Price of Anarchy in Supply Chains, Congested Systems and Joint Ventures

by

Wei Sun

B.E., National University of Singapore (2005)
M.S., Massachusetts Institute of Technology (2006)

Submitted to the SLOAN SCHOOL OF MANAGEMENT
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Operations Research

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2012

Author ....
SLOAN SCHOOL OF MANAGEMENT
August 15, 2012

Certified by ....
Georgia Perakis
William F. Pounds Professor of Management
Thesis Supervisor

Accepted by ....
Dimitris Bertsimas
Boeing Professor of Operations Research
Co-director, Operations Research Center

Price of Anarchy in Supply Chains, Congested Systems and Joint Ventures

by

Wei Sun

Submitted to the SLOAN SCHOOL OF MANAGEMENT
on August 15, 2012, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Operations Research

Abstract

This thesis studies the price of anarchy in supply chains, congested systems and joint ventures. It consists of three main parts. In the first part, we investigate the impact of imperfect competition with nonlinear demand. We focus on a distribution channel with a single supplier and multiple downstream retailers. To evaluate the performance, we consider several metrics, including market penetration, total profit, social welfare and rent extraction. We quantify the performance with tight upper and lower bounds. We show that with substitutes, while competition improves the efficiency of a decentralized supply chain, the asymmetry among the retailers deteriorates the performance. The reverse happens when retailers carry complements. We also show that efficiency of a supply chain with concave (convex) demand is higher (lower) than that with affine demand.

The second part of the thesis studies the impact of congestion in an oligopoly by incorporating convex costs. Costs could be fully self-contained or have a spillover component, which depends on others' output. We show that when costs are fully self-contained, the welfare loss in an oligopoly is at most 25% of the social optimum, even in the presence of highly convex costs. With spillover cost, the performance of an oligopoly depends on the relative magnitude of spillover cost to the marginal benefit to consumers. In particular, when spillover cost outweighs the marginal benefit, the welfare loss could be arbitrarily bad.

The third part of the thesis focuses on capacity planning with resource pooling in joint ventures under demand uncertainties. We distinguish heterogeneous and homogeneous resource pooling. When resources are heterogeneous, the effective capacity in a joint venture is constrained by the minimum individual contribution. We show that there exists a unique constant marginal revenue sharing scheme which induces the same outcome in a Nash equilibrium, Nash Bargaining and the system optimum. The optimal scheme rewards every participant proportionally with respect to his marginal cost. When resources are homogeneous, we show that the revenue sharing ratio should be inversely proportional to a participant's marginal cost.

Thesis Supervisor: Georgia Perakis
Title: William F. Pounds Professor of Management

Acknowledgments

First and foremost, I want to thank Georgia Perakis, my advisor, for her invaluable guidance, perpetual patience and words of encouragement throughout the past five years. It is a great fortune that I have the privilege to work with her, and know her not only as an advisor but also as a friend. Her advice (in both academic and nonacademic aspects), support and warmth have made the existence of this thesis possible.

I would also like to thank the members of my committee Charlie Fine, Retsef Levi and Mahesh Nagarajan for their insights and encouragement along the way. Apart from providing me with invaluable research feedback, they have been a continuous source of support and advice. Thank you, Retsef and Charlie, for allowing me to be your teaching assistant twice! I have learnt from you so much about supply chain management. Thank you, Mahesh, for your incredible kindness and warm thoughts during my entire job search process.

The Operations Research Center has been my home at MIT for the last five years. I would like to thank Dimitris Bertsimas, and Patrick Jaillet for serving as co-directors during my time at the center. My sincere appreciation also goes to Laura Rose and Andrew Carvalho for making sure the center runs smoothly and the students have everything they need to work. The best thing about the ORC has been the wonderful group of colleagues and friends, with whom I have shared many hours of work and fun. All the laughters, encouragement and (occasional) complaints made up the unforgettable memory of this place.

I also like to take this opportunity to thank a few others who greatly shaped my nonacademic life. Thank you, Abby Han Men, for being my closest friend and inspiring me to embark on this Ph.D journey. Thank you, Dmitry Malioutov, for your support and encouragement which have kept me going through the toughest of times.

Above all, no words can fully express my thankfulness to my parents. I am extremely grateful for their unconditional love and faith in me. I owe them more than I could ever express here. To them I dedicate this thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Decentralized systems with agents who operate within their sphere of interest are widely recognized as less efficient than their centralized counterparts. This phenomenon is pervasive in social, political, economical, and even biological systems. Papadimitriou (2001) in his seminal work, coined the term *Price of Anarchy* (PoA), which measures the difference in the performance of a decentralized system to a fully centralized system. Quantifying this value is essential in predicting system behavior and in designing appropriate rules of action to improve its performance. The main purpose of this research is to provide new ways for understanding and evaluating efficiency loss in complex competitive environments including supply chains, congested systems and joint ventures.

1.1 Thesis Outline

The thesis consists of three main parts. In the first part, we focus on a distribution channel with a single supplier and multiple competing retailers. While our application is in supply chain management, the base game-theoretic framework which we study is generic and captures many other applications where oligopolistic competition exists. To evaluate the performance, we consider several metrics, such as market penetration, total profit, social welfare and rent extraction. The second part of my dissertation incorporates a by-product of competition: externalities. We study the impact of selfish behavior on societal welfare. Our model consists of several service providers, competing for users who are sensitive to both prices and congestion. By identifying two types of congestion effects which depend on whether one service provider's service level could affect others' congestion cost, we measure the welfare loss in a decentralized setting and propose novel implementation of congestion pricing which appeals to the self-interest of participants. The third part of my thesis describes a joint work with Retsef Levi, Cong Shi and my advisor, Georgia Perakis. This work is motivated by the growing popularity of joint-ventures in the past decade. Decisions in such settings are distorted by self-interests and demand uncertainties. They are further complicated by joint business constraints.

1.2 Main Contributions

1.2.1 Supply chains under imperfect competition and nonlinear demand

In the first part of the dissertation, we study a distribution channel with a single supplier and serval downstream retailers, whose demand depends on the prices of all available products in the market. The main objective of this work is to understand how imperfect competition, demand nonlinearity and the nature of products (substitutable versus complementary products) affect the performance of a supply chain.

To analyze imperfect competition, we associate product substitutability (or complementarity) with a ratio between the *inter-firm* and the *intra-firm* price sensitivity coefficients. This measure captures the interdependence of one product with respect to other available products in the market, implying the strength of a retailer with respect to his competitors. Based on this measure, by only knowing information on two "representative" retailers, we are able to analytically quantify various performance metrics including market penetration, channel profit, social welfare and rent extraction for a supply chain with an arbitrary number of retailers with tight upper and lower bounds.

We show that when the retailers compete with substitutes, the two "strongest" retailers predominantly determine the supply chain performance. Although asymmetry between them deteriorates the performance, a decentralized supply chain with substitutes is fairly efficient as suggested by the performance metrics. For complementary products, we characterize how much the performance is affected by product complementarity. The bounds imply that the performance could be primarily determined by the two "weakest" retailers. Although asymmetry between them has some countervailing effect to combat this inefficiency, a decentralized supply chain with complements generally demonstrates a significant loss of efficiency.

With nonlinear demand, we utilize the concept of "Jacobian similarity" to describe the curvature of the demand function. We show that when the demand function is concave, the decentralized supply chain is more efficient than one with affine demand. To be more specific, with concave demand, the profit loss in a decentralized supply chain is at most 25% of the optimal profit in a coordinated setting and improves as the intensity of competition among retailers increases. On the other hand, with convex demand, the inefficiency is relatively higher than one with affine demand.

1.2.2 Congested systems with convex costs: self-contained versus spillover

In this part of the thesis, we consider competition in the presence of congestion effects. In contrast to many studies on congestion games which consider infinitesimal users who are price-takers with no market power (e.g., Dafermos and Sparrow 1969, Wardrop 1952), players in our model are competing oligopolists who have sufficient market power to influence prices. Our model consists of several service providers with differentiated services, competing for users who are sensitive to both prices and congestion costs which are convex and increasing with the output level. We study two types of congestion effects depending on whether one service provider's congestion cost could be influenced by other providers' output level. Congestion effect is said to be *self-contained*, when the congestion cost associated with one service provider only depends on his output level. Bandwidth congestion is one such example, where carriers use dedicated frequency bands for transmissions to avoid signal interference. As a result, service degradation is only experienced by users in the affected network. This is in contrast with airport congestion, where when one airline schedules an additional flight in a congested airport, it creates additional delays for every flight which attempts to land and take off. Congestion in this setting also has the *spillover* effect as everyone in the system experiences additional delay.

We compare the total welfare in an unregulated setting where the service providers have free access to the facility to that of the social optimum which maximizes the total surplus, so as to assess how much welfare is lost due to decentralization. We show that with fully self-contained cost, the maximum welfare loss in the unregulated setting is limited to 25% even in the presence of highly nonlinear convex costs. Moreover, the efficiency of the unregulated setting improves as the competition among the service providers increases. The main insight is that, when costs are self-contained, service providers take those into fully consideration when they determine their own output level. As a result, price of anarchy in this case is bounded. With spillover congestion cost, the performance highly depends on the relative magnitude of the marginal *uninternalized* congestion cost and the marginal consumer surplus when an additional user is enrolled. In particular, we show that when the marginal uninternalized congestion cost outweighs the marginal benefit, the welfare loss in the unregulated setting could be arbitrarily high even with affine costs. The latter validates the need of implementing some rationing mechanisms in airports to curb congestion since there are rooms for substantial potential welfare gains.

1.2.3 Joint ventures with resource pooling and demand uncertainties

The work is motivated by the growing popularity of joint-ventures in the recent years. In this work, we study capacity planning with resource pooling in a joint venture under demand uncertainties. We distinguish two types of resources pooling, based on whether the resources are heterogeneous or homogeneous. We assume each entity in a joint venture contributes one type of resources. When resources are heterogeneous, they are not fully substitutable. It implies that the contribution from one entity cannot be fully replaced by others. This is in contrast with homogenous resource pooling, where resources are fully substitutable.

We show that with heterogeneous resource pooling, the effective capacity in a joint venture is constrained by the minimum individual contribution. In addition, every participant is committed to make an equal contribution in a joint venture. We also show that, there exists a same efficient and fair revenue sharing scheme in both Nash equilibrium and Nash Bargaining solution. The optimal scheme rewards every participant proportionally to his marginal cost.

When resources are homogeneous, however, there does not exist a revenue sharing scheme which induces actions to achieve the optimum which maximizes the collective profit. Nonetheless, we propose some methods to share revenue with the worst case performance guarantee for general convex costs. The methods suggest that the reward should be inversely proportional to the marginal cost of each participant with homogeneous resources.

The rest of the thesis is organized as follows. The work on supply chains is broken into Chapter 2 and 3. Chapter 2 focuses on a setting with nonlinear demand while each retailer carries a single type of substitutable products. In Chapter 3, we consider affine demand while each retailer carries multiple complementary products. Chapter 4 is devoted to the study on congested systems. The work on joint ventures can be found in Chapter 5. We conclude the thesis in Chapter 6.

Chapter 2

Price of Anarchy in Supply Chains with Imperfect Competition and Nonlinear Demand

2.1 Introduction

In this paper, we develop tight upper and lower bounds to quantify the loss due to decentralization in a two-tier supply chain with price-only contracts when there is imperfect competition among downstream retailers with nonlinear demand. Browsing through a comparison shopping engine often reveals a surprisingly wide dispersion of prices for a same product. One of the reasons for price differentiation is imperfect competition induced by retailer asymmetry in the market. That is, retailers are perceived differently by consumers based on market share, brand name, reputation, availability, etc.

Despite a large and growing literature studying the issue of supply chain coordination, most papers make one of the following assumptions: (1) monopoly or a duopoly; (2) independent (noncompeting) retailers; or (3) symmetric retailers with homogenous products; (4) affine demand. Each of these assumptions imposes significant limitations. Some results immediately fail once we relax the assumptions.

For example, it is well-known that with noncompeting or symmetric retailers, every retailer in a decentralized setting experiences sales decline compared to a centralized setting due to "double marginalization". However, an example in Section 3.3 shows that in an asymmetric duopoly, one retailer may charge a higher price yet sell more in a decentralized setting. Under the symmetry assumption, it has been established in various settings that a decentralized supply chain achieves higher channel profit as competition in the retail market intensifies, where the intensity of competition is measured by the number of retailers (Tyagi 1999, Mahajan and Van Ryzin 2001, Cachon and Lariviere 2005). When retailers are asymmetric, for example, with dominant retailers such as Wal-Mart and fringe retailers like local grocery stores coexist in a market, it is ambiguous how to measure "competition", let alone quantify a supply chain's performance.

2.1.1 Contributions

In this work, we relax all the assumptions aforementioned. Our main contributions of this chapter are as follows.

Analytical upper and lower bounds with competition index. We associate the level of competition faced by a product with a ratio between its cross and own elasticities. This measure indicates the *relative* strength of a retailer's product with respect to his competitors, and captures the interdependence of the products in that market. Based on this measure, by only knowing information on two "representative" retailers, we are able to analytically quantify various performance metrics including *market penetration*, *channel profit*, *social welfare* and *rent extraction* for a supply chain with an arbitrary number of retailers by deriving their respective upper and lower bounds.

We present families of demand functions for which the bounds are *tight*. We also use simulation to show that the bounds achieve an accuracy of within 7% under a more general setting. To establish the analytical bounds, we develop the analysis by utilizing tools such as Cassini ovals of eigenvalues and copositivity from matrix analysis (Horn and Johnson 1985). We believe the methodology proposed in this work could potentially be used in other problems.

Impact of imperfect competition. By first focusing on affine demand, we quantify how much competition induced by product substitutability promotes output levels, channel profit and social welfare. We show that the two "strongest" retailers predominantly determine the channel performance. Although asymmetry between them deteriorates the performance, a decentralized supply chain with substitutes is fairly efficient as suggested by the performance metrics. We show that with affine demand, the profit loss in a decentralized supply chain is *always less than* 25%. In many real life scenarios supported by empirical evidences, the loss is well within 15%, which implies that price-only contracts are often "good enough" for supply chains in practice.

Impact of nonlinearity of demand. We show that when the demand function is concave, the decentralized supply chain is more efficient than that with affine demand. To be more specific, the profit loss in a decentralized supply chain is at most 25% of the optimal profit and improves as the intensity of competition among retailers increases. On the other hand, with convex demand, the inefficiency is relatively higher than that with affine demand. The intuition is that inefficiency in a decentralized supply chain is induced by the successive markups imposed by the supplier and retailers. When demand is concave (convex), a given price change induces a proportionally smaller (greater) change in demand than with affine demand.

Rent extraction in a decentralized chain. The study on the profit allocation reveals that the supplier is always guaranteed with a larger share of the channel profit. Moreover, the supplier enjoys a lion share of over 66% of the channel profit and her share increases further as the competition in the retail market intensifies. This result offers some explanation for the prevalence of price-only contracts in practice since they are desirable from the perspective of the supplier.

2.1.2 Relevant literature

Two sources of competition exist in a decentralized supply chain: (i) vertical competition between the supplier and the retailers and (ii) horizontal competition among the retailers. Since Spengler (1950) who identified the double marginalization effect, the problem of channel coordination and its relevant issues have generated considerable research in both the marketing and economic literature. (e.g., Choi 1991, Krishnan and Winter 2007, Moorthy 1987 and Pasternack 1985). During the last decade, the issue of coordination in supply chains has also gained a lot of attention in the operations management literature (see Cachon 2003 for a review).

Despite numerous coordinating contracts proposed in the academic literature (e.g., Corbett et al. 2005, Bernstein and Federgruen 2003, 2007, Cachon and Kok 2010), simple price-only contracts are more often observed in practice. Their prevalence suggests the importance of quantifying the "price of decentralization", that is, the decentralized channel performance in comparison to a centralized chain with price-only contracts. Such analysis measures the potential gains through coordination, and thus, allows managers to gauge whether there is a need to implement more complex contracts. Perakis and Roels (2007) formally characterize the profit loss for various supply chain configurations when retail prices are exogenous. Adida and DeMiguel (2010) study a supply chain with multiple suppliers and multiple risk-averse retailers in a Cournot oligopoly. Nevertheless, the authors primarily perform comparative analysis for symmetric manufacturers and retailers with affine demand. For the asymmetric setting, based on numerical simulations, they conclude that some results obtained for the symmetric chain do not hold for asymmetric retailers. Their observation reaffirms the need to study this problem with asymmetric retailers. Netessine and Zhang (2005) study the performance of a distribution channel from the supply side, where the demand of one retailer is concave with respect to the stock level of his competitors. They conclude that retailers with complements tend to understock compared to the centralized setting, whereas they tend to overstock with substitutes. Thus, Netessine and Zhang (2005) conclude that complements (substitutes) aggravates (compensates) the double-marginalization effect.

Linear, convex, and concave demand functions can be found in the literature. Robinson (1933)'s pioneering analysis, taken forward by Schmalensee (1981) studied how the curvature of demand functions affect the output level in an monopoly. The primary focus of the this body of literature (e.g., Varian 1985, Holmes 1989, Schwartz

1990, Yoshida 2000 etc.) is on how social welfare is affected when price discrimination exists in a market. It is well-known that with a affine demand, a profit-maximizing monopolist produces exactly half of the socially optimal output. Malueg (1994) shows that when the demand is concave (convex), the output level becomes at least (at most) half of the socially optimal quantity. Moreover, when demand is concave, the ratio between the welfare loss in the monopoly to the social optimum is bounded. The similar argument, however, does not apply when demand is convex. In this work, we will investigate the impact of demand curvature on the performance of a supply chain.

Lastly, our work which measures the performance of a decentralized system with respect to a centralized system is related to a stream of literature on *price of anarchy*, popularized by Koutsoupias and Papadimitriou (1999). It compares the performance of the worst-case Nash equilibrium with respect to the centralized system. The concept has been used in transportation networks (Roughgarden and Tardos 2002, Correa et al. 2004, 2007, Roughgarden 2005), network pricing (Acemoglu and Ozdaglar 2007, Johari and Van Roy 2009), oligopolistic pricing games in a single tier (Farahat and Perakis 2010a,b), and supply chain games with exogenous pricing (Perakis and Roels 2007, Martinez-de Alberniz and Simchi-Levi 2009, Martinez-de Alberniz and Roels 2010).

The rest of the chapter is organized as follows. Section 2.2 describes our model and assumptions. Section 2.3 begins with affine demand, and investigates the performance by quantifying various performance metrics to showcase the impact of imperfect competition. We will incorporate nonliear demand in Section 2.4. Several simulation experiments which evaluate the tightness of the bounds can be found in Section 2.5. Lastly, we will discuss some extensions of the current model and conclude the paper in Section 2.6.

2.2 Problem Formulation

2.2.1 Model and assumption

We consider a two tier supply chain with a single supplier and n retailers. We denote the retail price set by retailer $i = 1, 2, ..., n$ by p_i and we use $\mathbf{p} = (p_1, ..., p_n)$ to denote the entire price vector set by all the retailers. Each retailer purchases one type of product from the supplier. We will show that this assumption can be easily relaxed in Section 2.6. We denote the constant marginal production cost incurred by the supplier when she fulfills retailer i's order as c_i and let vector $\mathbf{c} = (c_1, ..., c_n)$. The marginal cost may vary across the retailers as different orders may require different levels of effort from the supplier. The supplier has to determine the contract, the wholesale price w_i for each retailer, and vector $\mathbf{w} = (w_1, ..., w_n)$. Each retailer faces a deterministic demand, $q_i(\mathbf{p})$, which depends on the prices set by all the retailers.

