}^{31*} = \left(4\psi - \lambda^2\right)\left[8\psi\left(v - 2t\right) + \lambda^2 t\right]\phi_{MX}^{-1},$$

in which $\phi_{MX} = 96\psi^2 - 52\lambda^2\psi + 3\lambda^4$.

Then, following the reasoning in Section 2.8.1.1, the optimal solution under MX is a unique global maximizer of π_{MX} on $\mathbb{R}_{++}^7$.

Proof of Non-Arbitrage Condition

Proof. The non-arbitrage condition is satisfied because of:

$$p_{MX}^{1*} + p_{MX}^{2*} - p_{MX}^{31*} \geqslant 0 \Leftrightarrow t \leqslant \frac{1}{2}\left(11 - \sqrt{97}\right)v,$$

in which the latter always holds, that is:

$$t < \overline{t_{MX}} \underbrace{<}_{\text{from } \psi\in\left(\psi_{MX}^{soc},+\infty\right)} \frac{1}{2}\left(11 - \sqrt{97}\right)v.$$

□

2.8.1.4 2-Hub Network

The Kuhn-Tucker conditions are:

$$\frac{\partial\pi_{2H}}{\partial f_{2H}^1} = \lambda\left(q_{2H}^1 + \frac{\sum_{i\in\{2,3\}} q_{2H}^{i1}}{2}\right) - 2\psi\left(2f_{2H}^1 + \sum_{i\in\{2,3\}} f_{2H}^i\right) \leqslant 0, \qquad f_{2H}^1 \geqslant 0, \quad f_{2H}^1\frac{\partial\pi_{2H}}{\partial f_{2H}^1} = 0; \qquad (2.59)$$

$$\frac{\partial\pi_{2H}}{\partial f_{2H}^i} = \lambda\left(q_{2H}^{i0} + \frac{q_{2H}^{j1}}{2}\right) - 2\psi\left(2f_{2H}^i + \sum_{j\in\{1,2,3\}\setminus\{i\}} f_{2H}^j\right) \leqslant 0, \qquad f_{2H}^i \geqslant 0, \quad f_{2H}^i\frac{\partial\pi_{2H}}{\partial f_{2H}^i} = 0; \qquad (2.60)$$

$$\frac{\partial \pi_{2H}}{\partial q_{2H}^{1}} = v + \lambda f_{2H}^{1} - 2q_{2H}^{1} \leqslant 0, \qquad q_{2H}^{1} \geqslant 0, \quad q_{2H}^{1} \frac{\partial \pi_{2H}}{\partial q_{2H}^{1}} = 0; \tag{2.61}$$

$$\frac{\partial \pi_{2H}}{\partial q_{2H}^{i0}} = v + \lambda f_{2H}^{i} - 2q_{2H}^{i0} - q_{2H}^{i1} \leqslant 0, \qquad q_{2H}^{i0} \geqslant 0, \quad q_{2H}^{i0} \frac{\partial \pi_{2H}}{\partial q_{2H}^{i0}} = 0; \tag{2.62}$$

$$\begin{aligned}\frac{\partial \pi_{2H}}{\partial q_{2H}^{i1}} &= v + \lambda \frac{\sum_{j \in \{1,2,3\} \setminus \{i\}} f_{2H}^{j}}{2} - t \\ &\quad - q_{2H}^{i0} - 2q_{2H}^{i1} \leqslant 0, \qquad q_{2H}^{i1} \geqslant 0, \quad q_{2H}^{i1} \frac{\partial \pi_{2H}}{\partial q_{2H}^{i1}} = 0,\end{aligned} \tag{2.63}$$

in which $i \in \{2,3\}$ and $j \in \{2,3\} \setminus \{i\}$.

Next, (I) according to (2.61), we have $q_{2H}^{1} > 0$ because otherwise $v > 0$ and $f_{2H}^{1} \geqslant 0$ imply $\frac{\partial \pi_{2H}}{\partial q_{2H}^{1}} > 0$, contradicting $\frac{\partial \pi_{2H}}{\partial q_{2H}^{1}} \leqslant 0$. (II) We have $f_{2H}^{1} > 0$ because otherwise $q_{2H}^{1} > 0$ cannot hold. (III) Analogously, $f_{2H}^{i} = 0$ and $q_{2H}^{i0} > 0$ ($q_{2H}^{j1} > 0$) cannot hold simultaneously. (IV) According to (2.62), $q_{2H}^{i0} = 0$ and $q_{2H}^{i1} = 0$ cannot hold simultaneously because otherwise $v > 0$ and $f_{2H}^{i} \geqslant 0$ imply $\frac{\partial \pi_{2H}}{\partial q_{2H}^{i0}} > 0$, contradicting $\frac{\partial \pi_{2H}}{\partial q_{2H}^{i0}} \leqslant 0$. (V) $f_{2H}^{i} > 0$, $q_{2H}^{i0} = 0$ and $q_{2H}^{j1} = 0$ cannot hold simultaneously because it is not optimal for the airline to operate empty flights. (VI) If there are $f_{2H}^{i} > 0$, $q_{2H}^{i0} > 0$, $q_{2H}^{i1} = 0$, $f_{2H}^{j} = 0$, $q_{2H}^{j0} = 0$, and $q_{2H}^{j1} > 0$, 2H is de facto HS. (VII) If all variables except (resp. one of) q_{2H}^{21} and q_{2H}^{31} are positive, 2H is de facto PP (resp. MX). (VIII) Cases that all variables except q_{2H}^{20} or (and) q_{2H}^{30} are positive imply that, there are non-stop services between the origin and destination airports, but the airline forbids passengers who wish to travel between them to choose these non-stop services, which are empirically irrelevant. (IX) If there is $f_{2H}^{1} = f_{2H}^{2} = 0$, PP will dominate 2H. Thus, only the interior solution matters, implying:

$$\begin{aligned}
f_{2H}^{1*} &= \lambda \left[12\psi\,(3v - 4t) - \lambda^{2}\,(5v - 4t)\right] \phi_{2H}^{-1}, \\
f_{2H}^{2*} = f_{2H}^{3*} &= \lambda \left[4\psi\,(v + 4t) - 5\lambda^{2} v\right] \phi_{2H}^{-1}, \\
q_{2H}^{1*} &= 2\left[48\psi^{2} v - 12\lambda^{2}\psi\,(v + t) + \lambda^{4} t\right] \phi_{2H}^{-1}, \\
q_{2H}^{20*} = q_{2H}^{30*} &= \left[64\psi^{2}\,(v + t) - 4\lambda^{2}\psi\,(8v + 3t) + \lambda^{4} t\right] \phi_{2H}^{-1}, \\
q_{2H}^{21*} = q_{2H}^{31*} &= 2\left(4\psi - \lambda^{2}\right)\left[8\psi\,(v - 2t) + \lambda^{2} t\right] \phi_{2H}^{-1},
\end{aligned}$$

in which $\phi_{2H} = 192\psi^{2} - 84\lambda^{2}\psi + 5\lambda^{4}$.

Then, following the reasoning in Section 2.8.1.1, the optimal solution under 2H is a unique global maximizer of π_{2H} on $\mathbb{R}^8_{++}$.

Proof of Non-Arbitrage Condition

Proof. The non-arbitrage conditions ($i \in \{2,3\}, j \in \{2,3\} \setminus \{i\}$) are satisfied because of:

$$p^{1*}_{2H} + p^{i0*}_{2H} - p^{j1*}_{2H} \geqslant 0 \Leftrightarrow \left(-43 + 4\sqrt{97}\right) t + 6\left(17 - 2\sqrt{97}\right) v \leqslant 0,$$

in which the latter always holds. □

2.8.2 Proofs of Network Structure

2.8.2.1 Proof of Proposition 2.2

Proof. Under $\psi \in (\psi^{soc}_{HS}, +\infty)$, the sign of $\pi^*_{HS} - \pi^*_{PP}$ depends on a quadratic function of t. As the coefficient of t^2 is positive, the parabola opens upwards. Moreover, the discriminant:

$$\Delta = 144\psi v^2 \left(\phi_{HS}\right)^2 \phi_{PP},$$

is positive.

Then, solving $\pi^*_{HS} - \pi^*_{PP} = 0$ gives two roots, in which the smaller root is less than $\overline{t_{HS}}$, while the larger root is greater than $\overline{t_{HS}}$. According to the constraint of t of HS $t \in \left(0, \overline{t_{HS}}\right)$, only the smaller root is relevant, denoted by:

$$t^{\pi}_{HS\sim PP} = \frac{6v}{12\psi - \lambda^2}\left[2\psi - \frac{\left(8\psi - \lambda^2\right)\sqrt{\left(16\psi - \lambda^2\right)\psi}}{16\psi - \lambda^2}\right].$$

Therefore, for the airline, π^*_{HS} is higher and then HS is more profitable if $t \in (0, t^{\pi}_{HS\sim PP}]$; π^*_{PP} is higher and then PP is more profitable if $t \in \left(t^{\pi}_{HS\sim PP}, \overline{t_{HS}}\right)$. Moreover, π^*_{HS} does not exist and then PP is more profitable if $t \in \left[\overline{t_{HS}}, +\infty\right)$. □

2.8.2.2 Proof of Proposition 2.4

Proof. Under $\psi \in (\psi^{soc}_{MX}, +\infty)$, the sign of $\pi^*_{HS} - \pi^*_{MX}$ depends on a quadratic function of t. As the coefficient of t^2 is negative, the parabola opens downwards. Moreover, the discriminant:

$$\Delta = -192\lambda^2\psi^2 v^2 \phi_{HS}\phi_{MX},$$

is negative. To summarize, the quadratic function of t is always negative. We can also verify $\overline{t_{MX}} < \overline{t_{HS}}$. Therefore, for the airline, π^*_{MX} is higher and then MX is more profitable if $t \in \left(\underline{t_{MX}}, \overline{t_{MX}}\right)$. Moreover, π^*_{MX} does not exist and then HS is more profitable if $t \in \left(0, \underline{t_{MX}}\right] \bigcup \left[\overline{t_{MX}}, \overline{t_{HS}}\right)$. This result is shown in Figure 2.23.

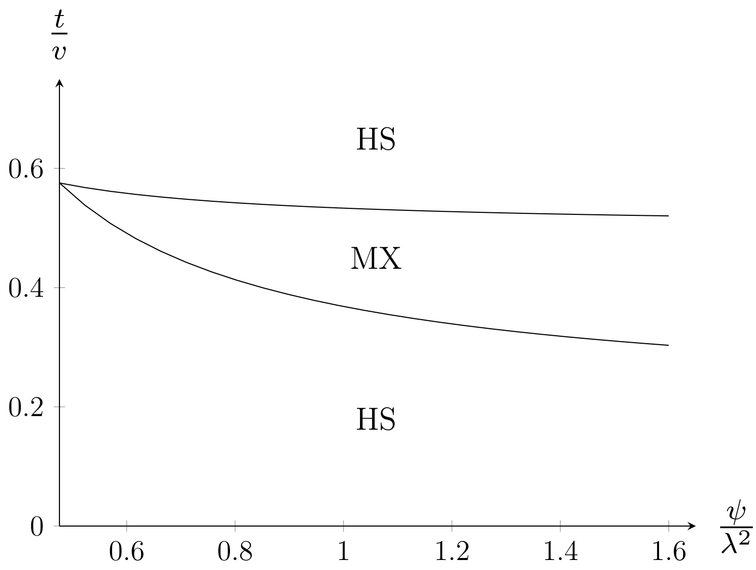

Figure 2.23: Comparison of Profits Between HS and MX

In addition, under $\psi \in (\psi^{soc}_{MX}, +\infty)$, $\pi^*_{PP} - \pi^*_{MX}$ is negative. Therefore, for the airline, π^*_{MX} is higher and then MX is more profitable if $t \in \left(\underline{t_{MX}}, \overline{t_{MX}}\right)$. Moreover, π^*_{MX} does not exist and then PP is more profitable if $t \in \left(0, \underline{t_{MX}}\right] \bigcup \left[\overline{t_{MX}}, +\infty\right)$. This result is shown in Figure 2.24.

□

2.8.2.3 Proof of Proposition 2.5

Proof. The comparison between $t^{\pi}_{HS\sim PP}$ and $\underline{t_{MX}}$ is the key to the proof.

Let $\lambda^2 = x\psi$, in which $x \in \left(0, \frac{\lambda^2}{\psi^{soc}_{MX}}\right)$. Then, we have:

$$t^{\pi}_{HS\sim PP} - \underline{t_{MX}} = \frac{\sigma_1(x)\, v}{(-x+20)(-x+16)(-x+12)},$$

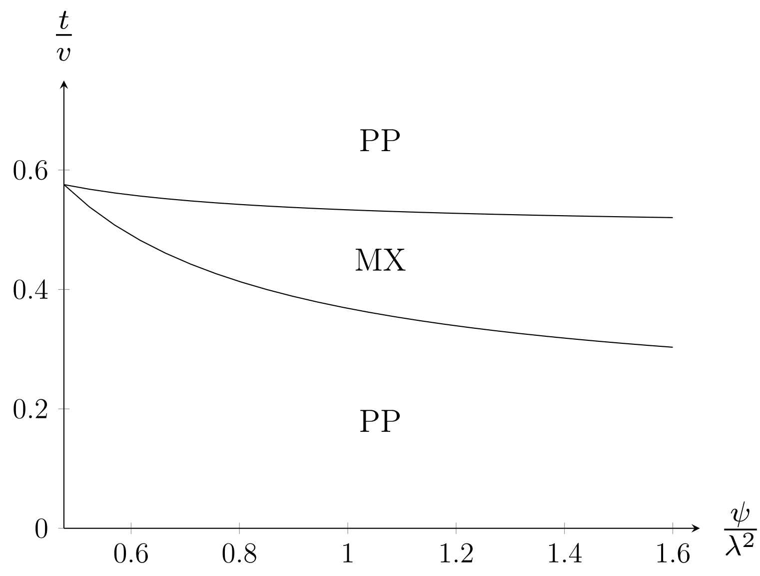

Figure 2.24: Comparison of Profits Between PP and MX

in which $(-x+20)(-x+16)(-x+12) > 0$ and:

$$\sigma_1(x) = -3x^3 + 92x^2 - 896x + 3072 - 6(-x+20)(-x+8)\sqrt{-x+16}.$$

By considering the gradients and extrema of $\frac{\mathrm{d}^n\sigma_1(x)}{\mathrm{d}x^n}$, we can find the gradients of $\frac{\mathrm{d}^{n-1}\sigma_1(x)}{\mathrm{d}x^{n-1}}$. Then, starting from $\frac{\mathrm{d}^4\sigma_1(x)}{\mathrm{d}x^4}$ and inferring step by step, there will be $\frac{\mathrm{d}^2\sigma_1(x)}{\mathrm{d}x^2} > 0$ and:

$$\sup_{x\in\left(0,\frac{\lambda^2}{\psi_{MX}^{soc}}\right)} \frac{\mathrm{d}\sigma_1(x)}{\mathrm{d}x} > 0, \qquad \inf_{x\in\left(0,\frac{\lambda^2}{\psi_{MX}^{soc}}\right)} \frac{\mathrm{d}\sigma_1(x)}{\mathrm{d}x} < 0. \tag{2.64}$$

According to (2.64), $\sigma_1(x)$ first decreases and then increases in x. Moreover, we can also find:

$$\lim_{x\to 0}\sigma_1(x) < 0, \qquad \lim_{x\to\frac{\lambda^2}{\psi_{MX}^{soc}}} \sigma_1(x) < 0. \tag{2.65}$$

According to (2.65), we can obtain $\sigma_1(x) < 0$. Therefore, we have $t^{\pi}_{HS\sim PP} < \underline{t_{MX}}$.