The decision sequence is as follows. The supplier initiates the process by proposing a wholesale price contract to each retailer. Each retailer then announces his retail price p_i and places his order quantity $q_i(\mathbf{p})$ from the supplier who fulfills all the orders without delay.

Assumption 2.2.1 *The demand function q_i is a continuous, twice differentiable function with respect to prices, $\mathbf{p}$. Furthermore, we assume that $\frac{\partial q_i}{\partial p_i} < 0$ for all i, and $\frac{\partial q_i}{\partial p_j} \geq 0$ for $j \neq i$.*

This assumption states that the demand decreases strictly if the retailer increases his price and increases if his competitors increase their prices. Denote the Jacobian matrix of the demand function as $-\mathbf{B}(\mathbf{p})$ such that,

$$\mathbf{B}(\mathbf{p}) = -\begin{bmatrix} \frac{\partial q_1}{\partial p_1} & \cdots & \frac{\partial q_1}{\partial p_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial q_n}{\partial p_1} & \cdots & \frac{\partial q_n}{\partial p_n} \end{bmatrix},$$

where $\mathbf{B}(\mathbf{p})$ is a $n \times n$ matrix. For ease of notation, we will use $\mathbf{B}$ instead of $\mathbf{B}(\mathbf{p})$ for the rest of the chapter. We will utilize subscripts to distinguish this matrix evaluated at different values of $\mathbf{p}$. When demand is affine, then $\mathbf{B}$ is independent of $\mathbf{p}$, and $\mathbf{q}(\mathbf{p}) = \mathbf{q}(\mathbf{0}) - \mathbf{B}\mathbf{p}$, where $\mathbf{q}(\mathbf{0})$ is the maximum demand when products are free. Denote β_{ij} as the (i,j)th element of matrix $\mathbf{B}$, which represents the demand sensitivity with respect to a price change. Assumption 2.2.1 states that matrix $\mathbf{B}$ has positive diagonal elements and nonnegative off-diagonal elements. Denote matrix $\mathbf{\Gamma}$ as the diagonal matrix of $\mathbf{B}$, i.e., $\mathbf{\Gamma} = diag(\beta_{ii}, \ldots, \beta_{nn})$.

Assumption 2.2.2 *Matrix* $\mathbf{B}$ *is symmetric and strictly diagonally dominant for all feasible price vectors* $\mathbf{p}$.

The symmetry of the Jacobian matrix of demand naturally arises from a representative consumer utility framework with a concave utility function. Furthermore, strict diagonal dominance of the matrix implies that a retailer's demand is more sensitive to his own price changes than to those of his competitors. This assumption follows from the law of demand for substitutable products where a price increase by any retailer leads to a decrease of the total sales in the market.

Assumption 2.2.1 and Definition 2.2.2 imply that the Jacobian matrix $\mathbf{B}$ belongs to the class of M-matrices, also referred to as Stieltjes matrices. This class of matrices has several interesting properties, e.g., positive definite and its inverse is componentwise nonnegative. We refer the reader to Horn and Johnson (1985) for more information on this topic.

Assumption 2.2.3 *The demand function* q_i *is a concave function of the price vector* $\mathbf{p}$. *Alternatively, the demand function* q_i *is a nonnegative, convex function of the prices and the individual profit for retailer* i $p_i q_i(p_i, \mathbf{p_{-i}})$ *is a quasi-concave function of the price* p_i.

This assumption guarantees the existence and uniqueness of the solution in both the decentralized and the centralized settings. It is possible to relax it to some extent, but it involves more tedious derivations. As a result, we impose this assumption to enhance the transparency of the model.

Assumption 2.2.4 $\mathbf{q}(\mathbf{c}) > \mathbf{0}$.

This assumption states that the demand must be positive when products are priced at cost. Products which do not satisfy this assumption are unprofitable and are expected to be removed from the market by the profit-seeking retailers. With affine demand, this assumption becomes $\mathbf{q}(\mathbf{c}) = \mathbf{q}(\mathbf{0}) - \mathbf{B}\mathbf{c} > \mathbf{0}$.

Assumption 2.2.5 $\mathbf{\Gamma}(\mathbf{\Gamma} + 2\mathbf{B})^{-1}\mathbf{B}(\mathbf{p} - \mathbf{c}) \geq \mathbf{0}$

This assumption implies that when the items are priced at the equilibrium retail prices, the expected demands are positive. For the case with symmetric retailers and affine demand, this assumption can be shown to be equivalent to Assumption 2.2.4. Adida and DeMiguel (2010) have imposed a similar assumption in their work with affine demand.

To facilitate the analysis with imperfect competition, we consider the following measure which captures the relative dependence of each retailer with respect to other competitors in the market.

Definition 2.2.6 ***Competition index***. *Given a price sensitivity matrix* $\mathbf{B}$*, we define* r_i *for retailer* i *as follows,* $r_i(\cdot) = \sum_{j \neq i} |\beta_{ji}|/\beta_{ii}$.

By Assumption 2.2.2, $0 \leq r_i(\mathbf{B}) < 1$ for retailer $i = 1, \ldots, n$. This index is known as *diversion ratio* in the economics literature (e.g., Bordley 1985), that is, the fraction of retailer i's customers who switch to other retailers when retailer i raises his price by 1 unit, given all other retailers keep their prices unchanged. This index is positively related to competition and negatively related to product differentiation. For instance, when retailers are noncompeting (e.g., the products are unrelated), i.e., $r_i(\mathbf{B}) = 0$. It means that if retailer i increases his price, there is no change in other retailers' demand. Another extreme case is when $r_i(\mathbf{B})$ approaches 1. That is, customers view retailer i's product and other products as fully substitutable. Thus, when retailer i raises his price, all his lost sales switch to other retailers.

With symmetry, all retailers share the same competition index, i.e., $r_i(\cdot) = \frac{n-1}{\beta}$ for all i (without the loss of generality, we set all $\beta_{ij} = -1$ for all $j \neq i$ and $\beta_{ii} = \beta$ for all i). When asymmetry exists, $r_i(\cdot)$ differs across retailers, which in turn, implies the relative strength of a retailer. A "strong" retailer is one whose demand is not affected much by external price disturbances, whereas a "weak" retailer is more susceptible to price changes. Suppose every retailer changes his price by ξ, the demand change experienced by retailer i selling substitutes is given by $\beta_{ii}(1 - r_i(\mathbf{B}))\xi$. High $r_i(\mathbf{B})$ results in a small demand change, implying a strong retailer.

Based on the competition index for every retailer, $r_i(\cdot)$, we introduce the following two indices to approximate a market with asymmetry.

Definition 2.2.7 *Given a market with n retailers whose competition indices are $r_1, \cdots, r_n$, we define*

(i) ***One-Firm Ratio.*** $r_{(1)}(\mathbf{B}) = \max_i r_i(\mathbf{B})$, *and*

(ii) ***Two-Firm Ratio.*** $r_{(2)}(\mathbf{B}) = \max_{i,j|j \neq i} \sqrt{r_i(\mathbf{B}) r_j(\mathbf{B})}$.

We are going to show that with either index, we are able to derive tight bounds on the channel performance. Though the One-Firm Ratio requires less information to compute (it only looks at one retailer whose product exhibits the highest competition index), the bounds in terms of the Two-Firm Ratio are more accurate as the ratio contains additional information on asymmetry. We can rewrite the Two-Firm Ratio as $r_{(2)} = \max_{i,j|j \neq i} \frac{r_i + r_j}{2} - \frac{(\sqrt{r_i} - \sqrt{r_j})^2}{2}$, which is essentially a geometric mean of the two retailers with the highest competition indices and a correction term which is determined by the asymmetry between them.

When the demand $\mathbf{q}(\mathbf{p})$ is nonlinear, the Jacobian matrix $\mathbf{B}$ depends on the value of $\mathbf{p}$. To establish a bound, it is key to introduce a constant that will measure the curvature, or the degree of nonlinearity of the demand function. As a result, we briefly introduce the concept of Jacobian similarity. We refer reader to Perakis (2007) for more information on this concept.

Definition 2.2.8 ***The Jacobian similarity property.*** *A positive semidefinite matrix $\mathbf{F}(\mathbf{p})$ satisfies the matrix similarity property if there exists a constant $\kappa \geq 1$ such that for all $\mathbf{w}, \mathbf{p}$ and $\mathbf{p}'$: $\kappa \mathbf{w}^T(\mathbf{F}(\mathbf{p}))\mathbf{w} \geq \mathbf{w}^T(\mathbf{F}(\mathbf{p}'))\mathbf{w} \geq \frac{1}{\kappa}\mathbf{w}^T(\mathbf{F}(\mathbf{p}))\mathbf{w}$.*

Notice that if $\mathbf{F}(\mathbf{p})$ is the Jacobian matrix of an affine function, i.e., $\mathbf{F}(\mathbf{p})$ does not depend on $\mathbf{p}$, then $\kappa = 1$. In general, the constant κ is easy to compute when matrix $\mathbf{F}(\mathbf{p})$ is positive definite for all $\mathbf{p}$. In that case, one choice for the constant $\kappa = \frac{\max_{\mathbf{p}} \max_i \lambda_i(\mathbf{F}(\mathbf{p}))}{min_{\mathbf{p}} \min_i \lambda_i(\mathbf{F}(\mathbf{p}))}$, that is, the ratio between the maximum and minimum eigenvalue of the matrix.

2.2.2 The decentralized and centralized problems

Throughout the paper, we compare the performance of a decentralized (uncoordinated) supply chain to a benchmark setting of a centralized (coordinated) system. Denote the wholesale prices, retail prices, order quantities and chain-wide profit as $\mathbf{w}$, $\mathbf{p}$, $\mathbf{q}$ and π respectively. We use subscripts c and d to differentiate the centralized and the decentralized settings.

In a decentralized setting, the supplier maximizes her profit by deciding the wholesale prices, anticipating the equilibrium order quantities from the retailers. Each retailer determines his retail prices in order to maximize his own profit. We assume a sub-game Nash Equilibrium has been reached where no single retailer can increase his profit by unilaterally changing his price. For each retailer i, given the supplier's equilibrium wholesale price, w_i, and competitors' equilibrium price vector, $\mathbf{p_{-i}}$, retailer i's optimal retail price p_i is a best response function to the maximization problem:

$$(\pi_d)_{r_i} \triangleq \max \quad (p_i - w_i)q_i(p_i, \mathbf{p_{-i}}), \text{ s.t. } p_i \geq 0, \quad q_i(p_i, \mathbf{p_{-i}}) \geq 0.$$

The supplier maximizes her profit by deciding the wholesale price vector, $\mathbf{w}$, given the retailers' order quantities obtained from vector $\mathbf{q}(\mathbf{p}(\mathbf{w}))$:

$$(\pi_d)_s \triangleq \max \quad \sum_i (w_i - c_i) \cdot q_i(\mathbf{p}(\mathbf{w})), \text{ s.t. } \mathbf{w} \geq 0.$$

The total profit in the supply chain π_d is given by $\pi_d \triangleq (\pi_d)_s + \sum_i (\pi_d)_{r_i}$.

In a coordinated supply chain, a central authority decides the retail prices across the chain with the objective to maximize the chain-wide profit:

$$\pi_c \triangleq \max \quad \sum_i (p_i - c_i) q_i(\mathbf{p}), \text{ s.t. } \mathbf{p} \geq 0, \quad \mathbf{q}(\mathbf{p}) \geq 0.$$

Proposition 2.2.9 *Under Assumptions 2.2.1 to 2.2.5, there exists a unique equilibrium to the decentralized supply chain problem. The equilibrium output level and the total profit are given by*

$$\mathbf{q}_d = \mathbf{\Gamma}_d(\mathbf{\Gamma}_d + 2\mathbf{B}_d)^{-1}\mathbf{B}_d(\mathbf{p}_d - \mathbf{c}), \text{ and} \tag{2.1}$$

$$\pi_d = (\mathbf{p}_d - \mathbf{c})^T \mathbf{\Gamma}_d(\mathbf{\Gamma}_d + 2\mathbf{B}_d)^{-1}\mathbf{B}_d(\mathbf{p}_d - \mathbf{c}). \tag{2.2}$$

Under Assumptions 2.2.1 to 2.2.4, there exists a unique solution to the centralized supply chain problem. The corresponding output and profit are given by

$$\mathbf{q_c} = \mathbf{B}_c(\mathbf{p}_c - \mathbf{c}), \text{ and} \tag{2.3}$$

$$\pi_c = (\mathbf{p}_c - \mathbf{c})^T \mathbf{B}_c(\mathbf{p}_c - \mathbf{c}). \tag{2.4}$$

Proof of Proposition 2.2.9. We first derive conditions on the decision variables based on the optimality condition. For both the centralized and decentralized settings, we will show that there exists a unique optimal solution to the *unconstrained* problem. Next, we show that with the nonnegativity constraints, the solution is feasible. Thus, it is also the unique optimal solution to the original *constrained* problem.

The decentralized problem:

Consider the decentralized supply chain problem for each retailer i, given a wholesale price $\mathbf{w}$, in the equilibrium, the output level must satisfy $\mathbf{q}_d = \mathbf{\Gamma}_d(\mathbf{p}_d - \mathbf{w}_d)$, where $\mathbf{\Gamma}_d$ is the diagonal matrix with matrix $\mathbf{B}_d$. Substitute $\mathbf{w}_d = \mathbf{p}_d - \mathbf{\Gamma}_d^{-1}\mathbf{q}_d$ into the supplier's objective, i.e., $(\pi_d)_S = \mathbf{q}_d^T(\mathbf{p}_d - \mathbf{\Gamma}_d^{-1}\mathbf{q}_d - \mathbf{c})$. Taking the first order condition with respect to $\mathbf{p}_d$, and since $\mathbf{B}_d$ is symmetric by Assumption 2.2.2, we obtain the following condition that

$$\mathbf{q}_d - \mathbf{B}_d(\mathbf{p}_d - \mathbf{c}) + 2\mathbf{B}_d\mathbf{\Gamma}_d^{-1}\mathbf{q}_d = 0 \Rightarrow (\mathbf{\Gamma}_d + 2\mathbf{B}_d)\mathbf{\Gamma}_d^{-1}\mathbf{q}_d = \mathbf{B}_d(\mathbf{p}_d - \mathbf{c}).$$

Thus, we have shown that $\mathbf{q}_d = \mathbf{\Gamma}_d(\mathbf{\Gamma}_d + 2\mathbf{B}_d)^{-1}\mathbf{B}_d(\mathbf{p}_d - \mathbf{c})$, which is nonnegative by Assumption 2.2.5. To show that nonnegative constraints are satisfied, notice that $\mathbf{p}_d = \mathbf{c} + (\mathbf{\Gamma}_d(\mathbf{\Gamma}_d + 2\mathbf{B}_d)^{-1}\mathbf{B}_d)^{-1}\mathbf{q}_d = \mathbf{c} + (\mathbf{B}_d^{-1} + 2\mathbf{\Gamma}_d^{-1})\mathbf{q}_d$. By Assumption 2.2.2 that matrix $\mathbf{B}$ is a M-matrix, thus, its inverse is nonnegative. $\mathbf{\Gamma}^{-1}$ is a positive diagonal matrix. Therefore, $\mathbf{p}_d \geq \mathbf{c} \geq \mathbf{0}$. In addition, substituting $\mathbf{q}_d$ and $\mathbf{q}_d$, the wholesale price is shown to be $\mathbf{w}_d = \mathbf{c} + (\mathbf{B}_d^{-1} + \mathbf{\Gamma}_d^{-1})\mathbf{q}_d \geq \mathbf{c}$. Moreover, we have established $\mathbf{p}_d \geq \mathbf{w}_d \geq \mathbf{c}$, that is, "double marginalization" exists with nonlinear demand and asymmetric retailers in a decentralized supply chain. Lastly, the chain-wide profit in a decentralized supply chain is given by $\pi_d = (\mathbf{p}_d - \mathbf{c})^T\mathbf{q}_d = (\mathbf{p}_d - \mathbf{c})^T\mathbf{\Gamma}_d(\mathbf{\Gamma}_d + 2\mathbf{B}_d)^{-1}\mathbf{B}_d(\mathbf{p}_d - \mathbf{c})$.

The centralized problem:

The first order optimality condition for the centralized supply chain problem is given by $\mathbf{B}_c(\mathbf{p}_c - \mathbf{c}) = \mathbf{q}_c$. To show that $\mathbf{q}_c$ satisfy the nonnegativity constraints, notice that when the demand function $\mathbf{q}(\mathbf{p})$ is concave,

$$\mathbf{q}(\mathbf{0}) - \mathbf{q_c} \leq -\mathbf{B}_c(\mathbf{0} - \mathbf{p}_c) \Rightarrow \mathbf{q}(\mathbf{0}) - \mathbf{q}_c \leq \mathbf{B}_c\mathbf{p}_c.$$

This implies that $\mathbf{q}_c \geq \mathbf{q}(\mathbf{0}) - \mathbf{q}_c - \mathbf{B}_c\mathbf{c} \Rightarrow 2\mathbf{q}_c \geq \mathbf{q}(\mathbf{0}) - \mathbf{B}_c\mathbf{c} \Rightarrow \mathbf{q}_c \geq \frac{1}{2}(\mathbf{q}(\mathbf{0}) - \mathbf{B}_c\mathbf{c}) > 0$. We obtain the last inequality by Assumption 2.2.4. To show $\mathbf{p}_c \geq 0$, by the optimality condition, $\mathbf{p}_c - \mathbf{c} = \mathbf{B}_c^{-1}\mathbf{q}_c$. Because $\mathbf{B}_c^{-1}$ is nonnegative componentwise (by property of a M-matrix) and $\mathbf{q}_c > 0$ which we have just shown, $\mathbf{p}_c - \mathbf{c} > \mathbf{0} \Rightarrow \mathbf{p}_c > \mathbf{c}$.

On the other hand, suppose the demand $\mathbf{q}(\mathbf{p})$ is convex, then

$$\mathbf{q}(\mathbf{p}_c) - \mathbf{q}(\mathbf{0}) \geq -\mathbf{B}_0\mathbf{p}_c \Rightarrow (\mathbf{B}_c + \mathbf{B}_0)\mathbf{p}_c \geq \mathbf{q}(\mathbf{0}),$$

where $\mathbf{B}_0$ is the matrix evaluated at $\mathbf{q} = \mathbf{0}$. $\mathbf{B}_c + \mathbf{B}_0$ is a M-matrix and its inverse is nonnegative. Thus, with positive $\mathbf{q}(\mathbf{0})$, $\mathbf{p}_c \geq (\mathbf{B}_c + \mathbf{B}_0)^{-1}\mathbf{q}(\mathbf{0}) > 0$

Lastly, substituting the optimality condition of $\mathbf{q}_c$ into the objective, we obtain that the total profit is $\pi_c = \mathbf{q}_c^T(\mathbf{p}_c - \mathbf{c}) = (\mathbf{p}_c - \mathbf{c})^T \mathbf{B}_c (\mathbf{p}_c - \mathbf{c})$. □

The proof of Proposition also shows that with nonlinear demand and imperfect competition, the existence of "double marginalization" in a decentralized supply chain continue to hold, i.e., $\mathbf{p}_d > \mathbf{w}_d > \mathbf{c}$.

2.3 Efficiency with Affine Demand

In this section, we investigate the performance of a decentralized supply chain with imperfect competition when the demand function is affine. By doing so, we isolate the impact of imperfect competition from demand nonlinearity and we will address nonlinear demand in the next section. Note that with affine demand, the Jacobian matrix $\mathbf{B}$ is constant and independent of the values of $\mathbf{p}_c$ or $\mathbf{p}_d$.

Before we delve into the analysis, we would like to highlight that the impact of imperfect competition could be ambiguous at the first glance. For example, with affine demand, one can show that $\mathbf{p}_d \geq \mathbf{p}_c$. One might conjecture that every retailer in the decentralized chain sells few units due to the higher retail price. The following numerical example shows, however, that this relationship may fail to hold. Consider the following setup with 2 retailers, let demand $q_1(\mathbf{p}) = 0.7413 - p_1 + 0.8039p_2$, $q_2(\mathbf{p}) = 0.1048 - p_2 + 0.8039p_1$, $c_1 = c_2 = 0$. This results in $\mathbf{p}_d = \begin{bmatrix} 1.4006 \\ 1.1107 \end{bmatrix} \geq \mathbf{p}_c = \begin{bmatrix} 1.1670 \\ 0.9906 \end{bmatrix}$. However, $\mathbf{q}_d = \begin{bmatrix} 0.2336 \\ 0.1201 \end{bmatrix}$ and $\mathbf{q}_c = \begin{bmatrix} 0.3706 \\ 0.0524 \end{bmatrix}$. In this case, $(q_d)_2 > (q_c)_2$. In general, lower sales volume for every retailer, in the decentralized setting, can only be guaranteed in special cases such as noncompeting or symmetric retailers.

2.3.1 Lower and upper bounds on performance metrics

We investigate three aspects of performance, the first two address the chain-wide behavior from the channel members' perspective and the last focuses on the consumers and the society as a whole. We begin with *chain-wide profit*, defined as the fraction of possible profit that the decentralized setting attains compared to the centralized setting, i.e., π_d/π_c. Next, we consider *market penetration*, that is, the total sales in a given market. The deeper the penetration, the higher the sales volume. We are interested in finding out the fraction of sales volume captured in the decentralized setting, i.e., $\mathbf{e}^T\mathbf{q}_d/\mathbf{e}^T\mathbf{q}_c$, where $\mathbf{e} = (1, ..., 1) \in \mathbb{R}^n$. Next, The third performance metric analyzes the behavior of the supply chain from a societal perspective. The quantities of interest are *consumer surplus* (CS) and *total surplus* (TS) which is defined by aggregating consumer surplus and producer surplus which is the channel profit in this model.

Theorem 2.3.1 *When the demand function is affine with a price sensitivity matrix* $\mathbf{B}$*, the chain-wide profit, market penetration, consumer surplus and total surplus of a decentralized supply chain with imperfect competition are bounded as follows,*

$$\begin{aligned}
&\frac{3}{4} \leq \frac{\pi_d}{\pi_c} \leq \frac{3-2r_{(2)}(\mathbf{B})}{(2-r_{(2)}(\mathbf{B}))^2} \leq \frac{3-2r_{(1)}(\mathbf{B})}{(2-r_{(1)}(\mathbf{B}))^2},\\
&\frac{1}{2} \leq \frac{\mathbf{e}^T\mathbf{q}_d}{\mathbf{e}^T\mathbf{q}_c} \leq \frac{1}{2-r_{(2)}(\mathbf{B})} \leq \frac{1}{2-r_{(1)}(\mathbf{B})},\\
&\frac{1}{4} \leq \frac{CS_d}{CS_c} \leq \frac{1}{(2-r_{(2)}(\mathbf{B}))^2} \leq \frac{1}{(2-r_{(1)}(\mathbf{B}))^2}, \quad \textit{and}\\
&\frac{7}{12} \leq \frac{TS_d}{TS_c} \leq \frac{7-4r_{(2)}(\mathbf{B})}{3(2-r_{(2)}(\mathbf{B}))^2} \leq \frac{7-4r_{(1)}(\mathbf{B})}{3(2-r_{(1)}(\mathbf{B}))^2}.
\end{aligned}$$

The lower bounds are tight with noncompeting retailers and the upper bounds are tight with symmetric retailers.

Since the techniques used to prove bounds on the performance metrics are similar, we will present the proof on π_d/π_c below. The proofs for the other metrics can be found in the Appendix.

Proof of Theorem 2.3.1 on π_d/π_c. This proof consists of three main steps.

Step 1: Preliminary work. Using Equation (2.2) and (2.4) derived in Propsition 2.2.9, we have an expression for the ratio of the profit obtained in the two settings, i.e.,

$$\frac{\pi_d}{\pi_c} = \frac{(\mathbf{p}_d - \mathbf{c})^T \boldsymbol{\Gamma}(\boldsymbol{\Gamma} + 2\mathbf{B})^{-1}\mathbf{B}(\mathbf{p}_d - \mathbf{c})}{(\mathbf{p}_c - \mathbf{c})^T \mathbf{B}(\mathbf{p}_c - \mathbf{c})}. \tag{2.5}$$

We will now find an expression to express $\mathbf{q}_d$ in terms of $\mathbf{p}_c$. With affine demand, $\mathbf{q}(\mathbf{p}) = \mathbf{q}(\mathbf{0}) - \mathbf{B}\mathbf{p}$, i.e., $\mathbf{q}_c = \mathbf{q}(\mathbf{0}) - \mathbf{B}\mathbf{p}_c$ and $\mathbf{q}_d = \mathbf{q}(\mathbf{0}) - \mathbf{B}\mathbf{p}_d$. Thus, $\mathbf{q}_d + \mathbf{B}\mathbf{p}_d - \mathbf{B}\mathbf{c} = \mathbf{q}_c + \mathbf{B}\mathbf{p}_c - \mathbf{B}\mathbf{c}$. Use the optimality conditions shown in Equation (2.1) and (2.3), we obtain

$$\begin{aligned}
& (\boldsymbol{\Gamma}(\boldsymbol{\Gamma} + 2\mathbf{B})^{-1}\mathbf{B} + \mathbf{B})(\mathbf{p}_d - \mathbf{c}) = 2\mathbf{B}(\mathbf{p}_c - \mathbf{c}) \\
\Rightarrow \quad & (\boldsymbol{\Gamma} + \mathbf{B})(\boldsymbol{\Gamma} + 2\mathbf{B})^{-1}\mathbf{B}(\mathbf{p}_d - \mathbf{c}) = \mathbf{B}(\mathbf{p}_c - \mathbf{c}) \\
\Rightarrow \quad & \mathbf{p}_d - \mathbf{c} = \mathbf{B}^{-1}(\boldsymbol{\Gamma} + 2\mathbf{B})(\boldsymbol{\Gamma} + \mathbf{B})^{-1}\mathbf{B}(\mathbf{p}_c - \mathbf{c}).
\end{aligned}$$

Substituting Equation (2.6) into Equation (2.5) and with some algebra, it is easy to show the following:

$$\frac{\pi_d}{\pi_c} = \frac{(\mathbf{p}_c - \mathbf{c})^T (2\mathbf{B} + \boldsymbol{\Gamma})(\mathbf{B} + \boldsymbol{\Gamma})^{-1}\mathbf{B}(\mathbf{B} + \boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}(\mathbf{p}_c - \mathbf{c})}{(\mathbf{p}_c - \mathbf{c})^T \mathbf{B}(\mathbf{p}_c - \mathbf{c})}. \tag{2.6}$$

Step 2: Lower bound. To prove π_d/π_c is always greater than 3/4, it is equivalent to prove a composite matrix $\Phi(\mathbf{B}) = 4\mathbf{B}^{-1}(\mathbf{B}+\boldsymbol{\Gamma})\boldsymbol{\Gamma}^{-1}(2\mathbf{B}+\boldsymbol{\Gamma})(\mathbf{B}+\boldsymbol{\Gamma})^{-1} - 3\mathbf{B}^{-1}(\mathbf{B}+\boldsymbol{\Gamma})\boldsymbol{\Gamma}^{-1}\mathbf{B}\boldsymbol{\Gamma}^{-1}(\mathbf{B}+\boldsymbol{\Gamma})\mathbf{B}^{-1}$ is a copositive matrix. To see this, let $\mathbf{x} = \mathbf{B}(\mathbf{B}+\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}(\mathbf{p}_c - \mathbf{c})$, which is nonnegative under Assumption 2.2.5. Notice that if $\Phi(\mathbf{B})$ is copositive, by its definition, $\mathbf{x}^T\Phi(\mathbf{B})\mathbf{x} \geq 0$ must hold. It implies that

$$\begin{aligned}
& \mathbf{x}^T\Phi(\mathbf{B})\mathbf{x} \geq 0 \\
\Leftrightarrow & \left(\mathbf{B}(\mathbf{B} + \boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}(\mathbf{p}_c - \mathbf{c})\right)^T \Phi(\mathbf{B}) \left(\mathbf{B}(\mathbf{B} + \boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}(\mathbf{p}_c - \mathbf{c})\right) \geq 0 \\
\Leftrightarrow & 4(\mathbf{p}_c - \mathbf{c})^T(2\mathbf{B} + \boldsymbol{\Gamma})(\mathbf{B} + \boldsymbol{\Gamma})^{-1}\mathbf{B}(\mathbf{B} + \boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}(\mathbf{p}_c - \mathbf{c}) \geq 3(\mathbf{p}_c - \mathbf{c})^T\mathbf{B}(\mathbf{p}_c - \mathbf{c}) \\
\Leftrightarrow & \frac{\frac{1}{4}(\mathbf{p}_c - \mathbf{c})^T(2\mathbf{B} + \boldsymbol{\Gamma})(\mathbf{B} + \boldsymbol{\Gamma})^{-1}\mathbf{B}(\mathbf{B} + \boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}(\mathbf{p}_c - \mathbf{c})}{\frac{1}{4}(\mathbf{p}_c - \mathbf{c})^T\mathbf{B}(\mathbf{p}_c - \mathbf{c})} \geq \frac{3}{4},
\end{aligned}$$

where the right-hand-side is the same expression as what we have derived for π_d/π_c

in Equation (2.6).

Denote $\mathbf{G} = \mathbf{\Gamma}^{-0.5}\mathbf{B}\mathbf{\Gamma}^{-0.5}$, whose diagonal elements are normalized to 1 and its $(i,j)^{th}$ element is defined as $\beta_{ij}/\sqrt{\beta_{ii}\beta_{jj}}$. It is straightforward to show that $\Phi(\mathbf{B}) = \mathbf{\Gamma}^{1/2}\Phi(\mathbf{G})\mathbf{\Gamma}^{1/2}$. Note that $\mathbf{\Gamma}$ is a diagonal matrix with positive elements, thus if we can show that $\Phi(\mathbf{G})$ is copositive, then $\Phi(\mathbf{B})$ must also be copositive. To prove this,

$$\begin{aligned}\Phi(\mathbf{G}) &= 4\mathbf{G}^{-1}(\mathbf{G}+\mathbf{I})(2\mathbf{G}+\mathbf{I})(\mathbf{G}+\mathbf{I})^{-1} - 3\mathbf{G}^{-1}(\mathbf{G}+\mathbf{I})\mathbf{G}(\mathbf{G}+\mathbf{I})\mathbf{G}^{-1} \\ &= 4\mathbf{G}^{-1}(\mathbf{G}+\mathbf{I}) + 4\mathbf{G}^{-1}(\mathbf{G}+\mathbf{I})\mathbf{G}(\mathbf{G}+\mathbf{I})^{-1} - 3\mathbf{G}^{-1}(\mathbf{G}+\mathbf{I})^2 \\ &= 4\mathbf{I} + 4\mathbf{G}^{-1} + 4\mathbf{I} - 3\mathbf{G}^{-1}(\mathbf{G}^2 + 2\mathbf{G} + \mathbf{I}) \\ &= 2\mathbf{I} + \mathbf{G}^{-1} - 3\mathbf{G} \\ &= (\mathbf{G}^{-1} + 3\mathbf{I})(\mathbf{I} - \mathbf{G}).\end{aligned}$$

Because $\mathbf{B}$ is a M-matrix, $\mathbf{G}$ is also a M-matrix. The first term is the sum of an inverse M-matrix $\mathbf{G}^{-1}$ and an identity matrix, therefore, it is also nonnegative. Also, $\mathbf{G}$ has diagonals equal to 1 and nonpositive off-diagonals, thus, $\mathbf{I} - \mathbf{G}$ is also nonnegative. $\Phi(\mathbf{G})$ which is a product of two nonnegative matrices is also nonnegative. A nonnegative matrix is copositive and this establishes 3/4 as a lower bound.

To show that this bound is tight, consider the case with noncompeting retailers, i.e., $\beta_{ij} = 0$ for all $j \neq i$. Thus, the sensitivity matrix $\mathbf{B}$ becomes a diagonal matrix, i.e., $\mathbf{\Gamma} = \mathbf{B}$. The profit expression in Equation (2.6) becomes,

$$\frac{\pi_d}{\pi_c} = \frac{(\mathbf{p}_c - \mathbf{c})^T(3\mathbf{\Gamma})(2\mathbf{\Gamma})^{-2}\mathbf{\Gamma}^2(\mathbf{p}_c - \mathbf{c})}{(\mathbf{p}_c - \mathbf{c})^T\mathbf{\Gamma}(\mathbf{p}_c - \mathbf{c})} = \frac{3}{4}.$$

This has shown that the lower bound is tight.

Step 3: Upper bound. Denote $\mathbf{w} = \mathbf{B}^{1/2}(\mathbf{p}_c - \mathbf{c})$. Rewrite the ratio $\pi_d/\ \pi_c$ in Equation (2.6) as follows,

$$\begin{aligned}\frac{\pi_d}{\pi_c} &= \frac{\mathbf{w}^T\mathbf{B}^{-1/2}(2\mathbf{B}+\mathbf{\Gamma})(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma}\mathbf{B}^{-1/2}\mathbf{w}}{\mathbf{w}^T\mathbf{w}} \\ &= \frac{\mathbf{w}^T\mathbf{B}^{-1/2}(\mathbf{I}+\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1})\mathbf{B}(\mathbf{I}-(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{B})\mathbf{B}^{-1/2}\mathbf{w}}{\mathbf{w}^T\mathbf{w}}\end{aligned}$$

$$
\begin{aligned}
&= \frac{\mathbf{w}^T\mathbf{B}^{-1/2}(\mathbf{B} - \mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{B})\mathbf{B}^{-1/2}\mathbf{w}}{\mathbf{w}^T\mathbf{w}} \\
&= \frac{\mathbf{w}^T(\mathbf{I} - \mathbf{B}^{1/2}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{B}^{1/2}\mathbf{B}^{1/2}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{B}^{1/2}\mathbf{w}}{\mathbf{w}^T\mathbf{w}} \\
&= \frac{\mathbf{w}^T(\mathbf{I} - \mathbf{A}^2)\mathbf{w}}{\mathbf{w}^T\mathbf{w}},
\end{aligned}
$$

where $\mathbf{A} = \mathbf{B}^{1/2}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{B}^{1/2}$. Note that matrix $\mathbf{I} - \mathbf{A^2}$ is symmetric and can be diagonalized with a unitary matrix which corresponds to the eigenvectors. Using the Rayleigh-Ritz Theorem, the ratio π_d/π_c must be bounded by the maximum eigenvalue, that is,

$$
\begin{aligned}
\frac{\pi_d}{\pi_c} &\leq \lambda_{\max}(\mathbf{I} - \mathbf{A}^2) \\
&= 1 - \lambda_{\min}(\mathbf{A}^2) \\
&= 1 - \lambda_{\min}^2(\mathbf{A}).
\end{aligned} \tag{2.7}
$$

The last equality holds because matrix $\mathbf{A}$ is positive definite.

$$
\begin{aligned}
\lambda_{\min}(\mathbf{A}) &= \lambda_{\min}(\mathbf{B}^{1/2}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{B}^{1/2}) \\
&= \frac{1}{\lambda_{\max}(\mathbf{B}^{-1/2}(\mathbf{B}+\mathbf{\Gamma})\mathbf{B}^{-1/2})} \\
&= \frac{1}{\lambda_{\max}(\mathbf{I} + \mathbf{B}^{-1/2}\mathbf{\Gamma}\mathbf{B}^{-1/2})} \\
&= \frac{1}{1 + \lambda_{\max}(\mathbf{B}^{-1/2}\mathbf{\Gamma}\mathbf{B}^{-1/2})}.
\end{aligned}
$$

Since $\mathbf{B}^{-1/2}\mathbf{\Gamma}\mathbf{B}^{-1/2} = \mathbf{B}^{-1/2}(\mathbf{\Gamma}\mathbf{B}^{-1})\mathbf{B}^{1/2}$ and $\mathbf{\Gamma}\mathbf{B}^{-1} = \mathbf{\Gamma}^{1/2}(\mathbf{\Gamma}^{1/2}\mathbf{B}^{-1}\mathbf{\Gamma}^{1/2})\mathbf{\Gamma}^{-1/2}$, which implies that $\mathbf{B}^{-1/2}\mathbf{\Gamma}\mathbf{B}^{-1/2}$, $\mathbf{\Gamma}\mathbf{B}^{-1}$ and $\mathbf{\Gamma}^{1/2}\mathbf{B}^{-1}\mathbf{\Gamma}^{1/2}$ are similar matrices, i.e., they all have the same eigenvalues.

$$
\begin{aligned}
\lambda_{\min}(\mathbf{A}) &= \frac{1}{1 + \lambda_{\max}(\mathbf{\Gamma}^{1/2}\mathbf{B}^{-1}\mathbf{\Gamma}^{1/2})} \\
&= \frac{1}{1 + \frac{1}{\lambda_{\min}(\mathbf{\Gamma}^{-1/2}\mathbf{B}\mathbf{\Gamma}^{-1/2})}} \\
&= \frac{1}{1 + \frac{1}{\lambda_{\min}(\mathbf{G})}}, \qquad \text{where } \mathbf{G} = \mathbf{\Gamma}^{-1/2}\mathbf{B}\mathbf{\Gamma}^{-1/2}
\end{aligned}
$$

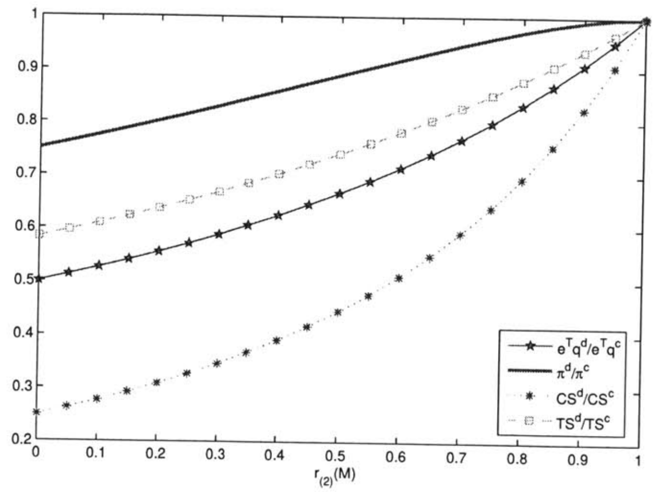

Figure 2-1: The upper bounds for the performance metrics (which are tight for symmetric retailers).

$$= \frac{\lambda_{\min}(\mathbf{G})}{1+\lambda_{\min}(\mathbf{G})}.$$

Substitute it into Equation (2.7) and we obtain the upper bound in terms of the minimum eigenvalue of $\mathbf{G}$ as shown below.

$$\begin{aligned} \frac{\pi_d}{\pi_c} &\leq 1 - \left(\frac{\lambda_{\min}(\mathbf{G})}{1+\lambda_{\min}(\mathbf{G})}\right)^2 \\ &= \frac{1+2\lambda_{\min}(\mathbf{G})}{(1+\lambda_{\min}(\mathbf{G}))^2}. \end{aligned} \tag{2.8}$$

Because the inequality (2.8) is decreasing in $\lambda_{\min}(\mathbf{G})$, we can further bound the ratio by lower bounding $\lambda_{\min}(\mathbf{G})$. Substitute the inequalities in Lemma A.2.1, we obtain the corresponding upper bounds on the chain-wide profits,

$$\begin{aligned} &\frac{\pi_d}{\pi_c} \leq \frac{1+2\lambda_{\min}(\mathbf{G})}{(1+\lambda_{\min}(\mathbf{G}))^2} \leq \frac{1+2(1-r_{(2)}(\mathbf{B}))}{(1+1-r_{(2)}(\mathbf{B}))^2} \leq \frac{1+2(1-r_{(1)}(\mathbf{B}))}{(1+1-r_{(1)}(\mathbf{B}))^2} \\ \Rightarrow\ &\frac{\pi_d}{\pi_c} \leq \frac{1+2\lambda_{\min}(\mathbf{G})}{(1+\lambda_{\min}(\mathbf{G}))^2} \leq \frac{3-2r_{(2)}(\mathbf{B}))}{(2-r_{(2)}(\mathbf{B}))^2} \leq \frac{3-2r_{(1)}(\mathbf{B}))}{(2+r_{(1)}(\mathbf{B}))^2}. \end{aligned}$$

□

Because $r_{(2)}(\mathbf{B}) \leq 1$, all four performance metrics considered are always lower in the decentralized supply chain than that of a centralized setting. Nonetheless, Theorem 2.3.1 states that as the competition index increases, the channel performance improves. In particular, the worst case for the decentralized setting occurs in the absence of horizontal competition (noncompeting retailers), i.e., $r_{(2)}(\mathbf{B}) = 0$. Note that as $r_{(2)}(\mathbf{B}) \to 1$, the performance of the decentralized setting approaches that of a centralized setting. Figure 2-1 depicts graphically how the performance metrics vary with competition index, $r_{(2)}(\mathbf{B})$.