Then, the order of relevant critical values is $0 < t^{\pi}_{HS\sim PP} < \underline{t_{MX}} < \overline{t_{MX}} < \overline{t_{HS}}$. Combining the results in Propositions 2.2 and 2.4, we can obtain Proposition 2.5. □

2.8.2.4 Proof of Proposition 2.7

Proof. Under $\psi \in (\psi_{2H}^{soc}, +\infty)$, the sign of $\pi_{HS}^* - \pi_{2H}^*$ depends on a quadratic function of t. Let $\lambda^2 = x\psi$, in which $x \in \left(0, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)$. Then, the coefficient of t^2 becomes $\sigma_2(x)\psi^3$, in which $\sigma_2(x)$ is a polynomial of degree 3 and is negative. Thus, the parabola opens downwards. Meanwhile, the discriminant:

$$\Delta = -48\psi v^2 \left(288\psi^2 - 140\lambda^2\psi + 21\lambda^4\right)\phi_{HS}\phi_{2H},$$

is negative. To summarize, the quadratic function of t is always negative. Moreover, because of $\overline{t_{2H}} = \overline{t_{MX}}$ and $\overline{t_{MX}} < \overline{t_{HS}}$, we have $\overline{t_{2H}} < \overline{t_{HS}}$. Therefore, for the airline, π_{2H}^* is higher and then 2H is more profitable if $t \in \left(\max\left\{0, \underline{t_{2H}}\right\}, \overline{t_{2H}}\right)$. Moreover, π_{2H}^* does not exist and then HS is more profitable if $t \in \left(0, \underline{t_{2H}}\right] \bigcup \left[\overline{t_{2H}}, \overline{t_{HS}}\right)$ when $\psi \in \left(\psi_{2H}^{soc}, \frac{5\lambda^2}{4}\right)$ and $t \in \left[\overline{t_{2H}}, \overline{t_{HS}}\right)$ when $\psi \in \left[\frac{5\lambda^2}{4}, +\infty\right)$. This result is shown in Figure 2.25.

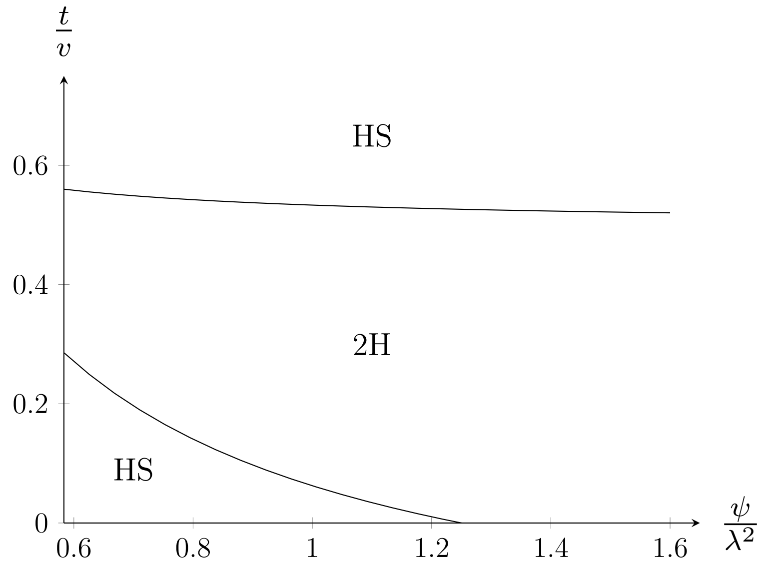

Figure 2.25: Comparison of Profits Between HS and 2H

In addition, under $\psi \in (\psi_{2H}^{soc}, +\infty)$, we have $\pi_{PP}^* - \pi_{2H}^* < 0$. Therefore, for the airline, π_{2H}^* is higher and then 2H is more profitable if $t \in \left(\max\left\{0, \underline{t_{2H}}\right\}, \overline{t_{2H}}\right)$. Moreover, π_{2H}^* does not exist and then PP is more profitable if $t \in \left(0, \underline{t_{2H}}\right] \bigcup \left[\overline{t_{2H}}, +\infty\right)$ when $\psi \in \left(\psi_{2H}^{soc}, \frac{5\lambda^2}{4}\right)$ and $t \in \left[\overline{t_{2H}}, +\infty\right)$ when $\psi \in \left[\frac{5\lambda^2}{4}, +\infty\right)$. This result is shown in Figure 2.26.

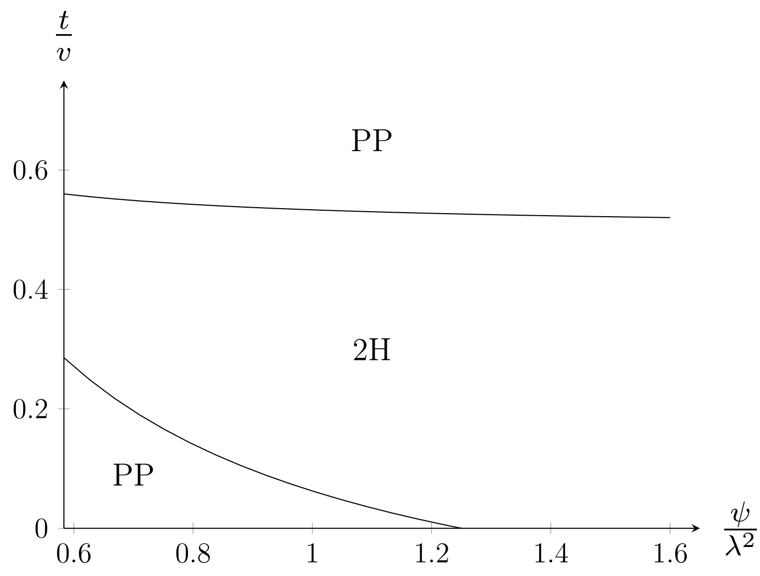

Figure 2.26: Comparison of Profits Between PP and 2H

Besides, under $\psi \in (\psi_{2H}^{soc}, +\infty)$, we have $\pi_{MX}^{*} - \pi_{2H}^{*} < 0$. We can then verify $\underline{t_{MX}} > \underline{t_{2H}}$ and $\overline{t_{MX}} = \overline{t_{2H}}$. Therefore, for the airline, MX is always less profitable than 2H. □

2.8.2.5 Proof of Proposition 2.8

Proof. The comparison between $t_{HS\sim PP}^{\pi}$ and $\underline{t_{2H}}$ is the key to the proof.

Let $\lambda^2 = x\psi$, in which $x \in \left(\frac{4}{5}, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)$. Then, we have:

$$t_{HS\sim PP}^{\pi} - \underline{t_{2H}} = \frac{\sigma_3(x)\, v}{16(-x+16)(-x+12)},$$

in which $(-x+16)(-x+12) > 0$ and:

$$\sigma_3(x) = -5x^3 + 144x^2 - 1264x + 3840 - 96(-x+8)\sqrt{-x+16}.$$

By considering the gradients and extrema of $\frac{\mathrm{d}^n \sigma_3(x)}{\mathrm{d}x^n}$, we can find the gradients of $\frac{\mathrm{d}^{n-1}\sigma_3(x)}{\mathrm{d}x^{n-1}}$. Then, starting from $\frac{\mathrm{d}^4\sigma_3(x)}{\mathrm{d}x^4}$ and inferring step by step, there will be $\frac{\mathrm{d}\sigma_3(x)}{\mathrm{d}x} < 0$ and:

$$\sup_{x\in\left(\frac{4}{5}, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)} \sigma_3(x) > 0, \qquad \inf_{x\in\left(\frac{4}{5}, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)} \sigma_3(x) < 0. \tag{2.66}$$

According to (2.66), $\exists\ \varepsilon_1, \varepsilon_2 > 0$ such that $\sigma_3\left(\frac{4}{5}+\varepsilon_1\right) > 0 > \sigma_3\left(\frac{\lambda^2}{\psi_{2H}^{soc}}-\varepsilon_2\right)$. Note that $\sigma_3(x)$ is real-valued and continuous on the compact interval $\left[\frac{4}{5}+\varepsilon_1, \frac{\lambda^2}{\psi_{2H}^{soc}}-\varepsilon_2\right]$ in $\mathbb{R}$. Then, according to *Bolzano's theorem*, $\exists\ x' \in \left(\frac{4}{5}+\varepsilon_1, \frac{\lambda^2}{\psi_{2H}^{soc}}-\varepsilon_2\right)$, such that $\sigma_3(x') = 0$. Moreover, because $\sigma_3(x)$ strictly decreases in x, x' is unique. Meanwhile, we have $\sigma_3(x) > 0,\ \forall\ x \in \left(\frac{4}{5}, \frac{4}{5}+\varepsilon_1\right)$, and $\sigma_3(x) < 0,\ \forall\ x \in \left(\frac{\lambda^2}{\psi_{2H}^{soc}}-\varepsilon_2, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)$. Therefore, $\exists$ a unique $x' \in \left(\frac{4}{5}, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)$, such that $\sigma_3(x') = 0$. Solving $\sigma_3(x) = 0$ when $x \in \left(\frac{4}{5}, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)$ gives $x' = -\frac{\xi_1 - 111}{15}$, in which:

$$\xi_1 = \sqrt{3(8\zeta_1+179)} + \sqrt{6\left(-4\zeta_1+179+7605\sqrt{\frac{3}{8\zeta_1+179}}\right)},$$
$$\zeta_1 = \sqrt[3]{190657-135\sqrt{1966530}} + \sqrt[3]{190657+135\sqrt{1966530}}.$$

Therefore, we have $t_{HS\sim PP}^{\pi} < \underline{t_{2H}}$ when $\psi \in \left(\psi_{2H}^{soc}, -\frac{15\lambda^2}{\xi_1-111}\right)$ and $t_{HS\sim PP}^{\pi} \geqslant \underline{t_{2H}}$ when $\psi \in \left[-\frac{15\lambda^2}{\xi_1-111}, \frac{5\lambda^2}{4}\right)$. Then, the order of relevant critical values is in Table 2.12. Combining the results in Propositions 2.2 and 2.7, we can obtain Proposition 2.8.

Interval of ψ	Order of Critical Values
$\left(\psi_{2H}^{soc}, -\frac{15\lambda^2}{\xi_1-111}\right)$	$0 < t_{HS\sim PP}^{\pi} < \underline{t_{2H}} < \overline{t_{2H}} < \overline{t_{HS}}$
$\left[-\frac{15\lambda^2}{\xi_1-111}, \frac{5\lambda^2}{4}\right)$	$0 < \underline{t_{2H}} \leqslant t_{HS\sim PP}^{\pi} < \overline{t_{2H}} < \overline{t_{HS}}$
$\left[\frac{5\lambda^2}{4}, +\infty\right)$	$0 < t_{HS\sim PP}^{\pi} < \overline{t_{2H}} < \overline{t_{HS}}$

Table 2.12: Order of Critical Values (Proposition 2.8)

□

2.9 Appendix C: Proofs and Solutions of Welfare Analysis

2.9.1 Proofs of Second-Best Socially Optimal Network Structure

2.9.1.1 Proof of Lemma 2.1

(I) Comparison Between sw_{HS}^* and sw_{PP}^*

Proof. Under $\psi \in (\psi_{2H}^{soc}, +\infty)$, the sign of $sw_{HS}^{*} - sw_{PP}^{*}$ depends on a quadratic function of t. As the coefficient of t^2 is positive, the parabola opens upwards. Meanwhile, let $\lambda^2 = x\psi$, in which $x \in \left(0, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)$. Then, for the discriminant, we have $\text{sign}(\Delta) = \text{sign}(\sigma_4(x))$, in which $\sigma_4(x)$ is a polynomial of degree 3 and is positive. Thus, the discriminant is positive.

Then, solving $sw_{HS}^{*} - sw_{PP}^{*} = 0$ gives two roots, in which the smaller root is less than $\overline{t_{HS}}$, while the larger root is greater than $\overline{t_{HS}}$. According to the constraint of t of HS $t \in \left(0, \overline{t_{HS}}\right)$, only the smaller root is relevant, denoted by:

$$t_{HS\sim PP}^{sw} = \frac{6v}{288\psi^2 - 56\lambda^2\psi + 3\lambda^4}\left[4\psi\left(12\psi - \lambda^2\right) - \frac{\left(8\psi - \lambda^2\right)\sqrt{2\psi\left(4608\psi^3 - 1344\lambda^2\psi^2 + 124\lambda^4\psi - 3\lambda^6\right)}}{16\psi - \lambda^2}\right].$$

Therefore, sw_{HS}^{*} is higher and then HS is more socially desirable if $t \in (0, t_{HS\sim PP}^{sw}]$; sw_{PP}^{*} is higher and then PP is more socially desirable if $t \in \left(t_{HS\sim PP}^{sw}, \overline{t_{HS}}\right)$. Moreover, sw_{HS}^{*} does not exist and then PP is more socially desirable if $t \in \left[\overline{t_{HS}}, +\infty\right)$. □

(II) Comparison Between sw_{HS}^{*} and sw_{MX}^{*}

Proof. Under $\psi \in (\psi_{2H}^{soc}, +\infty)$, the sign of $sw_{HS}^{*} - sw_{MX}^{*}$ depends on a quadratic function of t. Let $\lambda^2 = x\psi$, in which $x \in \left(0, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)$. Then, the coefficient of t^2 becomes $2\sigma_5(x)\psi^6$, in which $\sigma_5(x)$ is a polynomial of degree 5 and is negative. Thus, the parabola opens downwards. Meanwhile, for the discriminant, we have $\text{sign}(\Delta) = \text{sign}(\sigma_6(x))$, in which $\sigma_6(x)$ is a polynomial of degree 3 and is negative. Then, the discriminant is negative.

To summarize, the quadratic function of t is always negative. Therefore, sw_{MX}^{*} is higher and then MX is more socially desirable if $t \in \left(\underline{t_{MX}}, \overline{t_{MX}}\right)$. Moreover, sw_{MX}^{*} does not exist and then HS is more socially desirable if $t \in \left(0, \underline{t_{MX}}\right] \bigcup \left[\overline{t_{MX}}, \overline{t_{HS}}\right)$. □

(III) Comparison Between sw_{HS}^{*} and sw_{2H}^{*}

Proof. Under $\psi \in (\psi_{2H}^{soc}, +\infty)$, the sign of $sw_{HS}^{*} - sw_{2H}^{*}$ depends on a quadratic function of t. Let $\lambda^2 = x\psi$, in which $x \in \left(0, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)$. Then, the coefficient of t^2 becomes $\sigma_7(x)\psi^6$, in which $\sigma_7(x)$ is a polynomial of degree 6. By considering the gradients and extrema of

$\frac{\mathrm{d}^n\sigma_7(x)}{\mathrm{d}x^n}$, we can find the gradients of $\frac{\mathrm{d}^{n-1}\sigma_7(x)}{\mathrm{d}x^{n-1}}$. Then, starting from $\frac{\mathrm{d}^5\sigma_7(x)}{\mathrm{d}x^5}$ and inferring step by step, there will be $\frac{\mathrm{d}\sigma_7(x)}{\mathrm{d}x} > 0$ and:

$$\sup_{x \in \left(0, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)} \sigma_7(x) < 0. \tag{2.67}$$

According to (2.67), we can obtain $\sigma_7(x) < 0$. Thus, the parabola opens downwards.

Meanwhile, for the discriminant, we have $\mathrm{sign}(\Delta) = \mathrm{sign}(\sigma_8(x))$, in which $\sigma_8(x)$ is a polynomial of degree 5. By considering the gradients and extrema of $\frac{\mathrm{d}^n\sigma_8(x)}{\mathrm{d}x^n}$, we can find the gradients of $\frac{\mathrm{d}^{n-1}\sigma_8(x)}{\mathrm{d}x^{n-1}}$. Then, starting from $\frac{\mathrm{d}^4\sigma_8(x)}{\mathrm{d}x^4}$ and inferring step by step, there will be $\frac{\mathrm{d}\sigma_8(x)}{\mathrm{d}x} > 0$ and:

$$\sup_{x \in \left(0, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)} \sigma_8(x) < 0. \tag{2.68}$$

According to (2.68), we can obtain $\sigma_8(x) < 0$. Then, the discriminant is negative.

To summarize, the quadratic function of t is always negative. Therefore, sw_{2H}^* is higher and then 2H is more socially desirable if $t \in \left(\max\left\{0, \underline{t_{2H}}\right\}, \overline{t_{2H}}\right)$. Moreover, sw_{2H}^* does not exist and then HS is more socially desirable if $t \in \left(0, \underline{t_{2H}}\right] \bigcup \left[\overline{t_{2H}}, \overline{t_{HS}}\right)$ when $\psi \in \left(\psi_{2H}^{soc}, \frac{5\lambda^2}{4}\right)$ and $t \in \left[\overline{t_{2H}}, \overline{t_{HS}}\right)$ when $\psi \in \left[\frac{5\lambda^2}{4}, +\infty\right)$. □

(IV) Comparison Between sw_{PP}^* and sw_{MX}^*

Proof. Under $\psi \in (\psi_{2H}^{soc}, +\infty)$, the sign of $sw_{PP}^* - sw_{MX}^*$ depends on a quadratic function of t. Let $\lambda^2 = x\psi$, in which $x \in \left(0, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)$. Then, the coefficient of t^2 becomes $\sigma_9(x)\psi^6$, in which $\sigma_9(x)$ is a polynomial of degree 6. By considering the gradients and extrema of $\frac{\mathrm{d}^n\sigma_9(x)}{\mathrm{d}x^n}$, we can find the gradients of $\frac{\mathrm{d}^{n-1}\sigma_9(x)}{\mathrm{d}x^{n-1}}$. Then, starting from $\frac{\mathrm{d}^5\sigma_9(x)}{\mathrm{d}x^5}$ and inferring step by step, there will be $\frac{\mathrm{d}\sigma_9(x)}{\mathrm{d}x} > 0$ and:

$$\sup_{x \in \left(0, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)} \sigma_9(x) < 0. \tag{2.69}$$

According to (2.69), we can obtain $\sigma_9(x) < 0$. Thus, the parabola opens downwards.