Prior work on supply chains with symmetric retailers (see for example, Adida and DeMiguel 2010 and Cachon and Lariviere 2005) concluded that efficiency improves as the number of symmetric retailers increases. Intuitively, as the number of retailers grows to infinity, the retailers become price-takers with zero profit margin, and hence it eliminates the double marginalization effect. We extend the analysis to asymmetric retailers. We show that the level of competition in the retail market depends on the number of retailers *and* their relative price elasticities. Therefore, it is possible for a decentralized chain with a small number of retailers to also achieve high efficiency, if their products are highly substitutable. Moreover, the bounds in Theorem 2.3.1 also suggest that the performance of a decentralized supply chain predominantly depends on the two "strongest" retailers, whose products exhibit the highest levels of competition.

To explain this phenomenon, consider the comparison between the marginal gain of an individual retailer's profit in the decentralized setting $(\partial(\pi_d)_{r_i}/\partial p_i)$ and the marginal gain of the chain-wide profit in a centralized chain $(\partial \pi_c/\partial p_i)$:

$$\frac{\partial(\pi_{r_i})_d}{\partial p_i} = \frac{\partial \pi_c}{\partial p_i} + \overbrace{(w_i - c_i)\beta_{ii}}^{\text{vertical externality}} + \overbrace{\sum_{j \neq i}(p_j - c_j)\beta_{ji}}^{\text{horizontal externality}} . \tag{2.9}$$

As highlighted in Equation (2.9), an individual retailer in the decentralized setting faces two types of externalities: The vertical externality occurs when the supplier charges a price higher than her marginal cost, while the horizontal externality results from the cross-elasticity effect of the demand.

As $\beta_{ii} > 0$ for all i, the vertical externality is always positive, i.e., indicating

Source	Industry	Products	$r_{(2)}(\mathbf{B})$	π_d/π_c
Chintagunta et al. (2002)	Frozen pasta	5	0.45	87%
Ellison et al. (1997)	Pharmaceutical products	3	0.750	96%
Mela et al. (1998)	Cleaning supplies	9	0.875	98.7%

Table 2.1: Illustrative empirical results

that the retailer has an incentive to raise his price so as to increase his profit. The horizontal externality can be either positive or negative, depending on the type of products that the retailers are competing with. To be specific, when retailers are competing with substitutes as considered in this work, $\beta_{ji} \leq 0$, for all $j \neq i$. It means that the horizontal externality is negative, indicating a downward pressure to *offset* the vertical externality as the two externalities act in opposite directions. In other words, when products are highly substitutable, i.e., the intensity of horizontal competition is high, the retailers have to undercut prices to attract sales. As a result, it reduces the gap of double marginalization which in turn, promotes the channel efficiency. When the horizontal externality is sufficiently large to balance the vertical externality, the decentralized setting behaves as a centralized chain.

We conclude this subsection with three examples drawn from the literature that we use to estimate the chain-wide profit ratio for three different industries. The results are summarized in Table 2.1. The cited papers provide empirical estimates of price sensitivities of competing substitutes. We symmetrize the matrix by using $(\mathbf{B} + \mathbf{B}')/2$. We see that for all three examples, the performance of the decentralized supply chain is comparable to the centralized setting. It implies that in a supply chain with substitutable products, price-only contracts are quite "efficient". Thus, the room for benefits from using more elaborate contracts, which are often costly to implement, could be limited. This may partly explain the observation by Cachon (2003) that price-only contracts are commonly used in practice despite a proliferation of complex coordinating contracts proposed in the academic literature.

2.3.2 Rent extraction in decentralized supply chains

As with all supply chain structures, the retailers and the supplier are indirectly interested in the aggregate performance of the supply chain (e.g., chain-wide profit) and more directly interested in their own share of that profit. In this section, we analyze rent extraction of the individual channel members in a decentralized chain, that is, the profit allocation between the supplier and all the retailers. We use $(\pi_d)_S$ and $(\pi_d)_R$ to denote the profits earned by the supplier and all the retailers in a decentralized setting, where $(\pi_d)_R = \sum_{i=1}^{n}(\pi_d)_{r_i}$.

Proposition 2.3.2 *In a decentralized supply chain, the profit allocation between the retailers and the supplier is bounded as follows,*

$$\frac{1}{2} \geq \frac{(\pi_d)_R}{(\pi_d)_S} \geq \frac{1 - r_{(2)}(\mathbf{B})}{2 - r_{(2)}(\mathbf{B})} \geq \frac{1 - r_{(1)}(\mathbf{B})}{2 - r_{(1)}(\mathbf{B})},$$

where the lower bound is tight with symmetric retailers and the upper bound is tight with noncompeting retailers.

Price-only contracts guarantee the supplier with a higher profit share. Being a leader in a Stackelberg game, the supplier has the advantage of selecting the most favorable contractual terms by anticipating the retailers' responses. When retailers are independent (i.e., $r_{(2)}(\mathbf{B}) = 0$), the supplier's profit exactly doubles that of all the retailers.

The supplier also benefits from horizontal competition as her lion share of the profit grows further. We observe that the supplier's profit increases while the retailers' profits decrease with competition index. This result is consistent with our previous analysis: Intense competition induced by high level of product substitution essentially transfers the market power from the retailers to the supplier. For the case of symmetric retailers, in the limit as $r_{(2)}(\mathbf{B}) \to 1$, retailers become perfect price-takers, leaving the entire profit share to the supplier. Based on the analysis, we see that price-only contracts are desirable from the supplier's perspective, which offers another potential explanation to their popularity in practice.

2.3.3 Impact of retailer asymmetry

So far, we have developed several bounds that measure the performance of a decentralized supply chain relative to a centralized setting. We showed that the bounds are tight for special cases such as noncompeting or symmetric retailers. In this section, we will study the impact of asymmetry among retailers on the channel performance. As the behavior of the bounds is quite similar for all three performance metrics, we will use the chain-wide profit ratio, i.e., π_d/π_c, as an example.

Consider the following supply chain setting with n retailers:

$$\mathbf{c} = \begin{bmatrix} c \\ \vdots \\ c \end{bmatrix}, \quad \mathbf{q(0)} = \begin{bmatrix} q(0) \\ \vdots \\ q(0) \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} n & -1 & \dots & -1 \\ -1 & nk & & -1 \\ \vdots & \ddots & \ddots & \vdots \\ -1 & \dots & -1 & nk \end{bmatrix}.$$

Let k denote the asymmetry factor. The competition index is given by $r_1(\mathbf{B}) = \frac{n-1}{n} = \delta$ and $r_2(\mathbf{B}) = \dots = r_n(\mathbf{B}) = \frac{n-1}{nk} = \delta/k$. When $k = 1$, all the retailers are symmetric. Figure 2-2 shows a plot of π_d/π_c with respect to k for a setting with $n = 3$ retailers.

Proposition 2.3.3 *The channel performance decreases with the asymmetry factor k.*

As the asymmetry factor k increases, $r_i(\mathbf{B})$ decreases for all retailers except $i = 1$. Decreasing $r_i(\mathbf{B})$ implies weakening of retailer i or reduced competition intensity. We have shown in Theorem 2.3.1 that the channel performance is heavily dependent on the two "strongest" retailers in the market for substitutes, i.e., the "stronger" they are (or equivalently, more intense the competition is), the better the overall channel performance. As a result, as k increases, the asymmetry between the two "strongest" retailers increases (i.e., one of them becomes "weaker"), the channel performance deteriorates as shown in Figure 2-2.

Recall that the Two-Firm Ratio can be expressed as a geometric mean (see Definition 2.2.7) which captures the asymmetry between the two retailers with the highest

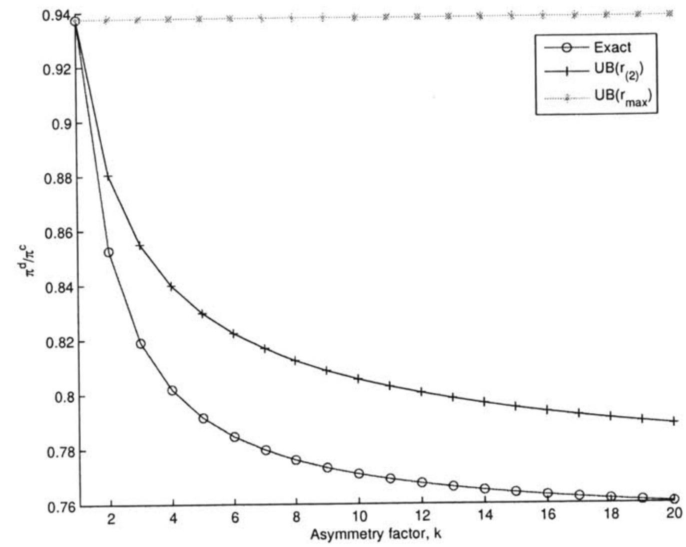

Figure 2-2: The plot shows the exact π_d/π_c and two upper bounds in terms of the Two-Firm Ratio $r_{(2)}(\mathbf{B})$, and the One-Firm ratio $r_{(1)}(\mathbf{B})$, with respect to the asymmetry factor k.

competition indices. For a fixed mean, as the asymmetry increases, $r_{(2)}$ decreases. Since π_d/π_c increases in $r_{(2)}$, the asymmetry between the two "representative" retailers has a reverse impact on the channel performance, i.e., π_d/π_c decreases in asymmetry. In other words, more imperfect the competition induces a more inefficient decentralized supply chain.

The analysis also highlights the benefit of using the Two-Firm ratio, especially in settings where there is a significant difference the two retailers with the highest competition indices (e.g., Walmart versus local grocery stores). For the given setting, the One-Firm ratio, $r_{(1)}(\mathbf{B}) = \delta$, is a constant, while the Two-Firm ratio, $r_{(2)}(\mathbf{B}) = \delta/\sqrt{k}$, decreases in k. Essentially, the One-Firm ratio estimates the performance of the entire supply chain based on a single retailer. As shown in Figure 2-2, the errors from using the bound in terms of $r_{(1)}(\mathbf{B})$ grow rapidly since $r_{(1)}(\mathbf{B})$ is independent of k and ignores retailer asymmetry. For instance, when $k = 20$, the exact ratio of π_d/π_c for complement is 0.71, whereas the bound predicted by $r_{(1)}(\mathbf{B})$ stays at 0.44. In contrast, the bound in terms of $r_{(2)}(\mathbf{B})$ estimates the exact value with errors around 5%, since $r_{(2)}(\mathbf{B})$ captures the asymmetry between the two "representative" retailers

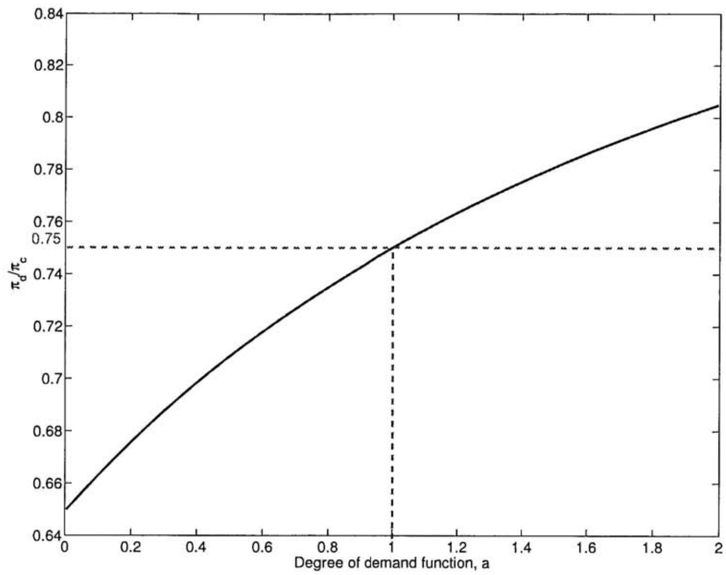

Figure 2-3: The impact of nonlinearity: π_d/π_c with respect to the degree a, where the demand function is given by $p(q) = p(0) - \beta p^a$.

who predominantly determine the channel performance and provides a more accurate estimation. We will present a more comprehensive computational experiment to compare the tightness of these two bounds in Section 2.5.

2.4 Efficiency with Nonlinear Demand

In this section, we will investigate the performance when the demand is nonlinear. We focus on the total profit as we have seen in the earlier section that the other metrics such as market penetration, social welfare and consumer surplus behave very similarly. We will begin with an example. Consider the following supply chain with a single retailer who faces a demand function: $p(q) = p(0) - \beta p^a$. This function can be convex, affine and concave, depending on the value of degree a. To be precise, when $0 < a < 1$, $p(q)$ is convex; $a = 1$, $p(q)$ is affine; when $a > 1$, the function is strictly concave. Figure 2-3 shows the value of π_d/π_c with respect to the degree a.

Figure 2-3 shows that the efficiency of the decentralized supply chain increases with a. In particular, when $a = 1$ (affine demand), the total profit of a decentralized supply with a monopoly is exactly 3/4 of the optimal profit. When $a < 1$ (convex demand), the efficiency is lower than 3/4. Contrast this with $a > 1$ (concave demand), the efficiency increases above 3/4.

We have already shown in Theorem 2.3.1 that with affine demand and noncompeting retailers, efficiency of a decentralized supply chain in terms of profit is exactly 3/4. Based on the example as shown in Figure 2-3, compared to affine demand, efficiency of a supply chain is higher with concave demand and lower with convex demand. In the following two subsections, we will generalize these results on nonlinear demand and derive upper and lower bounds on the profit ratio when is also imperfect competition.

With nonlinear demand function, the Jacobian matrix of the demand depends on the value of $\mathbf{p}$. We will use $\mathbf{B}_d$ and $\mathbf{B}_c$ to differentiate this matrix evaluated at $\mathbf{p}_d$ and $\mathbf{p}_c$ respectively. In particular, $r_{(1)}(\mathbf{B}_c)$ and $r_{(2)}(\mathbf{B}_c)$ refer to the One-Firm and Two-Firm ratio with respect to the Jacobian matrix $\mathbf{B}_c$.

2.4.1 Concave demand

When the demand function is concave, for a given price $\bar{\mathbf{p}}$, the demand $\mathbf{q}(\bar{\mathbf{p}}) \geq \mathbf{q}(\mathbf{0}) - \bar{\mathbf{B}}\bar{\mathbf{p}}$ where $\bar{\mathbf{B}}$ is the Jacobian matrix evaluated at $\bar{\mathbf{p}}$. To quantify the performance, we will make use of the Jacobian similarity property of a nonlinear function (see Definition 2.2.8).

Theorem 2.4.1 *When demand function $\mathbf{q}(\mathbf{p})$ is concave with $\mathbf{p}$, the efficiency of a decentralize supply chain is higher than that with affine demand. In particular, the total profit in a decentralized supply chain is bounded between*

$$\frac{3}{4} \leq \frac{\pi_d}{\pi_c} \leq \kappa_1 \frac{3 - r_{(2)}(\mathbf{B}_c)}{(2 - r_{(2)}(\mathbf{B}_c))^2} \leq \kappa_1 \frac{3 - r_{(1)}(\mathbf{B}_c)}{(2 - r_{(1)}(\mathbf{B}_c))^2},$$

where $\kappa_1 \geq 1$ is the Jacobian similarity factor.

Proof of Theorem 2.4.1. Without loss of generality, we will set the constant marginal cost $\mathbf{c} = \mathbf{0}$ in this proof.

Lower bound: By concavity of the demand function,

$$\mathbf{q}_c - \mathbf{q}_d \leq -\mathbf{B}_d(\mathbf{p}_c - \mathbf{p}_d)$$
$$\Rightarrow \mathbf{q}_c + \mathbf{B}_d\mathbf{p}_c \leq \mathbf{q}_d + \mathbf{B}_d\mathbf{p}_d.$$

Substituting the equilibrium condition derived in Proposition 2.2.9, we obtain

$$(\mathbf{B}_c + \mathbf{B}_d)\mathbf{p}_c \leq (\mathbf{\Gamma}_d(\mathbf{\Gamma}_d + 2\mathbf{B}_d)^{-1}\mathbf{B}_d + \mathbf{B}_d)\mathbf{p}_d$$
$$= 2(\mathbf{\Gamma}_d + \mathbf{B}_d)(\mathbf{\Gamma}_d + 2\mathbf{B}_d)^{-1}\mathbf{B}_d\mathbf{p}_d.$$

Since $\mathbf{B}_c + \mathbf{B}_d$ is a M-matrix by Assumption 2.2.2, $(\mathbf{B}_c + \mathbf{B}_d)^{-1}$ is nonnegative and we obtain the inequality,

$$\mathbf{p}_c \leq 2(\mathbf{B}_c + \mathbf{B}_d)^{-1} \underbrace{(\mathbf{\Gamma}_d + \mathbf{B}_d)(\mathbf{\Gamma}_d + 2\mathbf{B}_d)^{-1}\mathbf{B}_d}_{\mathbf{\Phi}_d} \mathbf{p}_d \tag{2.10}$$

One upper bound for $\pi_c = \mathbf{p}_c^T\mathbf{B}_c\mathbf{p}_c$ can be obtained by using Equation (2.10), i.e.,

$$\pi_c \leq 4\mathbf{p}_d^T\mathbf{\Phi}_d(\mathbf{B}_c + \mathbf{B}_d)^{-1}\mathbf{B}_c(\mathbf{B}_c + \mathbf{B}_d)^{-1}\mathbf{\Phi}_d\mathbf{p}_d$$

We will now multiply $(2\mathbf{B}_d)(2\mathbf{B}_d)^{-1} = \mathbf{I}$ and $\mathbf{B}_d\mathbf{B}_d^{-1} = \mathbf{I}$ to the inequality,

$$\pi_d \leq 4\mathbf{p}_d^T\mathbf{\Phi}_d(\mathbf{B}_c + \mathbf{B}_d)^{-1}(2\mathbf{B}_d)^{-1}(2\mathbf{B}_d)\mathbf{B}_c\mathbf{B}_d\mathbf{B}_d^{-1}(2\mathbf{B}_d)^{-1}(2\mathbf{B}_d)(\mathbf{B}_c + \mathbf{B}_d)^{-1}\mathbf{\Phi}_d\mathbf{p}_d.$$

Using the definition of the maximum eigenvalue, it implies that

$$\pi_c \leq \lambda_{\max}\{(\mathbf{B}_c + \mathbf{B}_d)^{-2}(2\mathbf{B}_d)^2\mathbf{B}_c\mathbf{B}_d^{-1}\}4\mathbf{p}_d^T\mathbf{\Phi}_d(2\mathbf{B}_d)^{-1}\mathbf{B}_d(2\mathbf{B}_d)^{-1}\mathbf{\Phi}_d\mathbf{p}_d$$
$$= \lambda_{\max}\{(\mathbf{B}_c + \mathbf{B}_d)^{-2}4\mathbf{B}_d\mathbf{B}_c\}\mathbf{p}_d^T\mathbf{\Phi}_d\mathbf{B}_d^{-1}\mathbf{\Phi}_d\mathbf{p}_d. \tag{2.11}$$

First notice that $\lambda_{\max}\{(\mathbf{B}_c+\mathbf{B}_d)^{-2}4\mathbf{B}_d\mathbf{B}_c\} > 0$ since the composite matrix is positive definite. Next, decompose the matrix,

$$\begin{aligned}
&\lambda_{\max}\{(\mathbf{B}_c+\mathbf{B}_d)^{-2}4\mathbf{B}_d\mathbf{B}_c\} \\
=&\lambda_{\max}\left\{\mathbf{I}-((\mathbf{B}_c+\mathbf{B}_d)^{-1}(\mathbf{B}_c-\mathbf{B}_d))^2\right\} \\
=&1-\lambda_{\max}\{((\mathbf{B}_c+\mathbf{B}_d)^{-1}(\mathbf{B}_c-\mathbf{B}_d))^2\} \\
\leq&1.
\end{aligned}$$

Thus, Equation (2.11) can be reduced to $\pi_c \leq \mathbf{p}_d^T\mathbf{\Phi}_d\mathbf{B}_d^{-1}\mathbf{\Phi}_d\mathbf{p}_d$. An lower bound on the profit can be written as

$$\frac{\pi_d}{\pi_c} \geq \frac{\mathbf{p}_d^T\mathbf{\Gamma}_d(\mathbf{\Gamma}_d+2\mathbf{B}_d)^{-1}\mathbf{B}_d(\mathbf{p}_d)}{\mathbf{p}_d^T\mathbf{\Phi}_d\mathbf{B}_d^{-1}\mathbf{\Phi}_d\mathbf{p}_d}. \tag{2.12}$$

Notice that every element in this expression only depends on the value of $\mathbf{q}_d$, the subscript d can be dropped for simplicity.

In the proof for Theorem 2.3.1 with affine demand (i.e., constant $\mathbf{B}$), we have shown that $\mathbf{p}_d = \mathbf{B}^{-1}(\mathbf{B}+\mathbf{\Gamma})(2+\mathbf{B}\mathbf{\Gamma})^{-1}\mathbf{B}\mathbf{p}_c$. Substitute this equation into Equation (2.6), followed by some algebraic manipulations, one can show that the expression is exactly the same with the right-hand-side in Equation (2.12). Therefore, we have shown that 3/4 is a lower bound with concave demand.

Upper bound: By concavity on the demand function,

$$\begin{aligned}
&\mathbf{q}_d-\mathbf{q}_c \leq -\mathbf{B}_c(\mathbf{p}_d-\mathbf{p}_c) \\
\Rightarrow\quad &\mathbf{q}_d+\mathbf{B}_c\mathbf{q}_d \leq \mathbf{q}_c+\mathbf{B}_c\mathbf{q}_c.
\end{aligned}$$

By the optimality condition shown in Proposition 2.2.9,

$$\underbrace{(\mathbf{\Gamma}_d(\mathbf{\Gamma}_d+2\mathbf{B}_d)^{-1}\mathbf{B}_d+\mathbf{B}_c)}_{\mathbf{\Omega}}\mathbf{p}_d \leq 2\mathbf{B}_c\mathbf{p}_c. \tag{2.13}$$

The composite matrix $\mathbf{\Omega}$ is a M-matrix and its inverse matrix is nonnegative. Therefore, $\mathbf{p}_d \leq 2\mathbf{\Omega}^{-1}\mathbf{B}_c\mathbf{p}_c$ and an upper bound on π_d is given by

$$\begin{aligned}\pi_d &= \mathbf{p}^T\mathbf{\Gamma}_d(\mathbf{\Gamma}_d + 2\mathbf{B}_d)^{-1}\mathbf{B}_d\mathbf{p}_d \\ &\leq 4\mathbf{p}_c^T\mathbf{B}_c\mathbf{\Omega}^{-1}\mathbf{\Gamma}_d(\mathbf{\Gamma}_d + 2\mathbf{B}_d)^{-1}\mathbf{B}_d(\mathbf{\Omega}^{-1}\mathbf{B}_c\mathbf{p}_c). \end{aligned} \tag{2.14}$$

Note that with the affine demand, i.e., $\mathbf{B}_c = \mathbf{B}_d$, Equation (2.13) becomes an equality and $\mathbf{\Omega} = 2(\mathbf{B}_c + \mathbf{\Gamma}_c)(\mathbf{\Gamma}_c + 2\mathbf{B}_c)^{-1}\mathbf{B}_c = \mathbf{\Omega_c}$. Equation (2.14) is simply,

$$\pi_d = \mathbf{p}_c^T(\mathbf{\Gamma}_c + 2\mathbf{B}_c)(\mathbf{B}_c + \mathbf{\Gamma}_c)^{-1}\mathbf{\Gamma}_c(\mathbf{B}_c + \mathbf{\Gamma}_c)^{-1}\mathbf{B}_c\mathbf{p}_c = f(\mathbf{p}_c).$$

With nonlinear demand, by using the Jacobian similarity property, there exists $\kappa_1 \geq 1$ such that $\pi_d \leq \kappa_1 f(\mathbf{p}_c)$. Using the definition of the maximum eigenvalue of a positive definite matrix, an upper bound on κ_1 is given by $\kappa_1 \leq \lambda_{\max}\{\mathbf{\Omega}^{-2}\mathbf{\Omega}_c^2(\mathbf{B}_d^{-1} + 2\mathbf{\Gamma}_d^{-1})^{-1}(\mathbf{B}_c^{-1} + 2\mathbf{\Gamma}_c^{-1})\}$. Note that when $\mathbf{B}_d = \mathbf{B}_c$, then $\kappa_1 = 1$. With this nonlinearity factor, we obtain an upper bound on the decentralized profit, i.e.,

$$\begin{aligned}\frac{\pi_d}{\pi_c} &\leq \kappa_1 \frac{f(\mathbf{p}_c)}{\pi_c} \\ &= \kappa_1 \frac{\mathbf{p}_c^T(\mathbf{\Gamma}_c + 2\mathbf{B}_c)(\mathbf{B}_c + \mathbf{\Gamma}_c)^{-1}\mathbf{\Gamma}_c(\mathbf{B}_c + \mathbf{\Gamma}_c)^{-1}\mathbf{B}_c\mathbf{p}_c}{\mathbf{p}_c^T\mathbf{B}_c\mathbf{p}_c}.\end{aligned}$$

Notice that the right-hand-side looks exactly the same as expression with affine demand in Equation (2.6), except matrix $\mathbf{B}_c$ and $\mathbf{\Gamma}_c$ depend on the optimal solution $\mathbf{p}_c$. With the same argument in the proof for affine demand, we can obtain an upper bound on π_d/π_c which is in terms of the competition index evaluated at the $\mathbf{p}_c$, i.e.,

$$\frac{\pi_d}{\pi_c} \leq \kappa_1 \frac{3 - r_{(2)}(\mathbf{B}_c)}{(2 - r_{(2)}(\mathbf{B}_c))^2} \leq \kappa_1 \frac{3 - r_{(1)}(\mathbf{B}_c)}{(2 - r_{(1)}(\mathbf{B}_c))^2}.$$

$\square$.

2.4.2 Convex demand

When the demand function is convex, for a given price $\bar{\mathbf{p}}$, the demand $\mathbf{q}(\bar{\mathbf{p}}) \leq \mathbf{q}(\mathbf{0}) - \bar{\mathbf{B}}\bar{\mathbf{p}}$, where $\bar{\mathbf{B}}$ is the Jacobian matrix evaluated at $\bar{\mathbf{p}}$.

Theorem 2.4.2 *When demand function $\mathbf{q}(\mathbf{p})$ is convex with $\mathbf{p}$, the efficiency of a decentralized supply chain is lower than that with affine demand. In particular, the total profit in a supply chain is bounded between*

$$\frac{3}{4\kappa_2} \leq \frac{\pi_d}{\pi_c} \leq \kappa_3 \frac{3 - r_{(2)}(\mathbf{B}_c)}{(2 - r_{(2)}(\mathbf{B}_c))^2} \leq \kappa_3 \frac{3 - r_{(1)}(\mathbf{B}_c)}{(2 - r_{(1)}(\mathbf{B}_c))^2},$$

where κ_2 and κ_3 are the Jacobian similarity factors.

Proof of Theorem 2.4.2. With convex demand, $\mathbf{q}_d - \mathbf{q}_c \geq -\mathbf{B}_c(\mathbf{p}_d - \mathbf{p}_c)$.

Lower bound : It is easy to show that when the demand is convex, the inequalities in Equation (2.13) and (2.14) switch its sign (compared to the case with concave demand), i.e.,

$$\pi_d \geq 4\mathbf{p}_c^T \mathbf{B}_c \boldsymbol{\Omega}^{-1} \boldsymbol{\Gamma}_d (\boldsymbol{\Gamma}_d + 2\mathbf{B}_d)^{-1} \mathbf{B}_d \boldsymbol{\Omega}^{-1} \mathbf{B}_c \mathbf{p}_c, \tag{2.15}$$

where the composite matrix $\boldsymbol{\Omega}$ is defined in Equation (2.13).

With nonlinear demand, by using the Jacobian similarity property, there exists $\kappa_2 \geq 1$ such that $\pi_d \geq \frac{1}{\kappa_2} f(\mathbf{p}_c)$, where $f(\mathbf{p}_c) = \mathbf{p}_c^T(\boldsymbol{\Gamma}_c + 2\mathbf{B}_c)(\mathbf{B}_c + \boldsymbol{\Gamma}_c)^{-1}\boldsymbol{\Gamma}_c(\mathbf{B}_c + \boldsymbol{\Gamma}_c)^{-1}\mathbf{B}_c\mathbf{p}_c$. By the definition of the minimum eigenvalue, we can obtain a bound on κ_2 as follows,

$$\begin{aligned}
\frac{1}{\kappa_2} &\geq \lambda_{\min}\{\boldsymbol{\Omega}^{-2}\boldsymbol{\Omega}_c^2(\mathbf{B}_d^{-1} + 2\boldsymbol{\Gamma}_d^{-1})^{-1}(\mathbf{B}_c^{-1} + 2\boldsymbol{\Gamma}_c^{-1})\} \\
&= \frac{1}{\lambda_{max}\{\boldsymbol{\Omega}^2\boldsymbol{\Omega}_c^{-2}(\mathbf{B}_d^{-1} + 2\boldsymbol{\Gamma}_d^{-1})(\mathbf{B}_c^{-1} + 2\boldsymbol{\Gamma}_c^{-1})^{-1}\}} \\
\Rightarrow \kappa_2 &\leq \lambda_{max}\{\boldsymbol{\Omega}^2\boldsymbol{\Omega}_c^{-2}(\mathbf{B}_d^{-1} + 2\boldsymbol{\Gamma}_d^{-1})(\mathbf{B}_c^{-1} + 2\boldsymbol{\Gamma}_c^{-1})^{-1}\}.
\end{aligned}$$

With this Jacobian similarity factor, we obtain a lower bound on the decentralized profit, i.e.,

$$\begin{aligned}\frac{\pi_d}{\pi_c} &\geq \frac{1}{\kappa_2}\frac{f(\mathbf{p}_c)}{\pi_c}\\ &= \frac{1}{\kappa_2}\frac{\mathbf{p}_c^T(\mathbf{\Gamma}_c+2\mathbf{B}_c)(\mathbf{B}_c+\mathbf{\Gamma}_c)^{-1}\mathbf{\Gamma}_c(\mathbf{B}_c+\mathbf{\Gamma}_c)^{-1}\mathbf{B}_c\mathbf{p}_c}{\mathbf{p}_c^T\mathbf{B}_c\mathbf{p}_c}.\end{aligned}$$

Notice that the right-hand-side after $\frac{1}{\kappa_2}$ looks exactly the same as expression with affine demand in Equation (2.6). We can use the same techniques to prove the lower bound in Theorem 2.3.1 to show that the lower bound on profit with concave demand is $\frac{3}{4\kappa_2}$, with $\kappa_2 \geq 1$.

Upper bound: By convexity,

$$\begin{aligned} &\mathbf{q}_c - \mathbf{q}_d \geq -\mathbf{B}_d(\mathbf{p}_c - \mathbf{p}_d)\\ \Rightarrow\ &(\mathbf{B}_c+\mathbf{B}_d)\mathbf{p}_c \geq 2\underbrace{(\mathbf{\Gamma}_d+\mathbf{B}_d)(\mathbf{\Gamma}_d+2\mathbf{B}_d)^{-1}\mathbf{B}_d}_{\Phi_d}\mathbf{p}_d.\end{aligned}$$

Since Φ_d is a M-matrix, its inverse is nonnegative and we obtain the inequality,

$$\mathbf{p}_d \leq \frac{1}{2}\Phi_d^{-1}(\mathbf{B}_c+\mathbf{B}_d)\mathbf{p}_c. \tag{2.16}$$

The profit in a decentralized setting is given by Equation (2.2). With Equation (2.16), an upper bound can be readily written as follows,

$$\pi_d \leq \frac{1}{4}\mathbf{p}_c^T(\mathbf{B}_c+\mathbf{B}_d)\Phi_d^{-1}\mathbf{\Gamma}_d(\mathbf{\Gamma}_d+2\mathbf{B}_d)^{-1}\mathbf{B}_d\Phi_d^{-1}(\mathbf{B}_c+\mathbf{B}_d)\mathbf{p}_c. \tag{2.17}$$

With the Jacobian similarity property, there exists a nonlinearity factor $\kappa_3 \geq 1$ such that $\pi_d \leq \kappa_3 f(\mathbf{p}_c)$. One upper bound on κ_3 is given by

$$\kappa_3 \leq \lambda_{\max}\{(\mathbf{B}_c+\mathbf{B}_d)^2(2\mathbf{B}_c)^{-2}\Phi_d^{-2}\Phi_c^2(2\Gamma_d^{-1}+\mathbf{B}_d^{-1})^{-1}(2\Gamma_c^{-1}+\mathbf{B}_c^{-1})\}.$$

The rest of the proof follows the same argument as that for the upper bound in Theorem 2.4.1. □

The bounds shown for concave and convex demand are very similar in the sense that the upper bounds suggests that a decentralized supply chain becomes more efficient as the competition in the market increases. The notable difference between Theorem 2.4.1 and Theorem 2.4.2 is that with convex demand, the worst case could be lower than 3/4.

To provide some explanation to the behavior with demand nonlinearity, note that the inefficiency in a decentralized supply chain is created by the successive markups imposed by the supplier and the retailers. A affine demand with a constant curvature produces a change in demand that is proportional to the price change. The same price change induces a change in demand which is proportionally smaller with concave demand, and larger with convex demand. Thus, the same markup in a decentralized supply chain will lead to a proportionally larger (smaller) demand drop with a convex (concave) demand, resulting in a larger (smaller) decrease in the total profit. Although it should be apparent that the prices are different with different demand functions, the intuitive explanation seems to agree with our analysis.

Our result on demand nonlinearity is consistent with the studies in oligopoly theory. For example, Malueg (1994) show that a monopolist facing a concave demand will produce a greater percentage of the efficient output level than if demand had been convex. Consequently, the relative welfare loss due to a monopoly is less when demand is concave rather than convex. We have shown that for supply chains, in terms of the worst-case performance, the inefficiency resulting from decentralization is relatively smaller when demand is concave rather than convex. Nonetheless, we would like to highlight that existing results in oligopoly theory often do not generalize to the multi-tier supply chain setting. For example, Farahat and Perakis (2009) have shown that the efficiency (in terms of total profit) of an oligopoly decreases with competition, whereas we have shown that a decentralized supply benefits from competition.

One interpretation of the shape of the demand curve is linked to the distribution of reserve price in the consumers' population. One could refer reserve prices to income, so that the shape of the demand curve can be derived from income distribution. In particular, three kinds of society give rise to three shapes of consumers' demand: A affine demand curve arises from a uniform distribution of reserve prices. A concave demand curve arises from a distribution of reserve price with a wide number of consumers having a similar middle reserve price, only few "rich" and few "poor". By contrast, a convex demand curve arises from a polarised distribution of reserve prices with most consumers having low reserve prices, few are "rich", and only slightly more are in the middle. Theorem 2.4.1 and 2.4.2 imply that a decentralized supply chain tends to be more efficient when the majority of the population has the similar reserve prices, and less efficient when the most consumers have low reserve prices and are sensitive to prices.

2.5 Tightness of Bounds

In the earlier sections, we have quantified several bounds on the performance of decentralized supply chains and proved that they are tight under special instances. This section addresses a natural questions that arises in the context of our analysis: How "good" are our bounds in general?

The first experiment is used to illustrate the tightness of the bounds for a general setting with an arbitrary number of asymmetric competing retailers with affine demand. For a comprehensive analysis, besides the bounds in terms of $r_{(2)}(\mathbf{B})$ and $r_{(1)}(\mathbf{B})$, we include an additional bounds in terms of the minimum eigenvalue of the normalized price sensitivity matrix as shown in Equation (2.9). It is crucial to note that we are able to compute bounds with information only based on the price elasticities. Neither the supplier's marginal cost $\mathbf{c}$ nor the maximum demand $\mathbf{q}(\mathbf{0})$ is needed.

The experiment consists of 35 scenarios and the results are shown in Figure 2-4. For each scenario, the number of retailers is uniformly picked between 2 and 20. Inputs, including the price sensitivity matrix, the vectors $\mathbf{q}(\mathbf{0})$ and $\mathbf{c}$ are also randomly generated. We first compute the optimal equilibrium solution and obtain the exact ratio, π_d/π_c. Based on the price sensitivity matrix, we also compute the corresponding eigenvalues, the degree of substitutability and their corresponding bounds.