Meanwhile, the discriminant:

$$\Delta = 256\lambda^4\psi^2v^2\left(4\psi - \lambda^2\right)^2(\phi_{PP})^2(\phi_{MX})^2,$$

is positive.

Then, solving $sw^*_{PP} - sw^*_{MX} = 0$ gives two roots, in which the smaller root equals $\overline{t_{MX}}$. Therefore, sw^*_{MX} is higher and then MX is more socially desirable if $t \in \left(\underline{t_{MX}}, \overline{t_{MX}}\right)$. Moreover, sw^*_{MX} does not exist and then PP is more socially desirable if $t \in \left(0, \underline{t_{MX}}\right] \bigcup \left[\overline{t_{MX}}, +\infty\right)$. □

(V) Comparison Between sw^*_{PP} and sw^*_{2H}

Proof. Under $\psi \in (\psi^{soc}_{2H}, +\infty)$, the sign of $sw^*_{PP} - sw^*_{2H}$ depends on a quadratic function of t. Let $\lambda^2 = x\psi$, in which $x \in \left(0, \frac{\lambda^2}{\psi^{soc}_{2H}}\right)$. Then, the coefficient of t^2 becomes $\sigma_{10}(x)\,\psi^6$, in which $\sigma_{10}(x)$ is a polynomial of degree 6. By considering the gradients and extrema of $\frac{\mathrm{d}^n\sigma_{10}(x)}{\mathrm{d}x^n}$, we can find the gradients of $\frac{\mathrm{d}^{n-1}\sigma_{10}(x)}{\mathrm{d}x^{n-1}}$. Then, starting from $\frac{\mathrm{d}^5\sigma_{10}(x)}{\mathrm{d}x^5}$ and inferring step by step, there will be $\frac{\mathrm{d}\sigma_{10}(x)}{\mathrm{d}x} > 0$ and:

$$\sup_{x\in\left(0,\frac{\lambda^2}{\psi^{soc}_{2H}}\right)} \sigma_{10}(x) < 0. \tag{2.70}$$

According to (2.70), we can obtain $\sigma_{10}(x) < 0$. Thus, the parabola opens downwards. Meanwhile, the discriminant:

$$\Delta = 256\lambda^4\psi^2v^2\left(4\psi - \lambda^2\right)^2(\phi_{PP})^2(\phi_{2H})^2,$$

is positive.

Then, solving $sw^*_{PP} - sw^*_{2H} = 0$ gives two roots, in which the smaller root equals $\overline{t_{2H}}$. Therefore, sw^*_{2H} is higher and then 2H is more socially desirable if $t \in \left(\max\left\{0, \underline{t_{2H}}\right\}, \overline{t_{2H}}\right)$. Moreover, sw^*_{2H} does not exist and then PP is more socially desirable if $t \in \left(0, \underline{t_{2H}}\right] \bigcup \left[\overline{t_{2H}}, +\infty\right)$ when $\psi \in \left(\psi^{soc}_{2H}, \frac{5\lambda^2}{4}\right)$ and $t \in \left[\overline{t_{2H}}, +\infty\right)$ when $\psi \in \left[\frac{5\lambda^2}{4}, +\infty\right)$. □

(VI) Comparison Between sw^*_{MX} and sw^*_{2H}

Proof. Under $\psi \in (\psi^{soc}_{2H}, +\infty)$, the sign of $sw^*_{MX} - sw^*_{2H}$ depends on a quadratic function of t. Let $\lambda^2 = x\psi$, in which $x \in \left(0, \frac{\lambda^2}{\psi^{soc}_{2H}}\right)$. Then, the coefficient of t^2 becomes $\sigma_{11}(x)\psi^8$, in which $\sigma_{11}(x)$ is a polynomial of degree 8. By considering the gradients and extrema of $\frac{\mathrm{d}^n\sigma_{11}(x)}{\mathrm{d}x^n}$, we can find the gradients of $\frac{\mathrm{d}^{n-1}\sigma_{11}(x)}{\mathrm{d}x^{n-1}}$. Then, starting from $\frac{\mathrm{d}^7\sigma_{11}(x)}{\mathrm{d}x^7}$ and inferring step by step, there will be $\frac{\mathrm{d}\sigma_{11}(x)}{\mathrm{d}x} > 0$ and:

$$\sup_{x\in\left(0, \frac{\lambda^2}{\psi^{soc}_{2H}}\right)} \sigma_{11}(x) > 0, \qquad \inf_{x\in\left(0, \frac{\lambda^2}{\psi^{soc}_{2H}}\right)} \sigma_{11}(x) < 0. \tag{2.71}$$

According to (2.71), $\exists\ \varepsilon_3, \varepsilon_4 > 0$ such that $\sigma_{11}(\varepsilon_3) < 0 < \sigma_{11}\left(\frac{\lambda^2}{\psi^{soc}_{2H}} - \varepsilon_4\right)$. Note that $\sigma_{11}(x)$ is real-valued and continuous on the compact interval $\left[\varepsilon_3, \frac{\lambda^2}{\psi^{soc}_{2H}} - \varepsilon_4\right]$ in $\mathbb{R}$. Then, according to *Bolzano's theorem*, $\exists\ x'' \in \left(\varepsilon_3, \frac{\lambda^2}{\psi^{soc}_{2H}} - \varepsilon_4\right)$, such that $\sigma_{11}(x'') = 0$. Moreover, because $\sigma_{11}(x)$ strictly increases in x, x'' is unique. Meanwhile, we have $\sigma_{11} < 0$, $\forall\ x \in (0, \varepsilon_3)$, and $\sigma_{11} > 0$, $\forall\ x \in \left(\frac{\lambda^2}{\psi^{soc}_{2H}} - \varepsilon_4, \frac{\lambda^2}{\psi^{soc}_{2H}}\right)$. Therefore, $\exists$ a unique $x'' \in \left(0, \frac{\lambda^2}{\psi^{soc}_{2H}}\right)$, such that $\sigma_{11}(x'') = 0$. Solving $\sigma_{11}(x) = 0$ when $x \in \left(0, \frac{\lambda^2}{\psi^{soc}_{2H}}\right)$ gives $x'' = \mu_1\ (\approx 1.47)$.

Thus, we can obtain $\sigma_{11}(x) > 0$ (resp. $\sigma_{11}(x) < 0$) and then the parabola opens upwards (resp. downwards) when $\psi \in \left(\psi^{soc}_{2H}, \frac{\lambda^2}{\mu_1}\right)$ (resp. $\psi \in \left(\frac{\lambda^2}{\mu_1}, +\infty\right)$). Moreover, we can obtain $\sigma_{11}(x) = 0$ and then the sign of $sw^*_{MX} - sw^*_{2H}$ depends on a linear function of t when $\psi = \frac{\lambda^2}{\mu_1}$.

When $\psi = \frac{\lambda^2}{\mu_1}$, we have an increasing linear function of t, which equals zero at $\overline{t_{2H}}\ (= \overline{t_{MX}})$. Therefore, sw^*_{MX} is lower than sw^*_{2H} if $t \in (\underline{t_{MX}}, \overline{t_{MX}})$ when $\psi = \frac{\lambda^2}{\mu_1}$.

When $\psi \in \left(\psi^{soc}_{2H}, \frac{\lambda^2}{\mu_1}\right) \bigcup \left(\frac{\lambda^2}{\mu_1}, +\infty\right)$, the discriminant of the quadratic function of t:

$$\Delta = 256\lambda^4\psi^2 v^2 \left(12\psi - 7\lambda^2\right)^2 \left(4\psi - \lambda^2\right)^2 (\phi_{MX})^2 (\phi_{2H})^2,$$

is positive. Then, solving $sw^*_{MX} - sw^*_{2H} = 0$ gives two roots, in which the first root equals $\overline{t_{2H}}$. However, we cannot determine the value of the second root.

Consider first $\psi \in \left(\psi^{soc}_{2H}, \frac{\lambda^2}{\mu_1}\right)$, that is, the parabola opens upwards. Let $\lambda^2 = x\psi$, in which $x \in \left(\mu_1, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)$. By comparing the second root with 0, $\underline{t_{MX}}$ and $\overline{t_{2H}}$, we have that the second root is non-positive when $x \in (\mu_1, \mu_2]$; positive but not greater than $\underline{t_{MX}}$ when $x \in (\mu_2, \xi_2]$; and greater than $\underline{t_{MX}}$ but less than $\overline{t_{2H}}$ when $x \in \left(\xi_2, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)$, in which $\mu_2\,(\approx 1.48)$ is a unique real root of:

$$105x^6 - 7168x^5 + 133664x^4 - 939264x^3 + 3071232x^2 - 4681728x + 2654208 = 0,$$

when $x \in \left(\mu_1, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)$; and $\xi_2\,(\approx 1.51)$ is a unique real root of:

$$63x^5 - 2172x^4 + 26192x^3 - 131136x^2 + 291840x - 221184 = 0,$$

when $x \in \left(\mu_2, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)$.

Therefore, sw_{MX}^* is higher (resp. lower) than sw_{2H}^* if $t \in \left(\underline{t_{MX}}, t_{MX\sim 2H}^{sw}\right]$ (resp. $t \in \left(t_{MX\sim 2H}^{sw}, \overline{t_{2H}}\right)$) when $\psi \in \left(\psi_{2H}^{soc}, \frac{\lambda^2}{\xi_2}\right)$ and sw_{MX}^* is lower than sw_{2H}^* if $t \in \left(\underline{t_{MX}}, \overline{t_{2H}}\right)$ when $\psi \in \left[\frac{\lambda^2}{\xi_2}, \frac{\lambda^2}{\mu_1}\right)$, in which $t_{MX\sim 2H}^{sw}$ denotes the second root when $\psi \in \left(\psi_{2H}^{soc}, \frac{\lambda^2}{\xi_2}\right)$, and $t_{MX\sim 2H}^{sw} = \frac{8\psi\zeta_2 v}{\zeta_3}$, with:

$$\begin{aligned}\zeta_2 &= 2654208\psi^6 - 4681728\lambda^2\psi^5 + 3071232\lambda^4\psi^4 \\ &\quad - 939264\lambda^6\psi^3 + 133664\lambda^8\psi^2 - 7168\lambda^{10}\psi + 105\lambda^{12}, \\ \zeta_3 &= 42467328\psi^7 - 79331328\lambda^2\psi^6 + 57028608\lambda^4\psi^5 - 20322048\lambda^6\psi^4 \\ &\quad + 3813632\lambda^8\psi^3 - 366688\lambda^{10}\psi^2 + 17216\lambda^{12}\psi - 315\lambda^{14}.\end{aligned}$$

Consider then $\psi \in \left(\frac{\lambda^2}{\mu_1}, +\infty\right)$, that is, the parabola opens downwards. Let $\lambda^2 = x\psi$, in which $x \in (0, \mu_1)$. By comparing the second root with $\overline{t_{2H}}$, we have that the second root is always greater than $\overline{t_{MX}}$. Therefore, sw_{MX}^* is lower than sw_{2H}^* if $t \in \left(\underline{t_{MX}}, \overline{t_{2H}}\right)$ when $\psi \in \left(\frac{\lambda^2}{\mu_1}, +\infty\right)$.

To summarize, first, sw_{MX}^* is higher (resp. lower) and then MX (resp. 2H) is more socially desirable if $t \in \left(\underline{t_{MX}}, t_{MX\sim 2H}^{sw}\right]$ (resp. $t \in \left(t_{MX\sim 2H}^{sw}, \overline{t_{2H}}\right)$) when $\psi \in \left(\psi_{2H}^{soc}, \frac{\lambda^2}{\xi_2}\right)$; second, sw_{MX}^* is lower and then 2H is more socially desirable if $t \in \left(\underline{t_{MX}}, \overline{t_{2H}}\right)$ when $\psi \in \left[\frac{\lambda^2}{\xi_2}, +\infty\right)$; and third, sw_{MX}^* does not exist and then 2H is more socially desirable if

$t \in \left(\max\left\{0, \underline{t_{2H}}\right\}, \underline{t_{MX}}\right]$ under $\psi \in (\psi_{2H}^{soc}, +\infty)$. □

(VII) Order of Critical Values

Proof. The comparison between $t_{HS\sim PP}^{sw}$ and $\underline{t_{MX}}$ ($\underline{t_{2H}}$) is the key to the proof.

Let $\lambda^2 = x\psi$, in which $x \in \left(0, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)$. Then, we have:

$$t_{HS\sim PP}^{sw} - \underline{t_{MX}} = \frac{\sigma_{12}(x)\, v}{(-x+20)(-x+16)(3x^2-56x+288)},$$

in which $(-x+20)(-x+16)(3x^2-56x+288) > 0$ and:

$$\begin{aligned}\sigma_{12}(x) &= 9x^4 - 324x^3 + 4288x^2 - 27136x + 73728 \\ &\quad - 6(-x+20)(-x+8)\sqrt{2(-3x^3+124x^2-1344x+4608)}.\end{aligned}$$

By considering the gradients and extrema of $\frac{\mathrm{d}^n\sigma_{12}(x)}{\mathrm{d}x^n}$, we can find the gradients of $\frac{\mathrm{d}^{n-1}\sigma_{12}(x)}{\mathrm{d}x^{n-1}}$. Then, starting from $\frac{\mathrm{d}^5\sigma_{12}(x)}{\mathrm{d}x^5}$, we have sign $\left(\frac{\mathrm{d}^5\sigma_{12}(x)}{\mathrm{d}x^5}\right) = \text{sign}(\sigma_{13}(x))$, in which $\sigma_{13}(x)$ is a polynomial of degree 12. By considering the gradients and extrema of $\frac{\mathrm{d}^n\sigma_{13}(x)}{\mathrm{d}x^n}$, we can find the gradients of $\frac{\mathrm{d}^{n-1}\sigma_{13}(x)}{\mathrm{d}x^{n-1}}$. Then, starting from $\frac{\mathrm{d}^{11}\sigma_{13}(x)}{\mathrm{d}x^{11}}$ and inferring step by step, there will be $\frac{\mathrm{d}\sigma_{13}(x)}{\mathrm{d}x} > 0$ and:

$$\sup_{x\in\left(0,\frac{\lambda^2}{\psi_{2H}^{soc}}\right)} \sigma_{13}(x) > 0, \qquad \inf_{x\in\left(0,\frac{\lambda^2}{\psi_{2H}^{soc}}\right)} \sigma_{13}(x) < 0. \tag{2.72}$$

According to (2.72), $\exists\ \varepsilon_5, \varepsilon_6 > 0$ such that $\sigma_{13}(\varepsilon_5) < 0 < \sigma_{13}\left(\frac{\lambda^2}{\psi_{2H}^{soc}} - \varepsilon_6\right)$. Note that $\sigma_{13}(x)$ is real-valued and continuous on the compact interval $\left[\varepsilon_5, \frac{\lambda^2}{\psi_{2H}^{soc}} - \varepsilon_6\right]$ in $\mathbb{R}$. Then, according to *Bolzano's theorem*, $\exists\ x''' \in \left(\varepsilon_5, \frac{\lambda^2}{\psi_{2H}^{soc}} - \varepsilon_6\right)$, such that $\sigma_{13}\left(x'''\right) = 0$. Moreover, because $\sigma_{13}(x)$ strictly increases in x, x''' is unique. Meanwhile, we have $\sigma_{13}(x) < 0$, $\forall\ x \in (0, \varepsilon_5)$, and $\sigma_{13}(x) > 0$, $\forall\ x \in \left(\frac{\lambda^2}{\psi_{2H}^{soc}} - \varepsilon_6, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)$. Therefore, $\exists$ a unique $x''' \in \left(0, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)$, such that $\sigma_{13}\left(x'''\right) = 0$. Solving $\sigma_{13}(x) = 0$ when $x \in \left(0, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)$ gives $x''' = \mu_3\,(\approx 0.98)$.

Then, we have $\frac{\mathrm{d}^5\sigma_{12}(x)}{\mathrm{d}x^5} \leqslant 0$ when $x \in (0, \mu_3]$ and $\frac{\mathrm{d}^5\sigma_{12}(x)}{\mathrm{d}x^5} > 0$ when $x \in \left(\mu_3, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)$.

Consequently, because of $\frac{\mathrm{d}^4\sigma_{12}(x)}{\mathrm{d}x^4} > 0$ when $x''' = \mu_3$, we have $\frac{\mathrm{d}^4\sigma_{12}(x)}{\mathrm{d}x^4} > 0$ when $x \in \left(0, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)$. Continuing to infer step by step, finally, there will be $\frac{\mathrm{d}\sigma_{12}(x)}{\mathrm{d}x} > 0$ and:

$$\sup_{x\in\left(0,\frac{\lambda^2}{\psi_{2H}^{soc}}\right)} \sigma_{12}(x) < 0. \tag{2.73}$$

According to (2.73), we can obtain $\sigma_{12}(x) < 0$. Therefore, we have $t_{HS\sim PP}^{sw} < \underline{t_{MX}}$.

Let $x \in \left(\frac{4}{5}, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)$. Then, we have:

$$t_{HS\sim PP}^{sw} - \underline{t_{2H}} = \frac{\sigma_{14}(x)\, v}{16(-x+16)(3x^2-56x+288)},$$

in which $(-x+16)(3x^2-56x+288) > 0$ and:

$$\begin{aligned}\sigma_{14}(x) = {} & 15x^4 - 532x^3 + 6720x^2 - 38528x + 92160 \\ & - 96(-x+8)\sqrt{2(-3x^3+124x^2-1344x+4608)}.\end{aligned}$$

By considering the gradients and extrema of $\frac{\mathrm{d}^n\sigma_{14}(x)}{\mathrm{d}x^n}$, we can find the gradients of $\frac{\mathrm{d}^{n-1}\sigma_{14}(x)}{\mathrm{d}x^{n-1}}$. Then, starting from $\frac{\mathrm{d}^5\sigma_{14}(x)}{\mathrm{d}x^5}$, we have sign $\left(\frac{\mathrm{d}^5\sigma_{14}(x)}{\mathrm{d}x^5}\right) = \mathrm{sign}(\sigma_{15}(x))$, in which $\sigma_{15}(x)$ is a polynomial of degree 11. By considering the gradients and extrema of $\frac{\mathrm{d}^n\sigma_{15}(x)}{\mathrm{d}x^n}$, we can find the gradients of $\frac{\mathrm{d}^{n-1}\sigma_{15}(x)}{\mathrm{d}x^{n-1}}$. Then, starting from $\frac{\mathrm{d}^{10}\sigma_{15}(x)}{\mathrm{d}x^{10}}$ and inferring step by step, there will be $\frac{\mathrm{d}\sigma_{15}(x)}{\mathrm{d}x} < 0$ and:

$$\inf_{x\in\left(\frac{4}{5},\frac{\lambda^2}{\psi_{2H}^{soc}}\right)} \sigma_{15}(x) > 0. \tag{2.74}$$

According to (2.74), we can obtain $\sigma_{15}(x) > 0$. Thus, we have $\frac{\mathrm{d}^5\sigma_{14}(x)}{\mathrm{d}x^5} > 0$ when $x \in \left(\frac{4}{5}, \frac{\lambda^2}{\psi_{2H}^{soc}}\right)$. Continuing to infer step by step, finally, there will be $\frac{\mathrm{d}\sigma_{14}(x)}{\mathrm{d}x} < 0$ and:

$$\sup_{x\in\left(\frac{4}{5},\frac{\lambda^2}{\psi_{2H}^{soc}}\right)} \sigma_{14}(x) > 0, \qquad \inf_{x\in\left(\frac{4}{5},\frac{\lambda^2}{\psi_{2H}^{soc}}\right)} \sigma_{14}(x) < 0. \tag{2.75}$$

According to (2.75), $\exists\ \varepsilon_7, \varepsilon_8 > 0$ such that $\sigma_{14}\left(\frac{4}{5}+\varepsilon_7\right) > 0 > \sigma_{14}\left(\frac{\lambda^2}{\psi_{2H}^{soc}} - \varepsilon_8\right)$. Note that $\sigma_{14}(x)$ is real-valued and continuous on the compact interval $\left[\frac{4}{5}+\varepsilon_7, \frac{\lambda^2}{\psi_{2H}^{soc}}-\varepsilon_8\right]$ in $\mathbb{R}$. Then, according to *Bolzano's theorem*, $\exists\, x^{(4)} \in \left(\frac{4}{5}+\varepsilon_7, \frac{\lambda^2}{\psi_{2H}^{soc}}-\varepsilon_8\right)$, such that $\sigma_{14}\left(x^{(4)}\right)=0$. Moreover, because $\sigma_{14}(x)$ strictly decreases in x, $x^{(4)}$ is unique. Meanwhile, we have $\sigma_{14}(x)>0$, $\forall\, x\in\left(\frac{4}{5},\frac{4}{5}+\varepsilon_7\right)$, and $\sigma_{14}(x)<0$, $\forall\, x\in\left(\frac{\lambda^2}{\psi_{2H}^{soc}}-\varepsilon_8,\frac{\lambda^2}{\psi_{2H}^{soc}}\right)$. Therefore, $\exists$ a unique $x^{(4)}\in\left(\frac{4}{5},\frac{\lambda^2}{\psi_{2H}^{soc}}\right)$, such that $\sigma_{14}\left(x^{(4)}\right)=0$. Solving $\sigma_{14}(x)=0$ when $x\in\left(\frac{4}{5},\frac{\lambda^2}{\psi_{2H}^{soc}}\right)$ gives $x^{(4)}=\xi_3\,(\approx 1.59)$.

Therefore, we have $t_{HS\sim PP}^{sw} < \underline{t_{2H}}$ when $\psi\in\left(\psi_{2H}^{soc},\frac{\lambda^2}{\xi_3}\right)$ and $t_{HS\sim PP}^{sw}\geqslant \underline{t_{2H}}$ when $\psi\in\left[\frac{\lambda^2}{\xi_3},\frac{5\lambda^2}{4}\right)$.

Then, the order of relevant critical values is in Table 2.13. Combining the results in Parts (I) through (VI), we can obtain Lemma 2.1.

Interval of ψ	Order of Critical Values
$\left(\psi_{2H}^{soc},\frac{\lambda^2}{\xi_3}\right)$	$0<t_{HS\sim PP}^{sw}<\underline{t_{2H}}<\underline{t_{MX}}<t_{MX\sim 2H}^{sw}<\overline{t_{2H}}<\overline{t_{HS}}$
$\left[\frac{\lambda^2}{\xi_3},\frac{\lambda^2}{\xi_2}\right)$	$0<\underline{t_{2H}}\leqslant t_{HS\sim PP}^{sw}<\underline{t_{MX}}<t_{MX\sim 2H}^{sw}<\overline{t_{2H}}<\overline{t_{HS}}$
$\left[\frac{\lambda^2}{\xi_2},\frac{5\lambda^2}{4}\right)$	$0<\underline{t_{2H}}<t_{HS\sim PP}^{sw}<\underline{t_{MX}}<\overline{t_{2H}}<\overline{t_{HS}}$
$\left[\frac{5\lambda^2}{4},+\infty\right)$	$0<t_{HS\sim PP}^{sw}<\underline{t_{MX}}<\overline{t_{2H}}<\overline{t_{HS}}$

Table 2.13: Order of Critical Values (Lemma 2.1)

□

2.9.1.2 Proof of Proposition 9

Proof. The comparison between $t_{HS\sim PP}^{sw}$ and $t_{HS\sim PP}^{\pi}$ is the key to the proof.

Let $\lambda^2=x\psi$, in which $x\in\left(0,\frac{\lambda^2}{\psi_{2H}^{soc}}\right)$. Then, we have:

$$t_{HS\sim PP}^{sw}-t_{HS\sim PP}^{\pi}=\frac{6(-x+8)\,\sigma_{16}(x)\,v}{(-x+16)(-x+12)\left(3x^2-56x+288\right)},$$

in which $-x+8>0$, $(-x+16)(-x+12)\left(3x^2-56x+288\right)>0$ and:

$$\sigma_{16}(x)=\zeta_4+\zeta_5-\zeta_6,$$

$$\zeta_4=32x-2x^2>0,$$

$$\zeta_5 = \left(3x^2 - 56x + 288\right)\sqrt{-x+16} > 0,$$
$$\zeta_6 = (-x+12)\sqrt{2\left(-3x^3 + 124x^2 - 1344x + 4608\right)} > 0.$$

Moreover, we have:

$$\begin{aligned} &\zeta_5 - \zeta_6 > 0 \\ &\quad \Leftrightarrow (\zeta_5)^2 - (\zeta_6)^2 = x(-x+8)\left(3x^3 - 64x^2 + 224x + 1152\right) > 0. \end{aligned} \tag{2.76}$$

According to (2.76), we can obtain $\sigma_{16}(x) > 0$. Therefore, we have $t^{sw}_{HS\sim PP} > t^{\pi}_{HS\sim PP}$.

Then, the order of relevant critical values is in Table 2.14. Combining the results in Proposition 2.8 and Lemma 2.1, we can obtain Proposition 2.9.

Interval of ψ	Order of Critical Values
$\left(\psi^{soc}_{2H}, \frac{\lambda^2}{\xi_3}\right)$	$0 < t^{\pi}_{HS\sim PP} < t^{sw}_{HS\sim PP} < \underline{t_{2H}} < \underline{t_{MX}} < t^{sw}_{MX\sim 2H} < \overline{t_{2H}} < \overline{t_{HS}}$
$\left[\frac{\lambda^2}{\xi_3}, \frac{\lambda^2}{\xi_2}\right)$	$0 < t^{\pi}_{HS\sim PP} < \underline{t_{2H}} \leqslant t^{sw}_{HS\sim PP} < \underline{t_{MX}} < t^{sw}_{MX\sim 2H} < \overline{t_{2H}} < \overline{t_{HS}}$
$\left[\frac{\lambda^2}{\xi_2}, -\frac{15\lambda^2}{\xi_1 - 111}\right)$	$0 < t^{\pi}_{HS\sim PP} < \underline{t_{2H}} < t^{sw}_{HS\sim PP} < \underline{t_{MX}} < \overline{t_{2H}} < \overline{t_{HS}}$
$\left[-\frac{15\lambda^2}{\xi_1 - 111}, \frac{5\lambda^2}{4}\right)$	$0 < \underline{t_{2H}} \leqslant t^{\pi}_{HS\sim PP} < t^{sw}_{HS\sim PP} < \underline{t_{MX}} < \overline{t_{2H}} < \overline{t_{HS}}$
$\left[\frac{5\lambda^2}{4}, +\infty\right)$	$0 < t^{\pi}_{HS\sim PP} < t^{sw}_{HS\sim PP} < \underline{t_{MX}} < \overline{t_{2H}} < \overline{t_{HS}}$

Table 2.14: Order of Critical Values (Proposition 2.9)

□

2.9.2 Solutions of First-Best Socially Optimal Network Structure

First, the solution of the hub-and-spoke network is:

$$F^{1*}_{HS} = F^{2*}_{HS} = \lambda(3v - t)(3\Phi_{HS})^{-1},$$
$$Q^{1*}_{HS} = Q^{2*}_{HS} = \left(12\psi v - \lambda^2 t\right)(3\Phi_{HS})^{-1},$$
$$Q^{3*}_{HS} = 2\left[6\psi(v-t) + \lambda^2 t\right](3\Phi_{HS})^{-1},$$

in which $\Phi_{HS} = 4\psi - \lambda^2$.

Second, the solution of the point-to-point network is:

$$F_{PP}^{i*} = \lambda v \Phi_{PP}^{-1},$$

$$Q_{PP}^{i*} = 8\psi v \Phi_{PP}^{-1},$$

in which $i \in \{1, 2, 3\}$ and $\Phi_{PP} = 8\psi - \lambda^2$.

Third, the solution of the mixed network is:

$$F_{MX}^{1*} = F_{MX}^{2*} = \lambda \left[6\psi (v - t) - \lambda^2 (3v - t)\right] \Phi_{MX}^{-1},$$

$$F_{MX}^{3*} = \lambda \left[2\psi (-v + 5t) - \lambda^2 (3v + t)\right] \Phi_{MX}^{-1},$$

$$Q_{MX}^{1*} = Q_{MX}^{2*} = \left[24\psi^2 v - 2\lambda^2 \psi (10v + 3t) + \lambda^4 t\right] \Phi_{MX}^{-1},$$

$$Q_{MX}^{30*} = 8\psi \left[2\psi (v + t) - 3\lambda^2 v\right] \Phi_{MX}^{-1},$$

$$Q_{MX}^{31*} = 2 \left(2\psi - \lambda^2\right) \left[4\psi (v - 2t) + \lambda^2 t\right] \Phi_{MX}^{-1},$$

in which $\Phi_{MX} = 24\psi^2 - 26\lambda^2\psi + 3\lambda^4$.

Finally, the solution of the 2-hub network is:

$$F_{2H}^{1*} = \lambda \left[6\psi (3v - 4t) - \lambda^2 (5v - 4t)\right] \Phi_{2H}^{-1},$$

$$F_{2H}^{2*} = F_{2H}^{3*} = \left[2\lambda\psi (v + 4t) - 5\lambda^3 v\right] \Phi_{2H}^{-1},$$

$$Q_{2H}^{1*} = 4 \left[12\psi^2 v - 6\lambda^2 \psi (v + t) + \lambda^4 t\right] \Phi_{2H}^{-1},$$

$$Q_{2H}^{20*} = Q_{2H}^{30*} = 2 \left[16\psi^2 (v + t) - 2\lambda^2 \psi (8v + 3t) + \lambda^4 t\right] \Phi_{2H}^{-1},$$

$$Q_{2H}^{21*} = Q_{2H}^{31*} = 4 \left(2\psi - \lambda^2\right) \left[4\psi (v - 2t) + \lambda^2 t\right] \Phi_{2H}^{-1},$$

in which $\Phi_{2H} = 48\psi^2 - 42\lambda^2\psi + 5\lambda^4$.

2.10 Appendix D: Extension and Critical Values

The solution of the 3-hub network is:

$$f_{3H}^{i*} = \lambda\left(2v - t\right)\left(2\phi_{3H}\right)^{-1},$$

$$q_{3H}^{i0*} = \left[8\psi\left(v + t\right) - \lambda^2 t\right]\left(2\phi_{3H}\right)^{-1},$$

$$q_{3H}^{i1*} = \left[8\psi\left(v - 2t\right) + \lambda^2 t\right]\left(2\phi_{3H}\right)^{-1},$$

in which $i \in \{1, 2, 3\}$ and $\phi_{3H} = 12\psi - \lambda^2$.

Moreover, Tables 2.15 through 2.17 give the critical values in this chapter.