The bound in terms of the eigenvalue of the price sensitivity matrix gives the

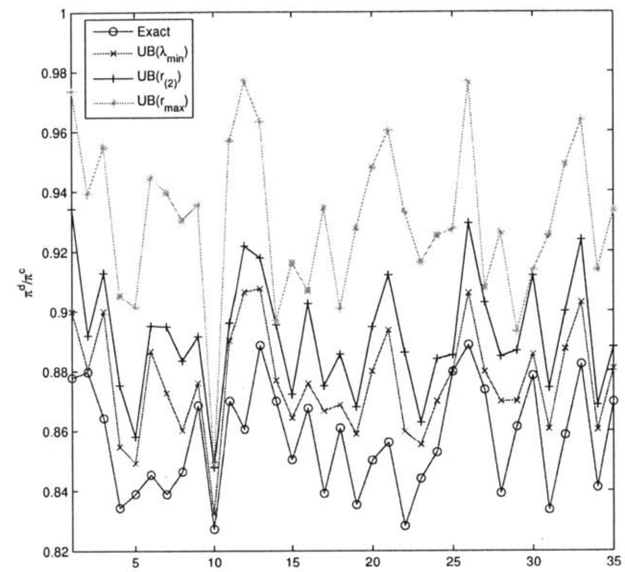

Figure 2-4: A simulation experiment for an arbitrary number of asymmetric retailers: The plot shows the exact ratio π_d/π_c and three upper bounds in terms of the eigenvalue, $r_{(2)}(\mathbf{B})$, and $r_{(1)}(\mathbf{B})$.

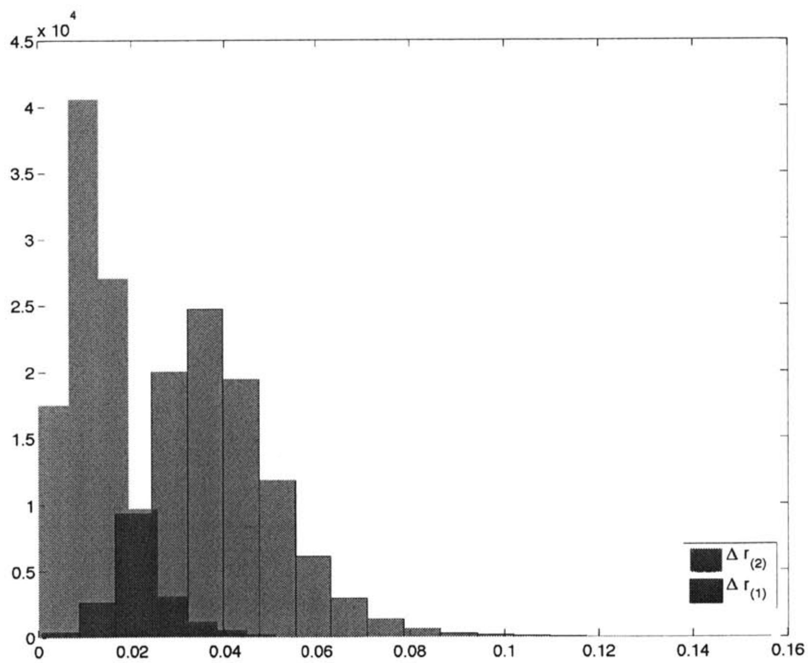

Figure 2-5: Histograms for errors between the exact values of π_d/π_c and the upper bounds in terms of $r_{(2)}(\mathbf{B})$ and $r_{(1)}(\mathbf{B})$ respectively for 10^5 instances.

most accurate estimation, with a maximum difference within 4% for all scenarios. The bounds in terms of $r_{(2)}(\mathbf{B})$ which uses information of only two "representative" retailers yield slightly looser estimation with accuracy within 7%. The bounds in terms of $r_{(1)}(\mathbf{B})$ which is based on a single retailer estimate the channel performance with accuracy within 12%.

We will now present a more comprehensive simulation experiment to compare the two bounds in terms of One-Firm and Two-Firm ratio. Figure 2.5 illustrates a result of a numerical simulation with 10^5 instances. For each instance, we generate a market structure of 2 to 20 retailers and randomly generate the inputs including $\mathbf{B}$, $\mathbf{c}$ and $\mathbf{p(0)}$. We then compute the exact values of the profit ratio π_d/π_c and the two upper bounds. Denote the error terms which are the differences between the exact value and the upper bound in terms of Two-Firm ratio as $\Delta r_{(2)}$ for all the instances (define $\Delta r_{(1)}$ similarly). For each array of the error terms, the elements are grouped into 20 equally spaced bins. The histograms for $\Delta r_{(2)}$ and $\Delta r_{(1)}$ are plotted in Figure 2.5. The x-axis reflects the range of error terms and the y-axis shows the number of instances that fall within the bins. Figure 2.5 depicts clearly the advantages of using $r_{(2)}$ over $r_{(1)}$: The spread of errors is smaller for $\Delta r_{(2)}$ than $\Delta r_{(1)}$ (i.e., [0, 0.131] vs. [0.003, 0.195]); the mean error is also lower for $\Delta r_{(2)}$ (i.e., 0.0117 vs. 0.0439).

The results highlight a trade-off between accuracy and complexity: In order to obtain a more accurate bound which is in terms of the minimum eigenvalue, we need information on the price elasticities of all the retailers. Taking a practical look, though it is relatively straightforward to estimate price elasticities, the task becomes increasingly challenging as the number of retailers in the market grows. Measurement errors in the data are unavoidable and eigenvalues are susceptible to perturbations (the computational procedure can be very inaccurate in the presence of round-off error). Furthermore, computing eigenvalues is as difficult as solving the original supply chain optimization problem and its complexity increases rapidly with the dimension of the matrix. On the other hand, for the bound in terms of $r_{(2)}(\mathbf{B})$, it is sufficient to select two retailers with the highest degree of complementarity (substitutability). Thus, the $r_{(2)}(\mathbf{B})$ bound is particularly useful when there is only enough data to make reasonable estimation about the most "representative" retailers in the market. Moreover, $r_{(2)}(\mathbf{B})$ provides an intuitive explanation on the channel performance as it implies the strength of the retailers and the degree of asymmetry in the market.

2.6 Extension and Conclusions

One of the assumptions used in this paper is that each retailer carries a single product. This assumption can be easily relaxed to incorporate a setting where each retailer carries multiple products. Consider the following supply chain, a single supplier offers wholesale price contracts to n retailers who carry a set m of products, where $m \geq n$. Retailer i offers product $\{m_{i-1}+1, \ldots, m_i\}$, where $m_0 := 0$, $m_n := m$, and $m_{i-1} < m_i$ for $i > 1$. When $m = n$, every retailer only carries a single product. Denote the price sensitivity matrix as $\mathbf{B}$, where $-\mathbf{B}$ is the Jacobian matrix of demand. Let $\mathbf{\Gamma}$ be a *block* diagonal matrix, consisting of n blocks, whose ith block is the square submatrix of $\mathbf{B}$ formed by the rows and columns indexed $m_{i-1}+1, \ldots, m_i$. $\mathbf{\Gamma}$ is referred to as the *intra-firm* price sensitivity matrix. Denote $\bar{\mathbf{B}} = \mathbf{B} - \mathbf{\Gamma}$ as the *inter-firm* price sensitivity matrix. For the setting when each retailer carries a single product, $\mathbf{\Gamma}$ and $\bar{\mathbf{B}}$ are simplified to the diagonal and off-diagonal matrix of $\mathbf{B}$ respectively.

This extended model allows each retailer to carry a bundle of imperfect *substitutes* or *complements*, as long as the retailers are viewed as competitors by consumers. That is, the inter-firm price sensitivity matrix $\bar{\mathbf{B}}$ is nonpositive. It can be shown that the lower bounds established in Theorem 2.3.1, 2.4.1 and 2.4.2 continue to hold with this extension.

With multiple products per retailer, the competition indices $r_{(1)}(\mathbf{B})$ and $r_{(2)}(\mathbf{B})$ need to be defined slightly differently to capture that fact that each retailer also carries other "competing" products.

Definition 2.6.1 ***Competition index for product*** i, $r_i(\mathbf{B})$ *for* $i \in \{1, \ldots, m\}$, $r_i = |[\mathbf{\Gamma}^{-1}\bar{\mathbf{B}}]_i|_1$, *where* $[\mathbf{\Gamma}^{-1}\bar{\mathbf{B}}]_i$ *refers to the* i^{th} *row of the matrix* $\mathbf{\Gamma}^{-1}\bar{\mathbf{B}}$ *and* $|\cdot|_1$ *is the* L_1 *vector norm (i.e.,* $|\mathbf{x}|_1 = \sum_j^m |x_j|$*).*

Remark When each retailer carries a single product, $r_i(\mathbf{B})$ is reduced to Defintion 2.2.7.

We normalize the price sensitivity matrix $\mathbf{B}$ with the intra-firm sensitivity matrix to capture the competing effect contributed by other products carried by the same retailer. With this definition for the competition index, the upper bounds developed in this chapter also carry through.

Inefficiency caused by competition has been the subject of extensive research in operations management in recent years. So far, however, primarily cases of symmetric equilibrium have been analyzed in the literature. In this paper, we provide an in-depth treatment of analyzing a general setting of a two-tier supply chain with imperfect competition and nonlinear demand. We present tractable and intuitive tight upper and lower bounds on performance metrics such as market penetration, chain-wide profit, consumer surplus and total social welfare.

We first propose a measure to evaluate competition in a supply chain with asymmetric retailers. We show that the performance of a decentralized supply chain improves with competition. Moreover, we characterize various performance metrics with tight bounds by using this measure. We conclude that the performance of a decentralized supply chain is predominantly determined by the two "strongest" retailers. Asymmetry between the retailers deteriorates the performance. The study on nonlinear demand suggests that compared to an affine demand, a supply chain is more efficient with concave demand and less efficient with convex demand.

An important take-away from our analysis is that for substitutes, price-only contracts are often "good enough" in the sense that there is limited room for potential gains from implementing other more complex contracts. In addition, as price-only contracts disproportionately favor the supplier, she has less incentive to adopt other contracts. The results provide some partial intuition that may help explain the popularity of price-only contracts in practice.

Chapter 3

Price of Anarchy for Supply Chains with Partial Positive Externalities

3.1 Introduction

The issue of inefficiency in a decentralized supply chain has attracted a lot of attention since Spengler (1950) introduced "double marginalization", i.e., two price markups, imposed by an upstream supplier and downstream retailers. The existing literature typically assumes that the demand of competing retailers exhibits *negative externalities* (or *substitutability*), i.e., an increase in one retailer's price induces an increase in the demand for other retailers' products. Research has shown that substitutability reduces the double marginalization effect, hence improves the channel performance Adida and DeMiguel (2010), Cachon and Lariviere (2001).

In this chapter, we focus on *positive externalities* (or *complementarity*), i.e., a decrease in the price of one product results in an increase in demand for all other products. We investigate the performance of a supply chain with a single supplier and several downstream retailers. The supplier offers wholesale price contracts to each retailer who carries multiple *imperfect complements*, inducing *partial* positive externalities. Most existing literature on supply chain performance with positive externalities typically assumes *perfect* complements Carr and Karmarkar (2005), Corbett and Karmarkar (2001), Wang and Gerchak (2003, 2004), which implies that whenever a purchase takes place, a consumer has to purchase one product from each and every retailer. Most complements in reality only exhibit partial complementarity – more complement goods are sold compared to the base goods, e.g., games versus game consoles, software versus hardware, ink cartridges versus printers, etc.

Based on a simple example, we illustrate a surprising phenomenon in a multi-product setting with partial complementarity – double marginalization may fail to exist in a decentralized supply chain! By characterizing the degree of complementarity for imperfect complements, we quantify the performance of a decentralized supply chain with respect to the centralized setting with upper and lower bounds. We show that the decentralized supply chain loses *at least* 25% of the optimal profit. We derive two lower bounds on its performance with respect to the complementarity effect which we will rigorously define later. We present the instances when the bounds are tight and demonstrate their performance in a general setting through numerical simulations.

Discussions on complements in the economics and the industrial organization literature tend to focus on single-tier oligopolistic settings Arora and Gambardella (1990), Bulow et al. (1985), Fudenberg and Tirole (1984), Milgrom and Roberts (1995). Most studies on supply chain performance with complements consider assembly chains Carr and Karmarkar (2005), Wang and Gerchak (2003, 2004) as opposed to distribution channels addressed in this work. Netessine and Zhang (2005) studies the impact of supply-side externalities on supply chains, where the complementary effect arises through product availability and prices are exogenous. Our work considers demand-side externalities that arise through prices which are endogenously determined.

3.2 The Model

We consider a supply chain with one supplier who offers wholesale price contracts to n retailers who carry a set m of products, where $m \geq n$. Retailer i offers product $\{m_{i-1}+1, \ldots, m_i\}$, where $m_0 := 0$, $m_n := m$, and $m_{i-1} < m_i$ for $i > 1$. When $m = n$, every retailer only carries a single product. In the notation below, we adopt a convention where vectors and matrices appear in boldface. As is traditional in the pricing literature Tirole (1988) Vives (1999), we consider affine demand functions, $\mathbf{q}(\mathbf{p}) = \bar{\mathbf{d}} - \mathbf{B}\mathbf{p}$ where $\bar{\mathbf{d}} \geq \mathbf{0}$. We assume that the price sensitivity matrix $\mathbf{B}$ is symmetric, which is a natural consequence of maximizing a quasilinear utility function of a representative consumer. To model the positive externalities or complements, it requires that $\partial q_i(\mathbf{p})/\partial p_j \leq 0$ for all i and j, $\partial p_i(\mathbf{q})/\partial q_i < 0$ for all i, and $\partial p_i(\mathbf{q})/\partial q_j \geq 0$ for all $j \neq i$ Vives (1999). The first condition implies that the demand for complements moves in the same direction when the price of one product changes, whereas the other conditions suggest that the prices of complements move in the opposite directions when the supply for one product changes. For instance, if the supply for product i increases, i's price decreases. This induces an increase in demand for product i, which then triggers an increase in demand for its complementary product j, resulting in an increase in product j's price. Together with the symmetry assumption on $\mathbf{B}$, it implies that $\mathbf{B}^{-1}$ belongs to the class of M-matrices and the reader is referred to Horn and Johnson (1985) for details. Let $\mathbf{\Gamma}$ be a *block* diagonal matrix, consisting of n blocks, whose ith block is the square submatrix of $\mathbf{B}$ formed by the rows and columns indexed $m_{i-1}+1, \ldots, m_i$. $\mathbf{\Gamma}$ is referred to as the *intra-firm* price sensitivity matrix. Denote $\bar{\mathbf{B}} = \mathbf{B} - \mathbf{\Gamma}$ as the *inter-firm* price sensitivity matrix. For the setting when each retailer carries a single product, $\mathbf{\Gamma}$ and $\bar{\mathbf{B}}$ are simplified to the diagonal and off-diagonal matrix of $\mathbf{B}$ respectively.

For each product, we assume that the production capacity is unlimited and marginal costs are constant. Let $\mathbf{c}$ denote the vector of marginal costs. Our final assumption is that $\tilde{\mathbf{d}} := \mathbf{B}^{-1}\bar{\mathbf{d}} > \mathbf{c}$, implying $\bar{\mathbf{d}} > \mathbf{B}\mathbf{c}$, i.e., the base demand at marginal costs must be positive. We will refer to this as Assumption $(\star)$.

Throughout the paper, we compare the performance of a decentralized supply chain to a benchmark setting of a centralized setting. Denote the wholesale prices, retail prices, order quantities and chain-wide profit as $\mathbf{w}$, $\mathbf{p}$, $\mathbf{q}$ and π respectively. We use superscripts d and c to differentiate the decentralized and the centralized settings.

In a decentralized supply chain, the supplier initiates the process by proposing a wholesale price contract $\mathbf{w_i}$ to every retailer i with the goal to maximize her profit. Each retailer then determines his retail prices $\mathbf{p_i}$, given the prices set by his competitors, $\mathbf{p_{-i}}$. We assume Nash Equilibrium has been reached where no single retailer can increase his profit by unilaterally changing his price. The problem for the supplier and the retailers can be written as follows:

$$
\begin{aligned}
(\pi_d)_s(\mathbf{w}) &\triangleq \max(\mathbf{w}-\mathbf{c})^T\mathbf{q}(\mathbf{p}(\mathbf{w})), \text{ s.t. } \mathbf{w} \geq \mathbf{0}. \\
(\pi_d)_{r_i}(\mathbf{p}) &\triangleq \max(\mathbf{p_i}-\mathbf{w_i})^T\mathbf{q_i}(\mathbf{p_i},\mathbf{p_{-i}}), \text{ s.t. } \mathbf{p} \geq \mathbf{0}, \quad \mathbf{q_i}(\mathbf{p_i},\mathbf{p_{-i}}) \geq \mathbf{0}.
\end{aligned}
$$

In a centralized supply chain, a central authority decides production quantities and retail prices across the chain with the objective to maximize the chain-wide profit by solving the following problem:.

$$
\pi_c(\mathbf{p}) \triangleq \max(\mathbf{p}-\mathbf{c})^T\mathbf{q}(\mathbf{p}), \text{ s.t. } \mathbf{p} \geq \mathbf{0}, \quad \mathbf{q}(\mathbf{p}) \geq \mathbf{0}.
$$

Proposition 3.2.1 *In a decentralized supply chain,* $\mathbf{w}_d = \frac{1}{2}(\mathbf{B}^{-1}\overline{\mathbf{d}}+\mathbf{c}), \mathbf{p}_d = \frac{1}{2}(\mathbf{B}+\mathbf{\Gamma})^{-1}(2\mathbf{B}+\mathbf{\Gamma})\tilde{\mathbf{d}}+\mathbf{c}, \mathbf{q}_d = \frac{1}{2}\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma}\tilde{\mathbf{d}}, \pi_d = \frac{1}{4}\tilde{\mathbf{d}}^T(2\mathbf{B}+\mathbf{\Gamma})(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma}\tilde{\mathbf{d}}.$ *In a centralized supply chain,* $\mathbf{p}_c = \frac{1}{2}(\mathbf{B}^{-1}\overline{\mathbf{d}}+\mathbf{c}), \mathbf{q}_c = \frac{1}{2}\mathbf{B}\tilde{\mathbf{d}}, \pi_c = \frac{1}{4}\tilde{\mathbf{d}}^T\mathbf{B}\tilde{\mathbf{d}}.$

Proof of Proposition 3.2.1 The decentralized problem is solved by backward induction. Since $\mathbf{B}^{-1}$ is a M-matrix, it is positive definite. This guarantees the existence and uniqueness of a pure strategy equilibrium to the unconstrained problem. We need to show that the solution obtained from the equilibrium condition also satisfies the nonnegativity constraint. The order quantities in the decentralized setting can be written as, $\mathbf{q}_d = \frac{1}{2}(\mathbf{B}^{-1}+\mathbf{\Gamma}^{-1})^{-1}\tilde{\mathbf{d}}$. Since $\mathbf{B}^{-1}$ is a M-matrix, $\mathbf{\Gamma}^{-1}+\mathbf{B}^{-1}$ is also a M-matrix, and its inverse is nonnegative. $\tilde{\mathbf{d}}$ is positive by Assumption ($\star$), thus, we have shown $\mathbf{q}_d \geq 0$. Similarly, we can show that $\mathbf{w}_d$ satisfies the nonnegativity constraint. For the centralized problem, $\mathbf{q}_c$ which is a product of a nonnegative matrix and a positive vector is clearly nonnegative. $\square$

3.3 Comparative Studies on Prices and Quantities

In this section, we begin with one example which highlights a interesting phenomenon that occurs in a decentralized supply chain with partial positive externalities. Consider the following setting with two retailers and three products. Retailer 1 carries the first two products and the retailer 2 carries the third product. The price sensitivity matrix, marginal cost, and the maximum demand are given as follows,

$$\mathbf{B} = \begin{bmatrix} 0.657 & 0.231 & 0.284 \\ 0.231 & 0.422 & 0.154 \\ 0.284 & 0.154 & 0.611 \end{bmatrix}, \mathbf{c} = \begin{bmatrix} 0.878 \\ 0.290 \\ 0.979 \end{bmatrix}, \bar{\mathbf{d}} = \begin{bmatrix} 1.432 \\ 1.373 \\ 1.267 \end{bmatrix}.$$

In a decentralized supply chain, the wholesale prices and the retail prices are given by $\mathbf{w}_d = (0.879, 1.336, 1.025)$ and $\mathbf{p}_d = (0.871, 1.856, 1.116)$ respectively. Notice $(p_d)_1 - (w_d)_1 = -0.008$. Retailer 1 is unable to cover the wholesale price and loses money every time when product 1 is sold!