$t_{MX}^{1} = \frac{16\psi v\left(48\psi^2-48\lambda^2\psi+13\lambda^4\right)}{384\psi^3-336\lambda^2\psi^2+112\lambda^4\psi-5\lambda^6}$
$t_{MX}^{2} = \frac{4\psi v\left(384\psi^3-336\lambda^2\psi^2+112\lambda^4\psi-5\lambda^6\right)}{3072\psi^4-1920\lambda^2\psi^3+496\lambda^4\psi^2-40\lambda^6\psi+\lambda^8}$
$t_{MX}^{3} = \frac{\left(384\psi^3+1072\lambda^2\psi^2-192\lambda^4\psi-9\lambda^6\right)v}{1920\psi^3+752\lambda^2\psi^2-128\lambda^4\psi+3\lambda^6}$
$t_{MX}^{4} = \frac{\left(80\psi^2+24\lambda^2\psi-9\lambda^4\right)v}{8\psi(26\psi-3\lambda^2)}$
$t_{MX}^{5} = \frac{3v\left(16\psi^2+24\lambda^2\psi-7\lambda^4\right)}{240\psi^2-24\lambda^2\psi-\lambda^4}$
$t_{MX}^{6} = \frac{\left(7680\psi^4-576\lambda^2\psi^3+1584\lambda^4\psi^2-412\lambda^6\psi+15\lambda^8+\phi_{MX}\sqrt{\Delta_{MX}^6}\right)v}{4(8448\psi^4-2592\lambda^2\psi^3+360\lambda^4\psi^2-32\lambda^6\psi+\lambda^8)}$
$\Delta_{MX}^{6} = -\left(16128\psi^4 - 22272\lambda^2\psi^3 + 4128\lambda^4\psi^2 - 48\lambda^6\psi - 25\lambda^8\right)$
$t_{MX}^{7} = -\frac{\left[64\psi^2+12\lambda^2\psi+3\lambda^4-\sqrt{3(3072\psi^4+5376\lambda^2\psi^3-1552\lambda^4\psi^2+168\lambda^6\psi+3\lambda^8)}\right]v}{8\psi(20\psi-3\lambda^2)}$
$t_{MX}^{8} = -\frac{\left(64\psi^2-32\lambda^2\psi+3\lambda^4-\sqrt{9216\psi^4+3072\lambda^2\psi^3-2240\lambda^4\psi^2+96\lambda^6\psi+9\lambda^8}\right)v}{16\psi(10\psi-\lambda^2)}$
$t_{2H}^{1} = \frac{8\psi v\left(768\psi^3-528\lambda^2\psi^2+112\lambda^4\psi-3\lambda^6\right)}{12288\psi^4-7680\lambda^2\psi^3+1648\lambda^4\psi^2-120\lambda^6\psi+3\lambda^8}$
$t_{2H}^{2} = -\frac{\left(768\psi^3-2544\lambda^2\psi^2+360\lambda^4\psi+25\lambda^6\right)v}{48\psi(64\psi^2+28\lambda^2\psi-5\lambda^4)}$
$t_{2H}^{3} = \frac{4v\left(24\psi^2+20\lambda^2\psi-5\lambda^4\right)}{384\psi^2-32\lambda^2\psi-3\lambda^4}$
$t_{2H}^{4} = -\frac{\left(48\psi^2-120\lambda^2\psi+25\lambda^4\right)v}{192\psi^2-5\lambda^4}$
$t_{2H}^{5} = \frac{\left(1792\psi^3+1616\lambda^2\psi^2-128\lambda^4\psi-5\lambda^6-\sqrt{\Delta_{2H}^5}\right)v}{32\psi(64\psi^2+28\lambda^2\psi-3\lambda^4)}$
$\Delta_{2H}^{5} = 5308416\psi^6 - 368640\lambda^2\psi^5 - 59136\lambda^4\psi^4 + 330240\lambda^6\psi^3 - 45856\lambda^8\psi^2 + 1280\lambda^{10}\psi + 25\lambda^{12}$
$t_{2H}^{6} = \frac{\left(224\psi^2+80\lambda^2\psi-17\lambda^4-\sqrt{82944\psi^4-46080\lambda^2\psi^3+13632\lambda^4\psi^2-1440\lambda^6\psi+49\lambda^8}\right)v}{4(64\psi^2-\lambda^4)}$

Table 2.15: Critical Values of Comparative Statics

$t^{\pi}_{HS\sim PP} = \frac{6v}{12\psi-\lambda^2}\left[2\psi - \frac{\left(8\psi-\lambda^2\right)\sqrt{(16\psi-\lambda^2)\psi}}{16\psi-\lambda^2}\right]$
$\xi_1 = \sqrt{3\left(8\zeta_1+179\right)} + \sqrt{6\left(-4\zeta_1+179+7605\sqrt{\frac{3}{8\zeta_1+179}}\right)}$
$\zeta_1 = \sqrt[3]{190657-135\sqrt{1966530}} + \sqrt[3]{190657+135\sqrt{1966530}}$
$t^{1}_{HS\sim PP} = \frac{12\lambda^2\psi v}{(16\psi-\lambda^2)(12\psi-\lambda^2)}$
$t^{1}_{HS\sim MX} = \frac{\left(40\psi+\lambda^2-\sqrt{640\psi^2-112\lambda^2\psi+13\lambda^4}\right)v}{2(20\psi-\lambda^2)}$
$t^{1}_{HS\sim 2H} = \frac{24\psi v\left(12\psi+\lambda^2\right)}{960\psi^2-156\lambda^2\psi+7\lambda^4}$
$t^{2}_{HS\sim 2H} = \frac{3\psi v\left(64\psi^2-4\lambda^2\psi-3\lambda^4\right)}{960\psi^3-420\lambda^2\psi^2+51\lambda^4\psi-2\lambda^6}$
$t^{3}_{HS\sim 2H} = \frac{60\lambda^2\psi v\left(4\psi-\lambda^2\right)}{2304\psi^3-1008\lambda^2\psi^2+120\lambda^4\psi-5\lambda^6}$
$t^{4}_{HS\sim 2H} = \frac{\left[3\left(36864\psi^4-6656\lambda^2\psi^3-144\lambda^4\psi^2-120\lambda^6\psi+15\lambda^8\right)-\phi_{HS}\phi_{2H}\sqrt{6(1200\psi^2+136\lambda^2\psi-21\lambda^4)}\right]v}{258048\psi^4-90624\lambda^2\psi^3+10992\lambda^4\psi^2-696\lambda^6\psi+23\lambda^8}$
$t^{5}_{HS\sim 2H} = \frac{\left(1184\psi^2+100\lambda^2\psi-7\lambda^4-\sqrt{418816\psi^4+40192\lambda^2\psi^3+5712\lambda^4\psi^2-1400\lambda^6\psi+49\lambda^8}\right)v}{64\psi(20\psi-\lambda^2)}$
$t^{1}_{PP\sim 2H} = \frac{16\lambda^2\psi v\left(8\psi-\lambda^2\right)}{3072\psi^3-1280\lambda^2\psi^2+116\lambda^4\psi-3\lambda^6}$
$t^{1}_{MX\sim 2H} = \frac{2v\left(64512\psi^5-131328\lambda^2\psi^4+82176\lambda^4\psi^3-21920\lambda^6\psi^2+2692\lambda^8\psi-105\lambda^{10}\right)}{36864\psi^5-140544\lambda^2\psi^4+83712\lambda^4\psi^3-16256\lambda^6\psi^2+1456\lambda^8\psi-47\lambda^{10}}$
$t^{2}_{MX\sim 2H} = \frac{\left[8448\psi^4+17664\lambda^2\psi^3-12928\lambda^4\psi^2+2176\lambda^6\psi-141\lambda^8-\sqrt{3\Delta^{21}_{MX\sim 2H}\Delta^{22}_{MX\sim 2H}}\right]v}{8(4\psi-\lambda^2)^2(240\psi^2-24\lambda^2\psi-\lambda^4)}$
$\Delta^{21}_{MX\sim 2H} = 2304\psi^4 - 3840\lambda^2\psi^3 + 3072\lambda^4\psi^2 - 640\lambda^6\psi + 51\lambda^8$
$\Delta^{22}_{MX\sim 2H} = 99072\psi^4 - 97536\lambda^2\psi^3 + 29952\lambda^4\psi^2 - 2944\lambda^6\psi + 97\lambda^8$

Table 2.16: Critical Values of Network Structure

$t^{sw}_{HS\sim PP} = \frac{6v}{288\psi^2-56\lambda^2\psi+3\lambda^4}\left[4\psi\left(12\psi-\lambda^2\right) - \frac{\left(8\psi-\lambda^2\right)\sqrt{2\psi(4608\psi^3-1344\lambda^2\psi^2+124\lambda^4\psi-3\lambda^6)}}{16\psi-\lambda^2}\right]$
$t^{sw}_{MX\sim 2H} = \frac{8\psi v\left(2654208\psi^6-4681728\lambda^2\psi^5+3071232\lambda^4\psi^4-939264\lambda^6\psi^3+133664\lambda^8\psi^2-7168\lambda^{10}\psi+105\lambda^{12}\right)}{42467328\psi^7-79331328\lambda^2\psi^6+57028608\lambda^4\psi^5-20322048\lambda^6\psi^4+3813632\lambda^8\psi^3-366688\lambda^{10}\psi^2+17216\lambda^{12}\psi-315\lambda^{14}}$
ξ_2 (≈ 1.51) is a solution of $63x^5 - 2172x^4 + 26192x^3 - 131136x^2 + 291840x - 221184 = 0$
ξ_3 (≈ 1.59) is a solution of $15x^4 - 532x^3 + 6720x^2 - 38528x + 92160$ $-96\left(-x+8\right)\sqrt{2\left(-3x^3+124x^2-1344x+4608\right)} = 0$
$T^{SW}_{HS\sim PP} = \frac{3v}{6\psi-\lambda^2}\left[2\psi - \frac{\left(4\psi-\lambda^2\right)\sqrt{2(8\psi-\lambda^2)\psi}}{8\psi-\lambda^2}\right]$

Table 2.17: Critical Values of Welfare Analysis

Chapter 3

Agricultural Land Marketization, Inverse Relationship and Land Productivity: Empirical Evidence from China

3.1 Introduction

For developing countries, especially those in transition from agricultural to non-agricultural economy, on the one hand, the transition of economy reduces the amount of agricultural labor significantly and thus decreases the utilization rate of agricultural land. On the other hand, the transition increases the demand of agricultural products of urban areas and thus further aggravates the balance between supply and demand. Given the reality that the domestic farmland cannot be enlarged easily, governments in many countries try to improve the output per unit of land, or land productivity, to increase the supply of agricultural products.

According to economic theory, the agricultural land marketization can improve the land allocation efficiency. After the agricultural land marketization, less efficient agricultural producers can rent out or sell some of their land at a price higher than their marginal production, while more efficient producers can rent in or buy some land at a price lower than their marginal production. Finally, the agricultural land will be allocated more efficiently through market mechanism (see Yao, 2000; Benjamin and Brandt, 2002; Carter and Yao, 2002; Deininger and Jin, 2005, 2008; Deininger et al., 2008a; Deininger et al., 2008b; Jin and Deininger, 2009; Barrett et al., 2010). Then, if the agricultural land marketization can improve the land allocation efficiency, can it also improve the average output per unit of land, or average land productivity? The conventional answer is affirmative because the higher land allocation efficiency implies the higher average land productivity (see, for example, Restuccia and Santaeulalia-Llopis, 2017[1]). However, if we consider the inverse relationship between farm size and land productivity, the answer is uncertain.

In many developing countries, there exists an inverse relationship between farm size and land productivity.[2] That is, compared to rural households with a large farm size, those with a small farm size have higher land productivity. This relationship has been found in the countries of Asia (see Sen, 1962; Lau and Yotopoulos, 1971; Bardhan, 1973; Rao and Chotigeat, 1981; Carter, 1984; Newell et al., 1997; Heltberg, 1998; Lamb, 2003), Africa (see Collier, 1983; Barrett, 1996; Byiringiro and Reardon, 1996; Kimhi, 2006; Carletto et al., 2013; Larson et al., 2014; Ali and Deininger, 2015), Europe (see Chayanov, 1926; Alvarez and Arias, 2004), and Latin America (see Berry and Cline, 1979; Cornia, 1985).[3]

After the agricultural land marketization, on the one hand, the previously unused land can be used again, and thus the total operational farm size may increase. Given the amount of rural households, the average operational farm size may also increase. On

[1]Restuccia and Santaeulalia-Llopis (2017) use household-level data from Malawi and find that a reallocation of production factors to their efficient use will result in higher average total factor productivity (TFP) of farmers, in which the farm TFP and the output per unit of land are found to be strongly positively correlated across farms because the allocation of land is not related to productivity so many productive farmers are constrained by size.

[2]The reasons explaining the existence of inverse relationship include, among others, land market imperfections (see Heltberg, 1998; Lamb, 2003), labor market imperfections (see Sen, 1966; Rosenzweig and Wolpin, 1985; Frisvold, 1994), credit market imperfections (see Feder, 1985; Eswaran and Kotwal, 1986; Carter, 1988), and risk (see Wiens, 1977; Rosenzweig and Binswanger, 1993; Kevane, 1996).

[3]Some studies show that, in USA and Japan, farm size is positively correlated with land productivity (see Sumner, 2014; Kawasaki, 2010).

the other hand, rural households can obtain monetary incomes from land transactions, which can provide a basic guarantee for their migrations to urban areas. In this way, the amount of rural households may decrease, and the average operational farm size may then increase. Under the inverse relationship, the increase of average operational farm size will reduce the average land productivity.

To summarize, the agricultural land marketization affects the average land productivity not only through improving the land allocation efficiency but also through increasing the average operational farm size.[4] The improvement of land allocation efficiency has a positive effect on average land productivity. However, when there exists an inverse relationship between farm size and land productivity, the increase of average operational farm size has a negative effect on average land productivity. Therefore, the agricultural land marketization does not necessarily improve the average land productivity. Only when the positive effect of the higher land allocation efficiency dominates the negative effect of the larger average operational farm size, the marketization will finally improve the average land productivity.

In this chapter, we use the year 2008 as the indicator of the agricultural land marketization in China to test empirically the effect of the marketization on average land productivity. The empirical framework is the one for the study of inverse relationship (see Binswanger et al., 1995; Assunção and Braido, 2007; Barrett et al., 2010; Carletto et al., 2013) and the data we use is from the China Health and Nutrition Survey (CHNS) database[5]. Finally, we find that: first, there exists an inverse relationship between farm size and land productivity in China; second, the agricultural land marketization in China improves the land allocation efficiency and increases the average operational farm size; third, the higher land allocation efficiency improves the average land productivity by 29.1% and the larger average operational farm size reduces the average land productivity by 9.2%, implying that the agricultural land marketization in China finally improves the

[4]The agricultural land marketization affects the average land productivity also through, for example, influencing indirectly the amount of labor input and intermediate inputs. However, in this chapter, we focus our discussions on the direct effects of the marketization, that is, improving the land allocation efficiency and increasing the average operational farm size.

[5]http://www.cpc.unc.edu/projects/china

average land productivity by 19.9%.

The rest of the chapter is organized as follows. Section 3.2 introduces the agricultural land marketization in China and proposes four hypotheses to be tested in the empirical analysis. Section 3.3 describes the empirical framework. Section 3.4 explains the data and the relevant descriptive statistics. Section 3.5 shows empirical results. Section 3.6 concludes.

3.2 Agricultural Land Marketization in China

In the collective periods beginning in the 1950s, the Chinese government prohibited transactions in land, labor and rental markets (see Lin, 1995). Since the rural reform in 1978, a household-based farming system, that is, the household responsibility system, was executed, and thus the prohibition on the transactions in labor was abandoned. However, the transactions in land were still prohibited. The Constitution of the People's Republic of China (1982) stipulates that no organization or individual may appropriate, buy, sell or lease land or otherwise engage in the transfer of land by unlawful means.[6]

With the rapid development of urbanization in the mid of 1980s, in order to satisfy the demand of urban development, the Amendment to the Constitution of the People's Republic of China (1988) stipulates that the right to the use of land may be transferred according to law and thus provides a legal basis for the transfer of land.[7] However, this amendment to the constitution emphasizes the transfer of industrial land and city construction land and does not provide detailed legal explanations for the transfer of agricultural land between rural households. In this period, more and more farmers move to urban areas in order to pursue a higher quality of life and income, making some land in rural areas unused (see Xu and Zhang, 1993; Wu, 1993; Ma, 2008).

In order to reduce the waste of agricultural resources and improve the utilization efficiency of agricultural land, the Law of the People's Republic of China on Land Contract in Rural Areas (2003) stipulates that the right to land contractual management obtained

[6] http://www.npc.gov.cn/englishnpc/Constitution/node_2830.htm
[7] http://www.npc.gov.cn/englishnpc/Constitution/node_2829.htm

through household contract may, according to law, be circulated by subcontracting, leasing, exchanging, transferring or other means.[8] For creating a better environment for the transfer of agricultural land, the Measures for the Administration of Circulation of Rural Land Contracted Management Right was carried out by the Ministry of Agriculture of China in 2005.[9] In 2008, the Decision of the Central Committee of the Communist Party of China on Several Big Issues on Promoting the Reform and Development of Rural Areas decided to strengthen the development of agricultural land transfer market and improve the transfer rate.[10]

Based on the above reform process in China, we use the year 2008 as the indicator of the agricultural land marketization. We first propose four hypotheses to be tested in the empirical analysis.

Hypothesis 1: *There exists an inverse relationship between farm size and land productivity in China.*

Hypothesis 2: *The agricultural land marketization in China improves the land allocation efficiency and thus the average land productivity.*

Hypothesis 3: *The agricultural land marketization in China increases the average operational farm size; given the existence of the inverse relationship, the marketization reduces the average land productivity.*

Hypothesis 4: *The positive effect of the higher land allocation efficiency dominates the negative effect of the larger average operational farm size, and thus the agricultural land marketization in China finally improves the average land productivity.*

The relationship between these four hypotheses is shown in Figure 3.1. Because the agricultural land market in China is imperfect, there is necessity to implement an agricultural land marketization reform and discuss whether or not the marketization can improve the average land productivity. Hypothesis 1 is another premise. Because of the existence of inverse relationship, the effect of the agricultural land marketization on average land productivity becomes uncertain. Hypotheses 2 and 3 are the two channels

[8] http://www.npc.gov.cn/englishnpc/Law/2007-12/06/content_1382125.htm

[9] http://www.moa.gov.cn/zwllm/tzgg/bl/200501/t20050126_311817.htm

[10] http://www.lawinfochina.com/display.aspx?lib=law&id=7542&CGid=

to test the effect. Hypothesis 4 is the net effect of these two channels and the central problem of this chapter.