This example illustrates a characteristic unique to pricing of complements: the base product (product 1 in the example) is priced low to generate sufficient sales volume which stimulates the demand for its complements (product 2). The objective is to create a level of profit which adequately covers losses sustained by the base product (otherwise, the retailer is better off by exiting the market). The almost universal tactic in the desktop printer business involves printers selling for as little as \$100 which include two ink cartridges, which themselves cost around \$30 each to replace. Thus the company prices low on the printers to create the anticipated revenue flow from selling the ink cartridges.

Proposition 3.3.1 *In a supply chain with partial positive externalities,*

(a) $\mathbf{w}_d = \mathbf{p}_c > \mathbf{c}$.

(b) $\mathbf{q}_d \leq \mathbf{q}_c$.

Proof of Proposition 3.3.1. (a) The inequality can be easily established under Assumption ($\star$). (b) The order quantities in the decentralized setting could be

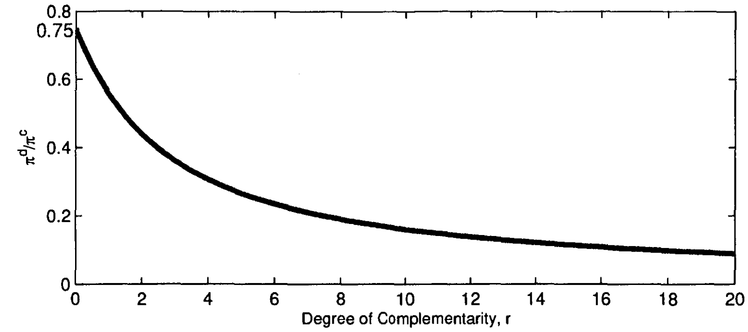

Figure 3-1: The total profit between the decentralized and the centralized settings with respect to the degree of complementarity for perfect complements. The curve serves as a lower bound for settings with imperfect complements, where r is replaced by $r_{(2)}$ and/or $r_{(1)}$.

expressed as $\mathbf{q}_d = \frac{1}{2}(\mathbf{B}^{-1} + \mathbf{\Gamma}^{-1})^{-1}\tilde{\mathbf{d}}$. Using the Inverse Binomial Theorem, we can expand the term $(\mathbf{B}^{-1} + \mathbf{\Gamma}^{-1})^{-1} = \mathbf{B} - (\mathbf{B}^{-1}\mathbf{\Gamma}^{1/2}\mathbf{\Gamma}^{1/2}\mathbf{B}^{-1} + \mathbf{B}^{-1})^{-1}$. $\mathbf{B}^{-1}\mathbf{\Gamma}^{1/2}$, $\mathbf{\Gamma}^{1/2}\mathbf{B}^{-1}$ and $\mathbf{B}^{-1}$ are M-matrices, thus the second term is nonnegative. It follows that $(\mathbf{B}^{-1} + \mathbf{\Gamma}^{-1})^{-1} \leq \mathbf{B}$. Since $\tilde{\mathbf{d}}$ is positive, we obtain the desired result. □

The supplier in the decentralized setting and the central planner in the centralized setting charge the same prices and keep a positive markup for every product they distribute. Any product whose price falls below the manufacturing cost will be dropped. The example above illustrates that for certain products, the retail price in a decentralized setting could be lower than in the centralized setting. Nonetheless, Proposition 3.3.1 states that for every product, fewer units are sold in the decentralized setting.

3.4 Comparative Studies on the Channel Profit

Decentralized supply chains are widely recognized as less efficient than centralized settings. The ratio π_d/π_c compares the channel profit in a decentralized setting relative to that in a centralized setting. Quantifying this ratio is essential in predicting system behavior and in designing appropriate rules of action to improve its performance. In the rest of this section, we will give a precise characterization of this ratio which solely depends on the price sensitivity information, assuming no knowledge of $\tilde{\mathbf{d}}$, aside from the fact that it is positive.

To facilitate the analysis with imperfect complements, we consider the following measure which captures the relative dependence of one product with respect to all other products available in the market.

Definition 3.4.1 ***The degree of complementarity*** *for product i, where $i \in \{1, \ldots, m\}$, $r_i = |[\mathbf{\Gamma}^{-1}\bar{\mathbf{B}}]_i|_1$, where $[\mathbf{\Gamma}^{-1}\bar{\mathbf{B}}]_i$ refers to the i^{th} row of the matrix $\mathbf{\Gamma}^{-1}\bar{\mathbf{B}}$ and $|\cdot|_1$ is the L_1 vector norm (i.e., $|\mathbf{x}|_1 = \sum_j^m |x_j|$).*

Remark When each retailer carries a single product, r_i could be simplified as $r_i = \sum_{j \neq i} \beta_{ij}/\beta_{ii}$, where β_{ij} is the (i,j)-element of $\mathbf{B}$.

The degree of complementarity of a product, measures the relative influence of positive externalities from other available products. It indicates how strongly one product is dependent on other retailers' products. To be precise, consider the setting when every retailer carries a single product, the numerator $\sum_{j \neq i} \beta_{ij}$ measures the aggregate demand change for retailer i's product triggered by retailer j's price change, the denominator β_{ii} reflects the demand change solely contributed by i's own price change. Thus, a high r_i implies a strong need to use i's product with other retailers' products. When some retailers carry multiple products, r_i compares the inter-firm positive externalities created by other retailers' products and the intra-firm externalities induced by other products which the same retailer carries.

By definition, $r_i \geq 0$. When products are perfect complements, then $r_i = r$ for all i. With imperfect complements, r_i varies across the products. In particular, as a unit change in demand of a base good (computer) induces a larger change in the demand for its complement good (software), the magnitude of r_i is higher for the latter.

Given a market with m products, each with the degree of complementarity r_i defined as before, we introduce the following two indices as "proxies" for the entire market.

Definition 3.4.2 *For a set of m products with the degree of complementarity given by $\mathbf{r} = (r_1, \ldots, r_m)$,*

(i) ***One-firm index:*** $r_{(1)} = \max_i r_i$.

(ii) ***Two-firm index:*** $r_{(2)} = \max_{i,j|j\neq i} \sqrt{r_i r_j}$.

The one-firm index represents the highest degree of complementarity among all available products, thus, it only requires price sensitivities with respect to that firm. The two-firm index is defined as a geometric mean of the two highest degrees of complementarities. This index captures the asymmetry effect between these two products. To be precise, rewrite $r_{(2)}$ as $\max_{i,j|j\neq i} \frac{1}{2}(r_i + r_j) - \frac{1}{2}(\sqrt{r_i} - \sqrt{r_j})^2$, where the second term decreases with the difference between r_i and r_j.

Theorem 3.4.3 *When retailers carry multiple imperfect complements, the total profit of a decentralized supply chain is bounded by*

$$\frac{3 + 2r_{(1)}}{(2 + r_{(1)})^2} \leq \frac{3 + 2r_{(2)}}{(2 + r_{(2)})^2} \leq \frac{\pi_d}{\pi_c} \leq \frac{3}{4}, \tag{3.1}$$

where the two lower bounds are tight with perfect complements, the upper bound is tight with independent products (i.e., noncompeting retailers).

Proof of Theorem 3.4.3. We first prove the lower bounds. Denote $\mathbf{w} = \mathbf{B}^{1/2}\tilde{\mathbf{d}}$ and $\mathbf{I}$ as the identity matrix. Rewrite the ratio $\pi_d/\ \pi_c$ as follows,

$$\begin{aligned}
\frac{\pi_d}{\pi_c} &= \frac{\tilde{\mathbf{d}}^T(2\mathbf{B}+\boldsymbol{\Gamma})(\mathbf{B}+\boldsymbol{\Gamma})^{-1}\mathbf{B}(\mathbf{B}+\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}\tilde{\mathbf{d}}}{\tilde{\mathbf{d}}^T\mathbf{B}\tilde{\mathbf{d}}} \\
&= \frac{\mathbf{w}^T\mathbf{B}^{-1/2}(2\mathbf{B}+\boldsymbol{\Gamma})(\mathbf{B}+\boldsymbol{\Gamma})^{-1}\mathbf{B}(\mathbf{B}+\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}\mathbf{B}^{-1/2}\mathbf{w}}{\mathbf{w}^T\mathbf{w}} \\
&= \frac{\mathbf{w}^T\mathbf{B}^{-1/2}(\mathbf{I}+\mathbf{B}(\mathbf{B}+\boldsymbol{\Gamma})^{-1})\mathbf{B}(\mathbf{I}-(\mathbf{B}+\boldsymbol{\Gamma})^{-1}\mathbf{B})\mathbf{B}^{-1/2}\mathbf{w}}{\mathbf{w}^T\mathbf{w}} \\
&= \frac{\mathbf{w}^T\mathbf{B}^{-1/2}(\mathbf{B}-\mathbf{B}(\mathbf{B}+\boldsymbol{\Gamma})^{-1}\mathbf{B}(\mathbf{B}+\boldsymbol{\Gamma})^{-1}\mathbf{B})\mathbf{B}^{-1/2}\mathbf{w}}{\mathbf{w}^T\mathbf{w}} \\
&= \frac{\mathbf{w}^T(\mathbf{I}-\mathbf{B}^{1/2}(\mathbf{B}+\boldsymbol{\Gamma})^{-1}\mathbf{B}^{1/2}\mathbf{B}^{1/2}(\mathbf{B}+\boldsymbol{\Gamma})^{-1}\mathbf{B}^{1/2}\mathbf{w}}{\mathbf{w}^T\mathbf{w}} \\
&= \frac{\mathbf{w}^T(\mathbf{I}-\mathbf{A}^2)\mathbf{w}}{\mathbf{w}^T\mathbf{w}}, \quad \text{where } \mathbf{A} = \mathbf{B}^{1/2}(\mathbf{B}+\boldsymbol{\Gamma})^{-1}\mathbf{B}^{1/2} \\
&\geq \lambda_{\min}(\mathbf{I}-\mathbf{A}^2) \\
&= 1-\lambda_{\max}(\mathbf{A}^2) \\
&= 1-\lambda_{\max}^2(\mathbf{A}),
\end{aligned} \tag{3.2}$$

where the last equality holds because matrix $\mathbf{A}$ is positive definite.

$$\begin{aligned}\lambda_{\max}(\mathbf{A}) &= \lambda_{\max}(\mathbf{B}^{1/2}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{B}^{1/2}) \\ &= \frac{1}{\lambda_{\min}(\mathbf{B}^{-1/2}(\mathbf{B}+\mathbf{\Gamma})\mathbf{B}^{-1/2})} \\ &= \frac{1}{\lambda_{\min}(\mathbf{I}+\mathbf{B}^{-1/2}\mathbf{\Gamma}\mathbf{B}^{-1/2})} \\ &= \frac{1}{1+\lambda_{\min}(\mathbf{B}^{-1/2}\mathbf{\Gamma}\mathbf{B}^{-1/2})}.\end{aligned}$$

Since $\mathbf{B}^{-1/2}\mathbf{\Gamma}\mathbf{B}^{-1/2} = \mathbf{B}^{-1/2}(\mathbf{\Gamma}\mathbf{B}^{-1})\mathbf{B}^{1/2}$ and $\mathbf{\Gamma}\mathbf{B}^{-1} = \mathbf{\Gamma}^{1/2}(\mathbf{\Gamma}^{1/2}\mathbf{B}^{-1}\mathbf{\Gamma}^{1/2})\mathbf{\Gamma}^{-1/2}$, which implies that $\mathbf{B}^{-1/2}\mathbf{\Gamma}\mathbf{B}^{-1/2}$, $\mathbf{\Gamma}\mathbf{B}^{-1}$ and $\mathbf{\Gamma}^{1/2}\mathbf{B}^{-1}\mathbf{\Gamma}^{1/2}$ are similar matrices, i.e., they all have the same eigenvalues.

$$\begin{aligned}\lambda_{\max}(\mathbf{A}) &= \frac{1}{1+\lambda_{\min}(\mathbf{\Gamma}^{1/2}\mathbf{B}^{-1}\mathbf{\Gamma}^{1/2})} \\ &= \frac{1}{1+\frac{1}{\lambda_{\max}(\mathbf{\Gamma}^{-1/2}\mathbf{B}\mathbf{\Gamma}^{-1/2})}} \\ &= \frac{1}{1+\frac{1}{\lambda_{\max}(\mathbf{G})}}, \qquad \text{where } \mathbf{G} = \mathbf{\Gamma}^{-1/2}\mathbf{B}\mathbf{\Gamma}^{-1/2} \\ &= \frac{\lambda_{\max}(\mathbf{G})}{1+\lambda_{\max}(\mathbf{G})}.\end{aligned}$$

Substituting this into Equation (3.2), we obtain the lower bound in terms of the maximum eigenvalue of $\mathbf{G}$,

$$\frac{\pi_d}{\pi_c} \geq 1 - \left(\frac{\lambda_{\max}(\mathbf{G})}{1+\lambda_{\max}(\mathbf{G})}\right)^2 = \frac{1+2\lambda_{\max}(\mathbf{G})}{(1+\lambda_{\max}(\mathbf{G}))^2}. \tag{3.3}$$

Because the inequality is decreasing in $\lambda_{\max}(\mathbf{G})$, we can further bound the ratio by upper bounding $\lambda_{\max}(\mathbf{G})$. We will show that $\lambda_{\max}(\mathbf{G}) \leq 1 + r_{(2)} \leq 1 + r_{(1)}$ and substitute this into Inequality (3.3) to establish the lower bounds.

To show $\lambda_{\max}(\mathbf{G})$ is bounded from below by the two indices, notice that the matrix $\mathbf{G}$ and $\mathbf{\Gamma}^{-1}\mathbf{B}$ are similar matrices, thus, $\lambda_{\max}(\mathbf{G}) = \lambda_{\max}(\mathbf{\Gamma}^{-1}\mathbf{B})$. Using Brauer's

Theorem (see Appendix), we obtain

$$\begin{aligned}
&(\lambda_{\max}(\mathbf{\Gamma}^{-1}\mathbf{B}) - 1)^2 \le \max_{i,j|j\neq i} r_i r_j \\
\Rightarrow\ & \lambda_{\max}(\mathbf{\Gamma}^{-1}\mathbf{B}) \le 1 + \max_{i,j|j\neq i} \sqrt{r_i r_j} \\
\Rightarrow\ & \lambda_{\max}(\mathbf{\Gamma}^{-1}\mathbf{B}) \le 1 + r_{(2)}(\mathbf{B}). \qquad (3.4)
\end{aligned}$$

Substituting Inequality (3.4) into (3.3), we obtain the first lower bound in terms of $r_{(2)}$. Notice $r_{(2)}(\mathbf{B}) \le r_{max}(\mathbf{B})$, we obtain the looser lower bound in terms of $r_{(1)}$.

To prove the upper bound $\pi_d/\pi_c \le 3/4$, it is equivalent to prove the matrix $\Phi(\mathbf{B}) = 3\mathbf{B} - 4(2\mathbf{B}+\mathbf{\Gamma})(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma}$ is copositive. By copositivity, $\mathbf{x}^T\Phi(\mathbf{B})\mathbf{x} \ge 0$, for all $\mathbf{x} \ge \mathbf{0}$. Thus, substituting $\tilde{\mathbf{d}}$ which is positive by Assumption ($\star$), we obtain

$$\begin{aligned}
& \tilde{\mathbf{d}}^T(3\mathbf{B} - 4(2\mathbf{B}+\mathbf{\Gamma})(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma})\tilde{\mathbf{d}} \ge 0 \\
\Leftrightarrow\ & \frac{\tilde{\mathbf{d}}^T(2\mathbf{B}+\mathbf{\Gamma})(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma}\tilde{\mathbf{d}}}{\tilde{\mathbf{d}}^T\mathbf{B}\tilde{\mathbf{d}}} \le \frac{3}{4} \\
\Leftrightarrow\ & \frac{\pi_d}{\pi_c} \le \frac{3}{4}.
\end{aligned}$$

To prove the copositivity of the matrix $\Phi(\mathbf{B})$, we first express $\mathbf{B} = \mathbf{\Gamma}^{1/2}\mathbf{G}\mathbf{\Gamma}^{1/2}$ (since $\mathbf{G} = \mathbf{\Gamma}^{-1/2}\mathbf{B}\mathbf{\Gamma}^{-1/2}$). Then we rewrite $\Phi(\mathbf{B})$ as $\Phi(\mathbf{B}) = \mathbf{\Gamma}^{1/2}\Phi(\mathbf{G})\mathbf{\Gamma}^{1/2}$, where $\Phi(\mathbf{G}) = 3\mathbf{G} - 4(2\mathbf{G}+\mathbf{I})(\mathbf{G}+\mathbf{I})^{-1}\mathbf{G}(\mathbf{G}+\mathbf{I})^{-1}$. $\mathbf{\Gamma}$ is a nonnegative block diagonal matrix, thus, the matrix $\Phi(\mathbf{B})$ is copositive if $\Phi(\mathbf{G})$ is copositive. Express $\Phi(\mathbf{G})$ as follows,

$$\begin{aligned}
\Phi(\mathbf{G}) &= \mathbf{G} - \mathbf{I} + 2\mathbf{G} + \mathbf{I} - 4(2\mathbf{G}+\mathbf{I})(\mathbf{G}+\mathbf{I})^{-1}(\mathbf{G}+\mathbf{I}-\mathbf{I})(\mathbf{G}+\mathbf{I})^{-1} \\
&= (\mathbf{G}-\mathbf{I}) + (2\mathbf{G}+\mathbf{I})\left(\mathbf{I} - 4(\mathbf{G}+\mathbf{I})^{-1}(\mathbf{I} - (\mathbf{G}+\mathbf{I})^{-1})\right) \\
&= (\mathbf{G}-\mathbf{I}) + (2\mathbf{G}+\mathbf{I})\left(\mathbf{I} - 4(\mathbf{G}+\mathbf{I})^{-1} + 4(\mathbf{G}+\mathbf{I})^{-2}\right) \\
&= (\mathbf{G}-\mathbf{I}) + (2\mathbf{G}+\mathbf{I})(\mathbf{G}+\mathbf{I})^{-2}(\mathbf{G}-\mathbf{I})^2 \\
&= (\mathbf{G}-\mathbf{I}) + (2\mathbf{G}+\mathbf{I})((\mathbf{G}+\mathbf{I})^{-1}(\mathbf{G}-\mathbf{I}))^2.
\end{aligned}$$

$\mathbf{G}-\mathbf{I}$ is a nonnegative matrix as $\mathbf{G}$ is a nonnegative matrix with diagonals equal to 1.

The second term is a product of a positive definite matrix, $(2\mathbf{G}+\mathbf{I})$, with a positive semi-definite matrix, $((\mathbf{G}+\mathbf{I})^{-1}(\mathbf{G}-\mathbf{I}))^2$, therefore, it is also positive semi-definite. Matrix $\Phi(\mathbf{G})$ is a sum of a nonnegative matrix and a positive semi-definite matrix, therefore, it is copositive. $\square$

The theorem suggests that the performance of a decentralized supply chain deteriorates with the degree of complementarity. The best scenario that one can hope for, arises when products are independent, i.e., $r_i = 0$ for all i. As the products exhibit higher complementarity effect, the total profit earned in a decentralized supply chain decreases with respect to the centralized setting and loses *at least* 25% of the optimal channel profit.