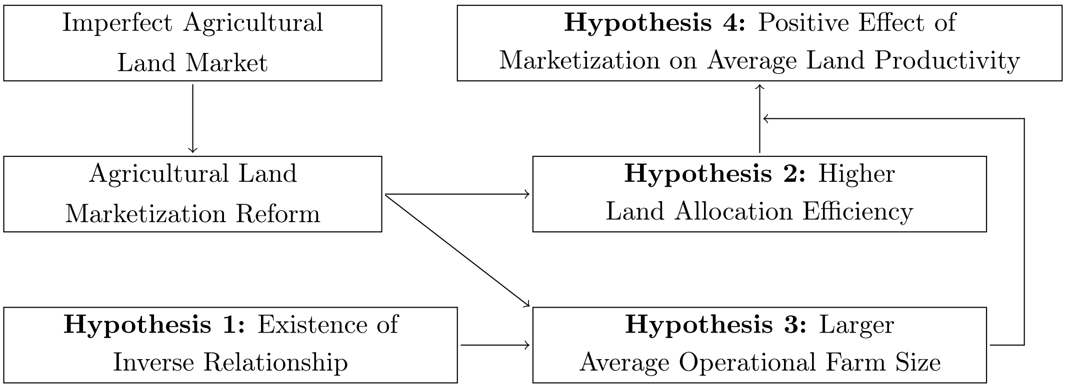

Figure 3.1: Relationship Between Different Hypotheses

3.3 Empirical Framework

For rural household i, consider a Cobb-Douglas agricultural production function:

$$Y_i = T A_i^{\alpha_A} L_i^{\alpha_L} R_i^{\alpha_R} \exp \varepsilon_i, \tag{3.1}$$

in which Y_i is the output; T is the technological level; A_i is the operational farm size, that is, the cropped area; L_i is the labor input; R_i is the intermediate inputs; and ε_i is an error term which accounts for unobserved and idiosyncratic determinants of the total output. α_A, α_L and α_R represent the output elasticities of operational farm size, labor input and intermediate inputs, respectively. In order to be consistent with our data, we need to represent the total output and the intermediate inputs in monetary units. Multiplying Y_i and R_i by their respective prices (namely p_Y and p_R). (3.1) becomes:

$$y_i = t A_i^{\alpha_A} L_i^{\alpha_L} r_i^{\alpha_R} \exp \varepsilon_i, \tag{3.2}$$

in which $y_i = p_Y Y$ is the value of output, $t = \frac{p_Y T}{(p_R)^{\alpha_R}}$ is the price-adjusted technological level and $r_i = p_R R$ is the value of intermediate inputs.

Taking the natural logarithm of both sides of (3.2), we can obtain:

$$\ln y_i = \ln t + \alpha_A \ln A_i + \alpha_L \ln L_i + \alpha_R \ln r_i + \varepsilon_i, \tag{3.3}$$

which is equivalent to:

$$\ln \frac{y_i}{A_i} = \ln t + (\alpha_A - 1) \ln A_i + \alpha_L \ln L_i + \alpha_R \ln r_i + \varepsilon_i. \tag{3.4}$$

Based on (3.4) and the framework for the study of inverse relationship (see Binswanger et al., 1995; Assunção and Braido, 2007; Barrett et al., 2010; Carletto et al., 2013), the empirical model is:

$$\ln \frac{y_i}{A_i} = Cons + \beta \ln A_i + \gamma_k CV_{ki} + \varepsilon_i, \tag{3.5}$$

in which $\frac{y_i}{A_i}$ is the output value per unit of land, that is, land productivity, of the household i; $Cons$ is the constant, implying the price-adjusted technological level of agricultural production; CV_{ki} are control variables, including other agricultural production factors, household characteristics and land quality differences. The other agricultural production factors include labor input ($Labor$) and intermediate inputs ($Raising$), both of which have a positive effect on the land productivity. The household characteristics include the household head's age (Age), gender ($Gend$), education level (Edu), marriage situation ($Marriage$), household size ($Hsize$), and dependency ratio ($Dratio$). An older household head implies richer agricultural production experience, which is conducive to higher land productivity; a divorced, single or female household head may bring lower land productivity; a household head with a higher education level implies richer knowledge of agricultural production, which is conducive to improving land productivity by using high and new technology; a household with a large size inclines to use internal labor, instead of hiring external labor; and a household with a high dependency ratio needs to use more labor input to feed dependants. Moreover, we introduce the village fixed effect to control the land quality differences (see Bhalla and Roy, 1988; Benjamin, 1995).

We will run OLS regressions for the full sample and all subsamples. The coefficient of interest is β, representing the relationship between farm size and land productivity. Then, we can use the following method to test Hypothesis 1.

Test of Hypothesis 1: When β is negative and statistically significant, there exists an inverse relationship between farm size and land productivity; when β is non-negative and statistically significant, the inverse relationship does not exist.

Moreover, when β is negative, the greater β is, the less significant the inverse relationship between farm size and land productivity will be.

As agricultural land transactions in China were prohibited or stayed at a low level during a long period, land market imperfections are one of the most important reasons to explain the inverse relationship in China (if it exists). Thus, the agricultural land marketization is conducive to improving the land allocation efficiency and then reducing the degree of the inverse relationship. In this way, the change of β can reflect the direction and degree of the variation of land allocation efficiency to a large extent. Denoting the periods before and after the agricultural land marketization in 2008 by the subscripts "before" and "after", we can use the following method to test Hypothesis 2.

Test of Hypothesis 2: When $|\beta_{after}| - |\beta_{before}|$ is negative and statistically significant, the land allocation efficiency improves after the agricultural land marketization, implying the higher average land productivity.

Moreover, when $|\beta_{after}| - |\beta_{before}|$ is negative and statistically significant, the less $|\beta_{after}| - |\beta_{before}|$ is, the higher the degree of the improvement of land allocation efficiency will be.

Denoting the average operational farm sizes before and after the marketization by $\overline{A_{i,before}}$ and $\overline{A_{i,after}}$, respectively, the test of Hypothesis 3 is rather straightforward.

Test of Hypothesis 3: When $\overline{A_{i,after}} > \overline{A_{i,before}}$, the average operational farm size increases after the agricultural land marketization, implying the lower average land productivity given that Hypothesis 1 is verified.

In order to test Hypothesis 4, we use the method of factor decomposition. Specifically, we first estimate (3.5) by using the samples before and after the marketization. The estimation results then give:

$$\ln \frac{y_{i,before}}{A_{i,before}} = Cons_{before} + \beta_{before} \ln A_{i,before} + \gamma_{k,before} CV_{ki,before}, \tag{3.6}$$

$$\ln \frac{y_{i,after}}{A_{i,after}} = Cons_{after} + \beta_{after} \ln A_{i,after} + \gamma_{k,after} CV_{ki,after}, \tag{3.7}$$

in which $\{Cons_{before}, \beta_{before}, \gamma_{k,before}\}$ and $\{Cons_{after}, \beta_{after}, \gamma_{k,after}\}$ are the estimation results of coefficients before and after the marketization, respectively.

Next, (3.7) minus (3.6) equals:

$$\begin{aligned} &\ln \frac{y_{i,after}}{A_{i,after}} - \ln \frac{y_{i,before}}{A_{i,before}} \\ &\qquad = \left[\beta_{after} \left(\ln A_{i,after} - \ln A_{i,before}\right) + \ln A_{i,before} \left(\beta_{after} - \beta_{before}\right)\right] \\ &\qquad + \left(\gamma_{k,after} CV_{ki,after} - \gamma_{k,before} CV_{ki,before}\right) + \left(Cons_{after} - Cons_{before}\right). \end{aligned} \tag{3.8}$$

Plugging the sample means before and after the marketization into (3.8), we can obtain:

$$\begin{aligned} &\ln \overline{\left(\frac{y_{i,after}}{A_{i,after}}\right)} - \ln \overline{\left(\frac{y_{i,before}}{A_{i,before}}\right)} \\ &\qquad = \left[\beta_{after} \left(\ln \overline{A_{i,after}} - \ln \overline{A_{i,before}}\right) + \ln \overline{A_{i,before}} \left(\beta_{after} - \beta_{before}\right)\right] \\ &\qquad + \left(\gamma_{k,after} \overline{CV_{ki,after}} - \gamma_{k,before} \overline{CV_{ki,before}}\right) + \left(Cons_{after} - Cons_{before}\right), \end{aligned} \tag{3.9}$$

in which $\overline{\left(\frac{y_{i,before}}{A_{i,before}}\right)}$ and $\overline{\left(\frac{y_{i,after}}{A_{i,after}}\right)}$ are the average land productivity before and after the marketization, respectively; and $\overline{CV_{ki,before}}$ and $\overline{CV_{ki,after}}$ are the sample means of control variables before and after the marketization, respectively.

The left-hand side of (3.9), $\ln \overline{\left(\frac{y_{i,after}}{A_{i,after}}\right)} - \ln \overline{\left(\frac{y_{i,before}}{A_{i,before}}\right)}$, is the growth rate of the average land productivity before and after the agricultural land marketization. The right-hand side of (3.9) is the main factors of the growth. Specifically, the term $\beta_{after} \left(\ln \overline{A_{i,after}} - \ln \overline{A_{i,before}}\right)$ shows the effect of the change of average operational farm size on average land productivity. When the average operational farm size increases, that is, $\ln \overline{A_{i,after}} - \ln \overline{A_{i,before}} > 0$, and the inverse relationship exists, that is, $\beta_{after} < 0$, the effect of the change of average operational farm size will be negative. Intuitively, given the degree of the inverse relationship after the marketization, the term $\beta_{after}\left(\ln\overline{A_{i,after}} - \ln\overline{A_{i,before}}\right)$ reflects how the change of average operational land size, due to the marketization, affects the average land productivity. The term $\ln\overline{A_{i,before}}\left(\beta_{after} - \beta_{before}\right)$ shows the effect of the change of land allocation efficiency on average land productivity. Because the average operational farm size is positive, that is, $\ln\overline{A_{i,before}} > 0$, when the land allocation efficiency improves, that is, $\beta_{after} - \beta_{before} > 0$, the effect of the change of land allocation efficiency will be positive. Intuitively, given the average operational land size before the marketization, the term $\ln\overline{A_{i,before}}\left(\beta_{after} - \beta_{before}\right)$ reflects how the change of land allocation efficiency, due to the marketization, affects the average land productivity. Then, the term $\beta_{after}\left(\ln\overline{A_{i,after}} - \ln\overline{A_{i,before}}\right) + \ln\overline{A_{i,before}}\left(\beta_{after} - \beta_{before}\right)$ is the essential problem to be tested in this chapter. That is to say, how the agricultural land marketization affects the average land productivity depends on the net effect of the changes of average operational farm size and land allocation efficiency. Moreover, terms $\gamma_{k,after}\overline{CV_{ki,after}} - \gamma_{k,before}\overline{CV_{ki,before}}$ and $Cons_{after} - Cons_{before}$ show the effects of control variables and price-adjusted technological level, respectively.

Then, we can test Hypothesis 4 as follows.

Test of Hypothesis 4: When $\beta_{after}\left(\ln\overline{A_{i,after}} - \ln\overline{A_{i,before}}\right) + \ln\overline{A_{i,before}}\left(\beta_{after} - \beta_{before}\right) > 0$, the agricultural land marketization improves the average land productivity.

3.4 Data and Descriptive Statistics

The data we use is from the China Health and Nutrition Survey (CHNS) database, which is created by the Carolina Population Center at the University of North Carolina at Chapel Hill and the National Institute for Nutrition and Health (NINH, former National Institute of Nutrition and Food Safety) at the Chinese Center for Disease Control and Prevention (CCDC). The CHNS database allows us to conduct a panel data analysis, in which, however, the sample size is too limited. Therefore, we pool observations in different years and conduct a cross-sectional data analysis. Specifically, we use a sample of 5313 observations in five years. The sample sizes in the years 2000, 2004, 2006, 2009, and 2011 are 512, 1088, 1227, 1251, and 1235, respectively. The data description is in Table 3.1.

Moreover, Tables 3.2 through 3.9 provide the summary statistics of different samples. According to Table 3.2, for the sample 2000-2011, the average household size ($Hsize$) is 2.776; the average dependency ratio ($Dratio$) is 0.282; and the household head's marriage rate ($Marriage$) is 99.1%, implying that a typical household consists of a married couple and a dependant. For the household head, male ($Gend$) accounts for 91%; the average age (Age) is 52.289 years old; and the average education level (Edu) is 6.532 years, implying that household heads are mainly male and relatively old, with relatively low education level. The average operational farm size (A) is 6.524 mu[11], implying a small scale of agricultural production. According to Tables 3.3 and 3.4, comparing the sample means before and after the agricultural land marketization (samples 2000-2006 and 2009-2011, respectively), we can find that the average land productivity ($\frac{y}{A}$) after the marketization is 1643.396 yuan/mu[12], which is 204.366 yuan/mu higher than the one before the marketization. Moreover, there is no significant change in labor input ($Labor$), while there is a significant growth of intermediate inputs ($Raising$). In particular, after the marketization, the average operational farm size is 7.162 mu, which is 1.198 mu higher than the one before the marketization.

3.5 Empirical Results

3.5.1 Existence of Inverse Relationship

According to the estimation results of full sample (sample 2000-2011) and samples 2000, 2004, 2006, 2009, and 2011 in Table 3.10, the operational farm size is negatively correlated with the land productivity, statistically significant at 1% level, implying the existence of an inverse relationship between farm size and land productivity in China, verifying Hypothesis 1.[13]

[11] One mu equals 1/15th of a hectare.

[12] At the time of surveys, approximately 6.5-8.2 Chinese yuan can be exchanged for each 1 US dollar.

[13] Some studies also find the existence of an inverse relationship in China (see Benjamin and Brandt, 2002; Chen et al., 2011; Li et al., 2013).

Specifically, for the full sample, the coefficient of operational farm size ($\ln A$) is -0.587, statistically significant at 1% level, implying the inverse relationship. The coefficients of labor input ($\ln Labor$) and intermediate inputs ($\ln Raising$) are 0.112 and 0.360, respectively, both statistically significant at 1% level, implying the positive effects of these two production factors on land productivity. For the control variables of household characteristics, the coefficients of the household head's age (Age) and household size ($Hsize$) are 0.003 and -0.033, respectively, both statistically significant at 1% level. However, the coefficients of the household head's gender ($Gend$), education level (Edu), marriage situation ($Marriage$), and dependency ratio ($Dratio$) are not statistically significant at 10% level. Therefore, these household characteristics do not significantly affect the land productivity.

Furthermore, for the samples 2000, 2004, 2006, 2009, and 2011, the coefficients of operational farm size are -0.549, -0.689, -0.671, -0.556, and -0.453, respectively, all statistically significant at 1% level, implying the existence of the inverse relationship.

As we have seen, the average operational farm size increases after the agricultural land marketization. Therefore, given the existence of the inverse relationship, the larger average operational farm size reduces the average land productivity, verifying then Hypothesis 3.

3.5.2 Higher Land Allocation Efficiency

For the samples 2000, 2004, 2006, 2009, and 2011, we depict the coefficients of operational farm size ($\ln A$) in Figure 3.2. We can find then that the coefficient of operational farm size, or the land allocation efficiency, decreases a lot from 2000 to 2006 and stays at a low level around 2004 and 2006. However, after the agricultural land marketization in China in 2008, there is a significant improvement of land allocation efficiency, verifying Hypothesis 2.

Moreover, the estimation results of samples before and after the agricultural land marketization in Table 3.11 also show an improvement of land allocation efficiency, verifying again Hypothesis 2. Specifically, the coefficients of operational farm size of samples

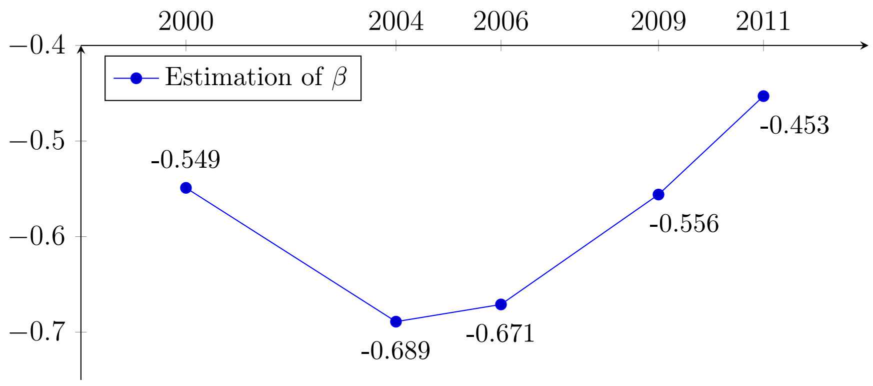

Figure 3.2: Coefficients of Operational Farm Size ($\ln A$)
Source: Authors' calculations based on CHNS 2000, 2004, 2006, 2009, and 2011

2000-2006 and 2009-2011 are -0.666 and -0.503, respectively, both statistically significant at 1% level. The difference between these two coefficients is 0.163, which is also statistically significant at 1% level, implying that the marketization improves the land allocation efficiency.

3.5.3 Positive Effect of Marketization on Productivity

In order to evaluate the net effect of the agricultural land marketization on average land productivity according to (3.9), we use the sample means in Tables 3.3 and 3.4 and the estimation results in Table 3.11. According to the calculation results of factor decomposition in Table 3.12, the marketization improves the land allocation efficiency by 0.163 units and thus improves the average land productivity by 29.1%. The marketization also increases the average operational farm size by 18.3% and thus reduces the average land productivity by 9.2%. Therefore, the net effect is that the marketization improves the average land productivity by 19.9%. These results thus verify Hypothesis 4, that is, the positive effect of the higher land allocation efficiency dominates the negative effect of the larger average operational farm size, and thus the agricultural land marketization in China finally improves the average land productivity.

In fact, the reason why the agricultural land marketization in China can affect the average land productivity negatively is that the inverse relationship between farm size and land productivity still exists after the marketization. On the one hand, in the short run, the agricultural land marketization may not eliminate the land market imperfections thoroughly. On the other hand, besides land market imperfections, there exist also other factors resulting in the inverse relationship, for example, credit market imperfections, that the agricultural land marketization cannot eliminate. Therefore, as long as the agricultural land marketization reform continues to deepen and other relevant reforms are carried out, the inverse relationship would probably disappear, and then the agricultural land marketization would not affect the average land productivity negatively.

3.6 Conclusion

Based on the perspective of the inverse relationship between farm size and land productivity, this chapter proposes that, besides improving the land allocation efficiency, the agricultural land marketization can also affect the average land productivity through increasing the average operational farm size.

By using the agricultural land marketization reform in China in 2008 as the indicator of marketization and the CHNS database, this chapter empirically finds that: first, there exists an inverse relationship between farm size and land productivity in China; second, the agricultural land marketization in China improves the land allocation efficiency and increases the average operational farm size; and third, the higher land allocation efficiency improves the average land productivity by 29.1% and the larger average operational farm size reduces the average land productivity by 9.2%, implying that the agricultural land marketization in China finally improves the average land productivity by 19.9%.

In fact, if land market imperfections are only an minor reason for the existence of the inverse relationship, the inverse relationship may still be quite strong after the agricultural land marketization. As a result, the positive effect of the higher land allocation efficiency may be dominated by the negative effect of the larger average operational farm size, implying that the agricultural land marketization may finally reduce the average land productivity. Then, a direction of future research is to explore how the inverse relationship between farm size and land productivity affects the relationship between agricultural land marketization and average land productivity in other developing countries.

3.7 Appendix

Table 3.1: Data Description

Variables	Description
Dependent Variable	
$\frac{y}{A}$	$\frac{y}{A}$ is the land productivity. This paper uses the output value per unit area to represent the land productivity. The unit is yuan/mu. Considering the effects of price levels in different years, we use the agricultural production price index in China Statistical Yearbook (2011 price=100) to adjust.
Independent Variable	
A	A is the farm size operated by a household. The unit is mu.
Control Variables	
$Labor$	$Labor$ is the total working time of a household in a year. The unit is month/household.
$Raising$	$Raising$ is the intermediate inputs, such as seeds, chemical fertilizers, pesticide, and rents of machines. The unit is yuan.
Age	Age is the household head's age. The unit is year.
$Gend$	$Gend$ is the household head's gender. "1" denotes male and "0" denotes female.
Edu	Edu is the education household head's education level. Primary school, junior high school, senior high school, secondary specialized school, undergraduate, postgraduate or above correspond to 6, 9, 12, 14, 16, and 19 years of education, respectively. The unit is year.
$Marriage$	$Marriage$ is the household head's marriage situation. "1" denotes married and "0" denotes unmarried.
$Hsize$	$Hsize$ is the size of a household. The unit is person.
$Dratio$	$Dratio$ is the ratio of the number of people under 16 or above 60 years old to the total number of people in a household.

Table 3.2: Descriptive Statistics of Sample 2000-2011 (Number of Obs. =5313)
Source: Authors' calculations based on CHNS 2000, 2004, 2006, 2009, and 2011

Variables	Mean	Std. Dev.	Min.	Max.
$\frac{y}{A}$	1534.654	1629.477	1.970	75544.700
A	6.524	9.971	1	200
$Labor$	20.598	11.341	1	79
$Raising$	1362.925	1672.807	10.209	20816.900
Age	52.289	11.194	18.920	88.180
$Gend$	0.910	0.286	0	1
Edu	6.532	3.971	0	16
$Marriage$	0.991	0.097	0	1
$Hsize$	2.776	1.088	1	9
$Dratio$	0.282	0.335	0	1

Table 3.3: Descriptive Statistics of Sample 2000-2006 (Number of Obs. =2827)
Source: Authors' calculations based on CHNS 2000, 2004, and 2006

Variables	Mean	Std. Dev.	Min.	Max.
$\frac{y}{A}$	1439.030	1754.537	1.970	75544.700
A	5.964	7.689	1	90
$Labor$	20.929	11.456	1	79
$Raising$	1092.652	1427.561	10.209	20816.900
Age	50.534	10.987	18.920	88.180
$Gend$	0.915	0.279	0	1
Edu	6.640	3.916	0	16
$Marriage$	0.987	0.114	0	1
$Hsize$	2.876	1.112	1	9
$Dratio$	0.254	0.303	0	1

Table 3.4: Descriptive Statistics of Sample 2009-2011 (Number of Obs. =2486)
Source: Authors' calculations based on CHNS 2009 and 2011

Variables	Mean	Std. Dev.	Min.	Max.
$\frac{y}{A}$	1643.396	1467.225	45.455	35000
A	7.162	12.022	1	200
$Labor$	20.222	11.199	1	76
$Raising$	1670.270	1867.218	20	9999
Age	54.285	11.094	22.030	88.180
$Gend$	0.905	0.293	0	1
Edu	6.409	4.030	0	16
$Marriage$	0.995	0.072	0	1
$Hsize$	2.663	1.050	1	7
$Dratio$	0.314	0.365	0	1

Table 3.5: Descriptive Statistics of Sample 2000 (Number of Obs. =512)

Source: Authors' calculations based on CHNS 2000

Variables	Mean	Std. Dev.	Min.	Max.
$\frac{y}{A}$	1060.061	888.813	29.800	13564.900
A	5.047	5.791	1	75
$Labor$	14.562	8.255	2	60
$Raising$	841.860	830.484	40.066	6868.390
Age	47.022	10.108	21.280	88.180
$Gend$	0.914	0.281	0	1
Edu	6.887	3.782	0	16
$Marriage$	0.980	0.139	0	1
$Hsize$	3.322	1.119	2	7
$Dratio$	0.224	0.218	0	1

Table 3.6: Descriptive Statistics of Sample 2004 (Number of Obs. =1088)

Source: Authors' calculations based on CHNS 2004

Variables	Mean	Std. Dev.	Min.	Max.
$\frac{y}{A}$	1362.206	865.082	1.970	8756.260
A	5.392	6.638	1	75
$Labor$	23.720	11.769	2	72
$Raising$	973.057	1122.650	10.209	10188.700
Age	50.639	10.995	18.920	88.180
$Gend$	0.921	0.270	0	1
Edu	6.778	3.703	0	16
$Marriage$	0.980	0.141	0	1
$Hsize$	2.851	1.109	1	9
$Dratio$	0.259	0.303	0	1

Table 3.7: Descriptive Statistics of Sample 2006 (Number of Obs. =1227)

Source: Authors' calculations based on CHNS 2006

Variables	Mean	Std. Dev.	Min.	Max.
$\frac{y}{A}$	1665.286	2446.763	37.772	75544.700
A	6.852	9.045	1	90
$Labor$	21.110	11.289	1	79
$Raising$	1303.349	1791.435	20.817	20816.900
Age	51.907	11.019	24.070	85.420
$Gend$	0.910	0.286	0	1
Edu	6.414	4.139	0	16
$Marriage$	0.996	0.064	0	1
$Hsize$	2.711	1.062	1	7
$Dratio$	0.264	0.331	0	1

Table 3.8: Descriptive Statistics of Sample 2009 (Number of Obs. =1251)
Source: Authors' calculations based on CHNS 2009

Variables	Mean	Std. Dev.	Min.	Max.
$\frac{y}{A}$	1643.996	1339.678	58.027	22050.300
A	7	10.943	1	130
$Labor$	19.963	11.282	1	76
$Raising$	1563.504	1799.716	29.286	9760.910
Age	53.786	11.246	22.830	88.180
$Gend$	0.900	0.300	0	1
Edu	6.368	4.030	0	16
$Marriage$	0.993	0.085	0	1
$Hsize$	2.678	1.056	1	7
$Dratio$	0.298	0.355	0	1

Table 3.9: Descriptive Statistics of Sample 2011 (Number of Obs. =1235)
Source: Authors' calculations based on CHNS 2011

Variables	Mean	Std. Dev.	Min.	Max.
$\frac{y}{A}$	1642.788	1586.553	45.455	35000
A	7.326	13.026	1	200
$Labor$	20.484	11.112	2	74
$Raising$	1778.420	1927.900	20	9999
Age	54.791	10.918	22.030	86.900
$Gend$	0.910	0.286	0	1
Edu	6.452	4.031	0	16
$Marriage$	0.997	0.057	0	1
$Hsize$	2.649	1.043	1	7
$Dratio$	0.329	0.375	0	1

Table 3.10: Estimation Results for Existence of Inverse Relationship

Note: 1. The standard deviation is inside the parenthesis: *, ** and *** denote the significant level of 10%, 5% and 1%; 2. Village dummies were included in regressions but not reported.

Samples	2000-2011	2000	2004	2006	2009	2011
Dependent Variable: $\ln \frac{y}{A}$						
$\ln A$	-0.587***	-0.549***	-0.689***	-0.671***	-0.556***	-0.453***
	(0.012)	(0.044)	(0.029)	(0.024)	(0.022)	(0.026)
$\ln Labor$	0.112***	-0.046	0.085**	0.217***	0.061**	-0.004
	(0.016)	(0.051)	(0.042)	(0.035)	(0.030)	(0.041)
$\ln Raising$	0.360***	0.218***	0.372***	0.380***	0.359***	0.314***
	(0.010)	(0.039)	(0.025)	(0.020)	(0.021)	(0.023)
Age	0.003***	0.000	0.001	0.004**	-0.001	0.000
	(0.001)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
$Gend$	0.037	-0.156	0.047	0.035	0.025	0.102*
	(0.028)	(0.097)	(0.072)	(0.058)	(0.054)	(0.062)
Edu	-0.004	0.006	-0.001	-0.003	-0.005	-0.009**
	(0.002)	(0.007)	(0.006)	(0.004)	(0.004)	(0.005)
$Marriage$	0.091	-0.128	0.137	-0.292	0.129	0.031
	(0.079)	(0.175)	(0.134)	(0.242)	(0.174)	(0.299)
$Hsize$	-0.033***	0.036	-0.008	-0.039**	-0.009	0.009
	(0.008)	(0.026)	(0.021)	(0.019)	(0.018)	(0.022)
$Dratio$	0.024	0.279**	-0.032	0.016	0.076	-0.027
	(0.027)	(0.119)	(0.067)	(0.055)	(0.052)	(0.057)
$Cons$	5.240***	5.975***	4.993***	5.151***	5.267***	5.596***
	(0.563)	(0.417)	(0.361)	(0.340)	(0.281)	(0.406)
Number of Obs.	5313	512	1088	1227	1251	1235
R^2	0.378	0.511	0.463	0.544	0.452	0.310
Adjusted R^2	0.361	0.349	0.387	0.489	0.386	0.228
F-statistic	21.840	3.156	6.077	9.897	6.916	3.781

Table 3.11: Estimation Results for Effect of Agricultural Land Marketization

Note: 1. The standard deviation is inside the parenthesis: *, ** and *** denote the significant level of 10%, 5% and 1%; 2. The standard deviation of the difference between the coefficients before and after the agricultural land marketization is calculated by $\sigma_\beta = \sqrt{(Std.Err_{2000-2006})^2 + (Std.Err_{2009-2011})^2}$ and the test statistics is $Z = \frac{\beta_{2009-2011} - \beta_{2000-2006}}{\sigma_\beta}$ (see Clogg et al., 1995); 3. Village dummies were included in regressions but not reported.

Samples	2000-2006	2009-2011	Difference
Dependent Variable: $\ln \frac{y}{A}$			
$\ln A$	-0.666***	-0.503***	0.163***
	(0.017)	(0.017)	(0.024)
$\ln Labor$	0.167***	0.030	-0.137***
	(0.021)	(0.024)	(0.032)
$\ln Raising$	0.369***	0.331***	-0.038*
	(0.014)	(0.015)	(0.021)
Age	0.070	0.050	-0.020
	(0.094)	(0.152)	(0.179)
$Gend$	0.014	0.046	0.032
	(0.040)	(0.040)	(0.057)
Edu	0.004***	-0.001	-0.005***
	(0.001)	(0.001)	(0.001)
$Marriage$	-0.001	-0.007**	-0.006
	(0.003)	(0.003)	(0.004)
$Hsize$	-0.046***	0.006	0.052***
	(0.011)	(0.014)	(0.018)
$Dratio$	0.037	0.021	-0.016
	(0.039)	(0.038)	(0.054)
$Cons$	5.159***	5.534***	0.375
	(0.576)	(0.230)	(0.620)
Number of Obs.	2827	2486	
R^2	0.440	0.339	
Adjusted R^2	0.411	0.300	
F-statistic	15.060	8.779	

Table 3.12: Factor Decomposition of Improvement of Average Land Productivity

Note: In the factor decomposition, we omit the variables such that their coefficients are not statistically significant before or after the agricultural land marketization and the variables such that the differences of their coefficients between before and after the marketization are not statistically significant. $\overline{X}$, β and ln denote the sample means of variables, coefficients and natural logarithm, respectively.

Variables	$\ln \overline{X}$			β			$\beta_{after}\,(\ln \overline{X_{after}} - \ln \overline{X_{before}})$	$\ln \overline{X_{before}}\,(\beta_{after} - \beta_{before})$	Contribution
	Before	After	Difference	Before	After	Difference			
$\ln \frac{y}{A}$	7.272	7.405	0.133				-	-	0.133
$\ln A$	1.786	1.969	0.183	-0.666	-0.503	0.163	-0.092	0.291	0.199
$\ln Raising$	6.996	7.421	0.425	0.369	0.331	-0.038	0.141	-0.266	-0.125

Bibliography

ALDERIGHI, M., A. CENTO, P. NIJKAMP, AND P. RIETVELD (2005): "Network Competition—the Coexistence of Hub-and-Spoke and Point-to-Point Systems," *Journal of Air Transport Management*, 11, 328–334.

ALI, D. A. AND K. DEININGER (2015): "Is There a Farm Size-Productivity Relationship in African Agriculture? Evidence from Rwanda," *Land Economics*, 91, 317–343.

ALVAREZ, A. AND C. ARIAS (2004): "Technical Efficiency and Farm Size: A Conditional Analysis," *Agricultural Economics*, 30, 241–250.

ASSUNÇÃO, J. J. AND L. H. B. BRAIDO (2007): "Testing Household-Specific Explanations for the Inverse Productivity Relationship," *American Journal of Agricultural Economics*, 89, 980–990.

BARDHAN, P. K. (1973): "Size, Productivity, and Returns to Scale: An Analysis of Farm-Level Data in Indian Agriculture," *Journal of Political Economy*, 81, 1370–1386.

BARLA, P. AND C. CONSTANTATOS (2005): "Strategic Interactions and Airline Network Morphology Under Demand Uncertainty," *European Economic Review*, 49, 703–716.

BARON, D. P. AND D. BESANKO (1984): "Regulation, Asymmetric Information, and Auditing," *The RAND Journal of Economics*, 15, 447–470.

BARON, D. P. AND R. B. MYERSON (1982): "Regulating a Monopolist with Unknown Costs," *Econometrica*, 50, 911–930.

BARRETT, C. B. (1996): "On Price Risk and the Inverse Farm Size-Productivity Relationship," *Journal of Development Economics*, 51, 193–215.

Barrett, C. B., M. F. Bellemare, and J. Y. Hou (2010): "Reconsidering Conventional Explanations of the Inverse Productivity-Size Relationship," *World Development*, 38, 88–97.

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

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

Benjamin, D. (1995): "Can Unobserved Land Quality Explain the Inverse Productivity Relationship?" *Journal of Development Economics*, 46, 51–84.

Benjamin, D. and L. Brandt (2002): "Property Rights, Labour Markets, and Efficiency in a Transition Economy: The Case of Rural China," *Canadian Journal of Economics/Revue canadienne d'économique*, 35, 689–716.

Berechman, J. and O. Shy (1998): *The Structure of Airline Equilibrium Networks*, Recent Advances in Spatial Equilibrium Modelling, Springer, Berlin.

Berry, R. A. and W. R. Cline (1979): *Agrarian Structure and Productivity in Developing Countries: A Study Prepared for the International Labour Office Within the Framework of the World Employment Programme*, Baltimore, Maryland, USA: Johns Hopkins Univ. Press.

Berry, S. and P. Jia (2010): "Tracing the Woes: An Empirical Analysis of the Airline Industry," *American Economic Journal: Microeconomics*, 2, 1–43.

Bhalla, S. S. and P. Roy (1988): "Mis-Specification in Farm Productivity Analysis: The Role of Land Quality," *Oxford Economic Papers*, 40, 55–73.

Bilotkach, V., X. Fageda, and R. Flores-Fillol (2013): "Airline Consolidation and the Distribution of Traffic Between Primary and Secondary Hubs," *Regional Science and Urban Economics*, 43, 951–963.

Binswanger, H. P., K. Deininger, and G. Feder (1995): "Chapter 42 Power, Distortions, Revolt and Reform in Agricultural Land Relations," *Handbook of Development Economics*, 3, 2659–2772.

Bittlingmayer, G. (1990): "Efficiency and Entry in a Simple Airline Network," *International Journal of Industrial Organization*, 8, 245–257.

Brueckner, J. K. (2002): "Airport Congestion When Carriers Have Market Power," *The American Economic Review*, 92, 1357–1375.

——— (2004): "Network Structure and Airline Scheduling," *The Journal of Industrial Economics*, 52, 291–312.

——— (2005): "Internalization of Airport Congestion: A Network Analysis," *International Journal of Industrial Organization*, 23, 599–614.

Brueckner, J. K., N. J. Dyer, and P. T. Spiller (1992): "Fare Determination in Airline Hub-and-Spoke Networks," *The RAND Journal of Economics*, 23, 309–333.

Brueckner, J. K. and R. Flores-Fillol (2007): "Airline Schedule Competition," *Review of Industrial Organization*, 30, 161–177.

Brueckner, J. K. and P. T. Spiller (1991): "Competition and Mergers in Airline Networks," *International Journal of Industrial Organization*, 9, 323–342.

——— (1994): "Economies of Traffic Density in the Deregulated Airline Industry," *The Journal of Law and Economics*, 37, 379–415.

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

Brueckner, J. K. and W. T. Whalen (2000): "The Price Effects of International Airline Alliances," *The Journal of Law and Economics*, 43, 503–546.

Brueckner, J. K. and Y. Zhang (2001): "A Model of Scheduling in Airline Networks: How a Hub-and-Spoke System Affects Flight Frequency, Fares and Welfare," *Journal of Transport Economics and Policy*, 35, 195–222.

Byiringiro, F. and T. Reardon (1996): "Farm Productivity in Rwanda: Effects of Farm Size, Erosion, and Soil Conservation Investments," *Agricultural Economics*, 15, 127–136.

Caillaud, B., R. Guesnerie, P. Rey, and J. Tirole (1988): "Government Intervention in Production and Incentives Theory: A Review of Recent Contributions," *The RAND Journal of Economics*, 19, 1–26.

Carletto, C., S. Savastano, and A. Zezza (2013): "Fact or Artifact: The Impact of Measurement Errors on the Farm Size-Productivity Relationship," *Journal of Development Economics*, 103, 254–261.

Carter, M. R. (1984): "Identification of the Inverse Relationship Between Farm Size and Productivity: An Empirical Analysis of Peasant Agricultural Production," *Oxford Economic Papers*, 36, 131–145.

——— (1988): "Equilibrium credit rationing of small farm agriculture," *Journal of Development Economics*, 28, 83–103.

Carter, M. R. and Y. Yao (2002): "Local versus Global Separability in Agricultural Household Models: The Factor Price Equalization Effect of Land Transfer Rights," *American Journal of Agricultural Economics*, 84, 702–715.

Chayanov, A. V. (1926): *The Theory of Peasant Economy*, Irwin: Homewood.

Chen, Z., W. E. Huffman, and S. Rozelle (2011): "Inverse Relationship Between Productivity and Farm Size: The Case of China," *Contemporary Economic Policy*, 29, 580–592.

Clogg, C. C., E. Petkova, and A. Haritou (1995): "Statistical Methods for Comparing Regression Coefficients Between Models," *American Journal of Sociology*, 100, 1261–1293.

Collier, P. (1983): "Malfunctioning of African Rural Factor Markets: Theory and A Kenyan Example," *Oxford Bulletin of Economics and Statistics*, 45, 141–172.

Cook, A. and G. Tanner (2011): "European Airline Delay Cost Reference Values, for EUROCONTROL Performance Review Unit," Tech. rep., University of Westminster.

——— (2015): "European Airline Delay Cost Reference Values - Updated and Extended Values," Tech. rep., University of Westminster.

Cornia, G. A. (1985): "Farm Size, Land Yields and the Agricultural Production Function: An Analysis for Fifteen Developing Countries," *World Development*, 13, 513–534.

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

Deininger, K., D. A. Ali, and T. Alemu (2008a): "Assessing the Functioning of Land Rental Markets in Ethiopia," *Economic Development and Cultural Change*, 57, 67–100.

Deininger, K. and S. Jin (2005): "The Potential of Land Rental Markets in the Process of Economic Development: Evidence from China," *Journal of Development Economics*, 78, 241–270.

——— (2008): "Land Sales and Rental Markets in Transition: Evidence from Rural Vietnam," *Oxford Bulletin of Economics and Statistics*, 70, 67–101.

Deininger, K., S. Jin, and H. K. Nagarajan (2008b): "Efficiency and Equity Impacts of Rural Land Rental Restrictions: Evidence from India," *European Economic Review*, 52, 892–918.

Dixit, A. (1979): "A Model of Duopoly Suggesting a Theory of Entry Barriers," *The Bell Journal of Economics*, 10, 20–32.

Doganis, R. (2009): *Flying Off Course IV: Airline Economics and Marketing*, New York: Routledge, fourth ed.

Douglas, G. W. and J. C. Miller (1974): "Quality Competition, Industry Equilibrium, and Efficiency in the Price-Constrained Airline Market," *The American Economic Review*, 64, 657–669.

Dunn, A. (2008): "Do Low-Quality Products Affect High-Quality Entry? Multiproduct Firms and Nonstop Entry in Airline Markets," *International Journal of Industrial Organization*, 26, 1074–1089.

Encaoua, D., M. Moreaux, and A. Perrot (1996): "Compatibility and Competition in Airlines Demand Side Network Effects," *International Journal of Industrial Organization*, 14, 701–726.

Eswaran, M. and A. Kotwal (1986): "Access to Capital and Agrarian Production Organisation," *The Economic Journal*, 96, 482–498.

EUROCONTROL (2010): *CODA Digest - Delays to Air Transport in Europe - Annual 2009*, Brussels: EUROCONTROL.

——— (2011): *CODA Digest - Delays to Air Transport in Europe - Annual 2010*, Brussels: EUROCONTROL.

——— (2016): *CODA Digest - All-Causes Delay and Cancellations to Air Transport in Europe - Annual 2015*, Brussels: EUROCONTROL.

European Union and EUROCONTROL (2015): *European ATM Master Plan*, Luxembourg: Publications Office of the European Union, 2015 ed.

Fageda, X. and R. Flores-Fillol (2012): "On the Optimal Distribution of Traffic of Network Airlines," *European Economic Review*, 56, 1164–1179.

——— (2015): "A Note on Optimal Airline Networks Under Airport Congestion," *Economics Letters*, 128, 90–94.

FEDER, G. (1985): "The relation between farm size and farm productivity," *Journal of Development Economics*, 18, 297–313.

FLORES-FILLOL, R. (2009): "Airline Competition and Network Structure," *Transportation Research Part B: Methodological*, 43, 966–983.

——— (2010): "Congested Hubs," *Transportation Research Part B: Methodological*, 44, 358–370.

FRISVOLD, G. B. (1994): "Does supervision matter? Some hypothesis tests using Indian farm-level data," *Journal of Development Economics*, 43, 217–238.

HELTBERG, R. (1998): "Rural Market Imperfections and the Farm Size-Productivity Relationship: Evidence from Pakistan," *World Development*, 26, 1807–1826.

HENDRICKS, K., M. PICCIONE, AND G. TAN (1995): "The Economics of Hubs: The Case of Monopoly," *The Review of Economic Studies*, 62, 83–99.

——— (1997): "Entry and Exit in Hub-Spoke Networks," *The RAND Journal of Economics*, 28, 291–303.

——— (1999): "Equilibria in Networks," *Econometrica*, 67, 1407–1434.

HU, Q. J. (2010): "Network Game and Capacity Investment Under Market Uncertainty," *Production and Operations Management*, 19, 98–110.

JIN, S. AND K. DEININGER (2009): "Land Rental Markets in the Process of Rural Structural Transformation: Productivity and Equity Impacts from China," *Journal of Comparative Economics*, 37, 629–646.

KAWASAKI, A. (2008): "Network Effects, Heterogeneous Time Value and Network Formation in the Airline Market," *Regional Science and Urban Economics*, 38, 388–403.

KAWASAKI, K. (2010): "The Costs and Benefits of Land Fragmentation of Rice Farms in Japan," *Australian Journal of Agricultural and Resource Economics*, 54, 509–526.

Kevane, M. (1996): "Agrarian Structure and Agricultural Practice: Typology and Application to Western Sudan," *American Journal of Agricultural Economics*, 78, 236–245.

Kimhi, A. (2006): "Plot Size and Maize Productivity in Zambia: Is There An Inverse Relationship?" *Agricultural Economics*, 35, 1–9.

Laffont, J.-J. and J. Tirole (1986): "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, 94, 614–641.

Lamb, R. L. (2003): "Inverse Productivity: Land Quality, Labor Markets, and Measurement Error," *Journal of Development Economics*, 71, 71–95.

Larson, D. F., K. Otsuka, T. Matsumoto, and T. Kilic (2014): "Should African Rural Development Strategies Depend on Smallholder Farms? An Exploration of the Inverse-Productivity Hypothesis," *Agricultural Economics*, 45, 355–367.

Lau, L. J. and P. A. Yotopoulos (1971): "A Test for Relative Efficiency and Application to Indian Agriculture," *The American Economic Review*, 61, 94–109.

Li, G., Z. Feng, L. You, and L. Fan (2013): "Re-Examining the Inverse Relationship Between Farm Size and Efficiency: The Empirical Evidence in China," *China Agricultural Economic Review*, 5, 473–488.

Lin, J. Y. (1995): "Endowments, Technology, and Factor Markets: A Natural Experiment of Induced Institutional Innovation from China's Rural Reform," *American Journal of Agricultural Economics*, 77, 231–242.

Lin, M. H. (2012): "Airlines-Within-Airlines Strategies and Existence of Low-Cost Carriers," *Transportation Research Part E: Logistics and Transportation Review*, 48, 637–651.

Liu, S., M. R. Carter, and Y. Yao (1998): "Dimensions and Diversity of Property Rights in Rural China: Dilemmas on the Road to Further Reform," *World Development*, 26, 1789–1806.

Ma, G. (2008): "Stability of Leasing Farmland and Reality of Abandoning Farmland," *Rural Economy*, 7, 43–45.

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

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

Mussa, M. and S. Rosen (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18, 301–317.

Newell, A., K. Pandya, and J. Symons (1997): "Farm Size and the Intensity of Land Use in Gujarat," *Oxford Economic Papers*, 49, 307–315.

OAG (2016): *OAG flightview: Airline Monthly OTP - July 2016*.

Oum, T. H., A. Zhang, and Y. Zhang (1995): "Airline Network Rivalry," *The Canadian Journal of Economics/Revue canadienne d'Economique*, 28, 836–857.

Panzar, J. C. (1979): "Equilibrium and Welfare in Unregulated Airline Markets," *The American Economic Review*, 69, 92–95.

Pearce, B. (2013): *Profitability and the Air Transport Value Chain*, Switzerland: IATA.

Pels, E., P. Nijkamp, and P. Rietveld (2000): "A Note on the Optimality of Airline Networks," *Economics Letters*, 69, 429–434.

Pels, E. and E. T. Verhoef (2004): "The Economics of Airport Congestion Pricing," *Journal of Urban Economics*, 55, 257–277.

Rao, V. and T. Chotigeat (1981): "The Inverse Relationship Between Size of Land Holdings and Agricultural Productivity," *American Journal of Agricultural Economics*, 63, 571–574.

Restuccia, D. and R. Santaeulalia-Llopis (2017): "Land Misallocation and Productivity," *NBER Working Paper.*

Rietveld, P. and M. Brons (2001): "Quality of Hub-and-Spoke Networks; The Effects of Timetable Co-ordination on Waiting Time and Rescheduling Time," *Journal of Air Transport Management*, 7, 241–249.

Rosenzweig, M. R. and H. P. Binswanger (1993): "Wealth, Weather Risk and the Composition and Profitability of Agricultural Investments," *The Economic Journal*, 103, 56–78.

Rosenzweig, M. R. and K. I. Wolpin (1985): "Specific Experience, Household Structure, and Intergenerational Transfers: Farm Family Land and Labor Arrangements in Developing Countries," *The Quarterly Journal of Economics*, 100, 961–987.

Sen, A. (1962): "An Aspect of Indian Agriculture," *Economic Weekly*, Annual Number 14, 243–246.

Sen, A. K. (1966): "Peasants and Dualism With or Without Surplus Labor," *Journal of Political Economy*, 74, 425–450.

Silva, H. E., E. T. Verhoef, and V. A. van den Berg (2014): "Airline Route Structure Competition and Network Policy," *Transportation Research Part B: Methodological*, 67, 320–343.

Singh, N. and X. Vives (1984): "Price and Quantity Competition in a Differentiated Duopoly," *The RAND Journal of Economics*, 15, 546–554.

Smyth, M. and B. Pearce (2007): *Airline Liberalisation*, Switzerland: IATA.

Starr, R. M. and M. B. Stinchcombe (1992): "Efficient Transportation Routing and Natural Monopoly in the Airline Industry: An Economic Analysis of Hub-Spoke and Related Systems," *Working Paper.*

STATFOR (2013): *Challenges of Growth 2013*, EUROCONTROL.

Sumner, D. A. (2014): "American Farms Keep Growing: Size, Productivity, and Policy," *Journal of Economic Perspectives*, 28, 147–166.

University of Westminster (2015): "The Cost of Passenger Delay to Airlines in Europe - Consultation Document," Tech. rep., University of Westminster.

US Federal Aviation Administration (1969): *Airport Capacity Handbook*, Washington DC: US Government Printing Office, second ed.

Wang, C. (2017): "Congestion Delays, Horizontal Product Differentiation and Airline Networks," *Working Paper*.

Wang, X. (2016): "1-Hub, 2-Hub or Fully Connected Network? A Theoretical Analysis of the Optimality of Airline Network Structure," *Economics of Transportation*, 5, 12–23.

Wiens, T. B. (1977): "Uncertainty and Factor Allocation in a Peasant Economy," *Oxford Economic Papers*, 29, 48–60.

Wu, J. (1993): "A Review for Abandoning Farmland in Rural Areas," *China Rural Economy*, 9.

Xu, W. and H. Zhang (1993): "A Serious Problem in Current China's Rural Areas: A Survey for Abandoning Farmland in Anhui Province," *China Rural Economy*, 3.

Yang, H. and A. Zhang (2011): "Price-Cap Regulation of Congested Airports," *Journal of Regulatory Economics*, 39, 293–312.

Yao, Y. (2000): "The Development of the Land Lease Market in Rural China," *Land Economics*, 76, 252–266.

Zhang, A. and Y. Zhang (1997): "Concession Revenue and Optimal Airport Pricing," *Transportation Research Part E: Logistics and Transportation Review*, 33, 287–296.

——— (2003): "Airport Charges and Capacity Expansion: Effects of Concessions and Privatization," *Journal of Urban Economics*, 53, 54–75.

——— (2006): "Airport Capacity and Congestion When Carriers Have Market Power," *Journal of Urban Economics*, 60, 229–247.