The theorem reveals two sources of distortions in a decentralized supply chain with partial positive externalities. The first distortion is attributed to vertical competition between the supplier and the downstream retailers. This alone costs a decentralized setting 25% of the optimal profit. The second distortion stems from the "neglected" complementary effect across the retailers. In contrast with a centralized setting where all products are priced to generate a demand level that maximizes the total profit, a retailer in a decentralized supply chain is only interested in his own profit. By ignoring the complementary effect induced by other retailers, the retailers in a decentralized setting charge prices which induce a lower demand level for every product as compared to the centralized setting (as shown in Proposition 3.3.1). As the complementary effect across the retailers grows, the second distortion increases and gives rise to a larger profit loss in a decentralized setting.

The bounds in Theorem 3.4.3 also suggest that the rate of decrease is more substantial when the degree of complementarity is small. As illustrated in Figure 3-1, when the degree of complementarity increases from 0 to 1, the relative profit in the decentralized chain decreases from 75% to 55.6% of the optimal profit; when the complementarity effect increases to 2, the decentralized setting captures less than half of the optimal profit. The main implication is that the decentralized setting could experience significant profit loss even when the products exhibit a small degree of complementarity.

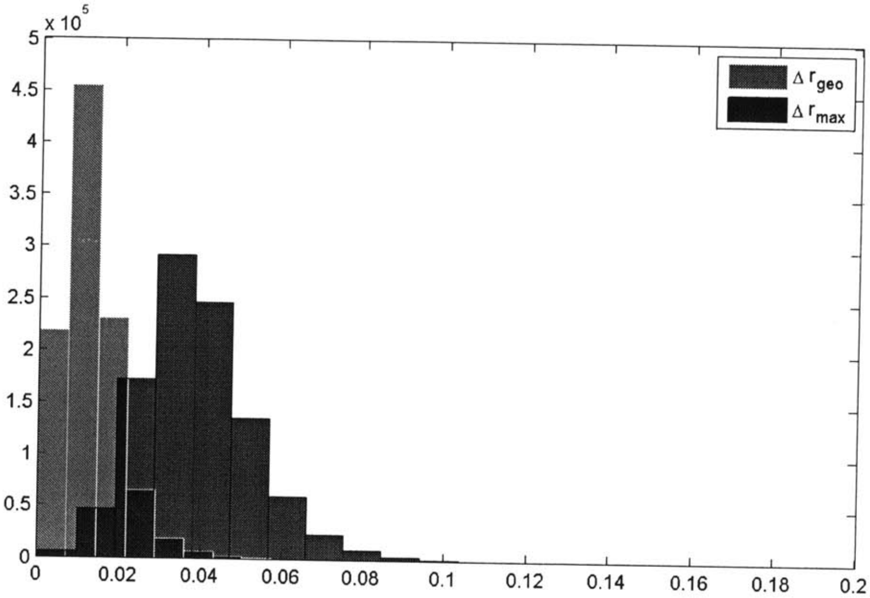

Figure 3-2: Histograms for errors between the exact values of π_d/π_c and the lower bounds in terms of $r_{(2)}$ and $r_{(1)}$ respectively for 10^6 instances.

An interesting observation from Theorem 3.4.3 is that the performance of a decentralized supply chain with complements is predominantly determined by the products which exhibit the highest degree of complementarity. Since the complement good has a much higher degree of complementarity than the base good, Theorem 3.4.3 implies that in order to boost the performance of a decentralized chain, it is important to "strengthen" the complementary products.

Our analysis thus far shows that the bounds in Theorem 3.4.3 are tight when the retailers carry perfect complements, i.e., $r_{(1)} = r_{(2)} = r_i$, for all i. In the following two simulation experiments, we will investigate the performance of the bounds with imperfect complements, i..e., there exists some i and j, where $r_i \neq r_j$. In particular, we will highlight the strength of using the bound in terms of $r_{(2)}$ which captures the asymmetry effect for products with partial complementarity.

Figure 3-2 illustrates a result of a numerical simulation with 10^6 instances. For each instance, we generate a market structure of 2 to 20 retailers and randomly generate the inputs including $\mathbf{B}$, $\mathbf{c}$ and $\bar{\mathbf{d}}$. We then compute the exact values of the profit ratio π_d/π_c and the two lower bounds. Denote the error terms which are the

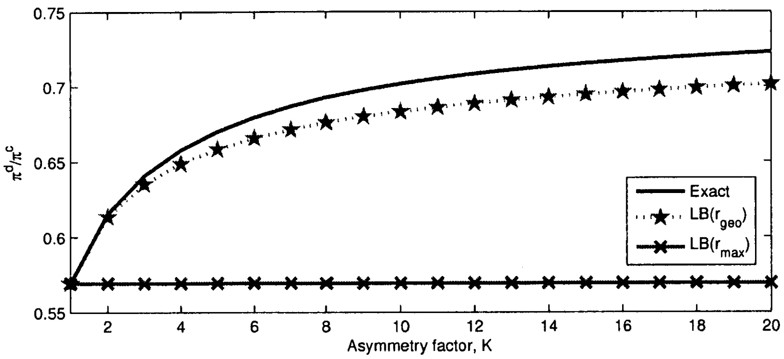

Figure 3-3: The exact value of π_d/π_c and the two lower bounds as k increases from 1 to 20.

differences between the exact value and the lower bound $LB(r_{(2)})$ as $\Delta r_{(2)}$ for all the instances (define Δr_{max} similarly). For each array of the error terms, the elements are grouped into 20 equally spaced bins. The histograms for $\Delta r_{(2)}$ and $\Delta r_{(1)}$ are plotted in Figure 3-2. The x-axis reflects the range of error terms and the y-axis shows the number of instances that fall within the bins. Figure 3-2 depicts clearly the advantages of using $r_{(2)}$ over $r_{(1)}$: The spread of errors is smaller for $LB(r_{(2)})$ than $LB(r_{(1)})$ (i.e., [0, 0.143] vs. [0.003, 0.188]); the mean error is also lower for $LB(r_{(2)})$ (i.e., 0.0126 vs. 0.0389).

The advantage of using $r_{(2)}$ is even more prominent as the asymmetry between the two products with the highest degree of complementarity increases. Consider a setting with two retailers who each carries one product, where $\mathbf{c} = (c, c)$, $\bar{\mathbf{d}} = (\bar{d}, \bar{d})$ and the price sensitivity matrix is defined as $\begin{bmatrix} \beta & 1 \\ 1 & k\beta \end{bmatrix}$, where $k \geq 1$ represents the asymmetry factor between the two retailers. When $k = 1$, the setting described above is fully symmetric. The exact expression of the profit ratio can be written as, $\pi_d/\pi_c = \frac{b(12\beta^4k^3+12\beta^4k^2+16\beta^3k^2-9\beta^2k^2-9\beta^2k-6\beta k+2k+2)}{(\beta+\beta k+2)(4\beta^2k-1)^2}$. Note that, in general, it is hard to express the profit ratio for problems with imperfect complements in higher dimensions. The degree of complementarity for two retailers is defined as $r_1 = 1/\beta$, and $r_2 = 1/(k\beta)$. Thus, the two indices are $r_{(1)} = \max_i r_i = 1/\beta$ and $r_{(2)} = \sqrt{r_1 r_2} = 1/(\sqrt{k}\beta)$. As k increases, $r_{(1)}$ remains the same while $r_{(2)}$ decreases. We obtain the two lower bounds, namely, $LB(r_{(1)})$ and $LB(r_{(2)})$ in Theorem 3.4.3 by substituting the corresponding index. Figure 3-3 records the exact value of π_d/π_c and the two lower bounds as k increases from 1 to 20. When $k = 1$, all expressions yield the same value. As k increases, the gap between the exact value and $LB(r_{(1)})$ widens to 15% since $r_{(1)}$ ignores the impact of asymmetry. At the same time, the differences between the exact value and $LB(r_{(2)})$ stays within 2% for this experiment.

We would like to point out that to compute the exact values of π_d/π_c, one would have to estimate the pair-wise price sensitivity across all products in the market, marginal costs $\mathbf{c}$ and maximum demand $\bar{\mathbf{d}}$ (when products are free), which is a challenging task to begin with. Both lower bounds are independent of $\mathbf{c}$ and $\bar{\mathbf{d}}$ and only require information of one or two products with the highest degree of complementarity in the market. In addition, we have also shown that using the two-firm index allows us to estimate the supply chain performance with high accuracy based on the simulations.

Chapter 4

Price of Anarchy for Congested Systems

4.1 Introduction

In a recent study conducted by the Federal Communication Committee (FCC), it predicts that the demand for wireless bandwidth is expected to surpass the available spectrum by as early as 2014. When the amount of wireless traffic exceeds the available bandwidth, users in the affected network would experience spotty services, dropped calls, and sluggish data speeds. With the rapid proliferation of new content-rich multimedia devices such as smartphones and tablet PCs, network congestion is becoming a common problem in many urban areas. For instance, many AT&T users in San Francisco and New York were angered by service degradation, when AT&T, the exclusive seller of iPhone in the United States until 2011, did not adequately anticipate the demand surge in bandwidth accompanied by the device's huge popularity (WSJ December 9, 2009). In the U.S., FCC currently auctions off spectrum bands that then become the property of the purchaser. Given the demand for wireless bandwidth is expected to grow between 25 and 50 times the current levels within 5 years (FCC Technical Paper 2010), one may wonder what kind of measure FCC should take to control network congestion and ensure an efficient usage of the finite bandwidth.

Besides the field of wireless networks, another area routinely plagued by congestion is airports. In July 2009, for example, 30% of the flights in the U.S. domestic market arrived late, up from 20% in July 2005. It has been reported that flight delays cost passengers, airlines and the U.S. economy more than $40 billion in 2007 (The Joint Economic Committee Report 2008). The rise in delays, not surprisingly, correlates with a significant increase in the number of flights. However, research shows that the vast majority is coming from substitution away from large aircraft in favor of more frequent operations of smaller planes. Whalen et al. (2008) found out that at LaGuardia, for example, the average number of seats per aircraft was 143 in the first quarter of 1998, but that fell to 94 as of the first quarter of 2007. Similar trend was also observed in other congested airports such as O'Hare. The existing system across the U.S. is unlikely to promote efficient use of the scarce runway space: Current landing fees depend only on aircraft weight and do not vary by time of day; runway allocation is mostly done on first-come, first-serve basis. In fact, the weight-based landing fees are often blamed for creating the wrong incentives for airlines to use small aircrafts. The widespread problem of airport congestion raises the issue of finding a rationing mechanism to use the runway space efficiently.

One common feature in the two examples mentioned above is oligopolistic competition with congestion effects. In the example of wireless communication, as of the fourth quarter in 2008, Verizon, AT&T, Sprint, Nextel, and T-Mobile together control 89% of the U.S. market. While airports vary in sizes, a small number of airlines usually dominates an airport in terms of flight share. For both industries,, Cournot has received both theoretical and empirical support as a good model of competition. Theoretically, Kreps and Scheinkman (1983) show that Cournot models best approximate the long run results of two-stage competition with capacity choice followed by price setting. Airlines compete by first setting schedules and later set prices to fill seats. Wireless service providers first purchase bandwidth and then determine prices for subscriptions. Empirical work also supports Cournot models reflecting actual competition in these two industries, see for example, Weisman (1990), Brander and Zhang (1990, 1993), Oum et al. (1993), Parker and Roller (1997) and Faulhaber and Hogendorn (2000).

However, we would like to point out a subtle difference associated with the cost effects in the two examples: As wireless service providers use different frequency bands to reduce signal interference, congestion in one provider's network does not affect other networks. Congestion effect in this case is *self-contained*, i.e., the congestion cost associated with one service provider only depends on his service level, e.g., the number of users, the amount of bandwidth consumed. For the example of wireless communication, service degradation is only experienced by users in the affected network (a bogged-down AT&T network does not affect the service quality of other carriers). This is in contrast with the airport example: When one airline schedules an additional flight in a congested airport, it creates additional delays for every flight which attempts to land and take off. Besides the self-contained cost component, cost in this setting also has the *spillover* effect as an increase in delay affects everyone in the system.

In this chapter, we study both scenarios, depending on whether spillover cost is present. We consider congestion pricing as a control mechanism to ration demand, with the goal to improve the societal welfare. Our model consists of several service providers with differentiated services, competing for users who are sensitive to both prices and congestion costs. A facility manager imposes an admission-level pricing scheme on the service providers. Relating this model to the two examples, FCC could play the role of the facility manager and determines a unit price for bandwidth allocated to each wireless network operator, who in turn determines how many users to enroll. Similarly, FAA would be responsible for imposing a landing fee and each airline determines its flight frequency subsequently.

Our main contributions of the paper are the following:

Welfare analysis with nonlinear convex costs. We compare the total welfare in an unregulated setting where the service providers have free access to the facility to that of the social optimum, so as to assess how much welfare is lost due to the lack of coordination. Congestion effects are modeled as nonlinear convex cost functions which increase with the output level. We show that with self-contained costs, the maximum welfare loss in the unregulated setting is limited. In fact, this loss is capped at 25%

of the optimal welfare, even in the presence of highly nonlinear costs. Moreover, the efficiency of the unregulated setting improves as the competition among the service providers increases. When spillover cost is also present, the performance highly depends on *net marginal externality*, which is measured by the relative magnitude of the marginal spillover cost and the marginal consumer surplus when an additional user is enrolled. With positive net externality, we show that the unregulated setting could be just as efficient as the social optimum. However, we also show that when net externality becomes negative due to high spillover costs, the welfare loss in the unregulated setting could be arbitrarily high, even with affine marginal cost. The latter validates the need of implementing some rationing mechanisms in airports to curb congestion since there are rooms for substantial potential welfare gains.

An alternative perspective on mergers in the absence of cost synergy. Most arguments which support mergers rely on the internal efficiency gain in the form of cost reduction (e.g., economies of scale, technological progress and eliminating redundancy etc). We show that even in the absence of such internal cost synergy, the society could potentially benefit from mergers. We show that it only happens when congestion has the spilling effect. The key idea is that the merged firm internalizes more congestion, thus, leaving a smaller amount of uninternalized congestion onto others. This optimistic view on mergers, however, does not apply to the case when congestion cost is self-contained. Under this setting, our analysis agrees with the conventional wisdom that reduced competition leads to a lowered total welfare in the absence of cost reduction. While we recognize that our model only addresses one aspect on mergers from the perspective of total welfare, our analysis could offer an explanation (at least partially) to the Court's rejection for the proposal between T-Mobile with AT&T versus a clearance for the consolidation of United with Continental airlines.

Social acceptance and a novel implementation. We address some of the issues which traditionally have made the adoption of congestion pricing a challenging task in practice. Firstly, to address the question of whether congestion pricing is necessary, one needs to measure how much potential benefit such a scheme could offer. We quantify the welfare loss in an unregulated setting to a social optimum by deriving tight upper and lower bounds which depend on at most two parameters. The parameters measure the degree of competition in the oligopoly market and the relative magnitude of the spillover cost. We demonstrate with simulations that the bounds are able to predict the gain computed from the model for a high amount of accuracy. Thus, instead of doing a full-blown estimation of all the inputs to the model (which is a challenging task to begin with), it is possible to quantify the welfare gain with confidence by only estimating a couple of parameters. We identify one obstacle for congestion pricing to be implemented, that is, without a proper channel of investing the revenue collected from congestion pricing, users and service providers will be strictly worse-off. Clearly, such consequences leave individuals with little desire to adopt the scheme. We propose a proportional rule which appeals to the self-interest of participants by ensuring a positive welfare improvement for everyone. As the proposed scheme ensures not only social optimality but also individual welfare improvement, we believe it has a greater likelihood of gaining support in practice.

4.1.1 Related literature

Motivated by congestion management in transportation and communication networks, there has been a huge body of literature to analyze traffic in a congested network (e.g., Hamdouch et al. 2007, Hayrapetyan et al. 2007 and Maille and Stier-Moses 2009). Acemoglu and Ozdaglar 2007, Ozdaglar 2008 study competition among profit-maximizing oligopolists who set prices on the links and congestion cost on each link only depends on the traffic volume on that link (this corresponds to self-contained congestion cost in this chapter). Users choose the path based on a notion of *full price*, which is the sum of the price paid to the oligopolists and the congestion cost. In these models, a single price prevails in equilibrium as a result of homogeneous services. In this work, we adopt the full price concept to capture user equilibrium. Our model differentiates itself from prior work by incorporating several new features, including differentiated services, elastic demand and spillover congestion cost.

Compared to the vast amount of literature which focuses on road congestion, there is less work addressing congestion pricing in other settings where users have oligopolistic power. Daniel (1995, 2001), Brueckner (2002, 2005) and Mayer and Sinai (2003) have considered congestion pricing for airports. Most of the studies derive their findings either through simulations with empirical data or theoretical analysis on a symmetric duopoly. In general, the topic on congestion pricing has long been recognized as controversial and challenging. Jones (1991) suggests that the support for the scheme is much higher when it is presented as a complete financial package with explicit proposals for using the revenues in his survey paper. Economists such as Goodwin (1990) and Small (1992) analyze how to use revenues collected from road tolls and propose dividing the money equally as an reimbursement to road users, tax rebates and investment for new transportation services. In this work, we quantify how much each service provider and user should gain from adopting congestion pricing based on a proportional rule, with a goal that everyone should experience a positive welfare improvement.

Several papers in the operations management literature have addressed the issue of congestion in service industries (see for example, Allon and Federgruen 2007, 2008, Cachon and Harker 2002, Weintraub et al. 2010, etc). They assume each firm owns a resource and is only sensitive to congestion caused by users using his resource. In our model, differentiated services belong to different service providers who participate in an oligopolistic quantity competition to maximize their own profit. In addition, the facility manager acts as the Stackelberg leader who determines optimal access fees by anticipating that service providers will select their service levels according to the equilibrium.

The optimal congestion pricing considered in our model resembles a "coordinating contract" in the supply chain literature (e.g., Corbett et al. 2005, Bernstein and Federgruen 2003, 2007, Martinez-de Alberniz and Simchi-Levi 2009, Cachon and Kok 2010). A key difference is that the objective in our model is to maximize the societal welfare which sums over consumer surplus, producer surplus as well as the congestion pricing revenue. We show that despite a higher societal welfare after implementing a coordinating contract, service providers and users might be worse off compared to their counterparts in the unregulated setting. Therefore, it leaves little incentives for adoption, in addition to the high administrative costs which are commonly associated with coordinating contracts. To bypass this difficulty, based on an n-person bargaining game, we show that there exists an alternative implementation which achieves the social optimum and ensures that every entity in the system could gain from this scheme.

Recently, several talks on mega-mergers have created quite a stir in the media, e.g., AT&T with T-Mobile in the mobile marketplace, as well as United with Continental in the airline industry. The primary argument against mergers is that the reduced competition could lead to many undesirable repercussions such as price increases, which ultimately hurt consumers. For example, based on a symmetric setting, Deneckere and Davidson (1985) and Farrell and Shapiro (1990) have shown exactly such behavior under both price and quantity competition, in the absence of cost reduction resulting from a merger. Some economists have challenged this pessimistic view on mergers which is based on "consumer welfare" and propose the use of "total welfare" to evaluate mergers (see for example, Neven and Roller (2005), Heyer (2006)). Perhaps the most influential contribution which advocated the total welfare approach in merger analysis is by Williamson (1968). By focusing on the net welfare impact, the author argues that when the benefit from cost saving (from realization of internal efficiency) offsets the welfare loss due to a price increases (from greater market power), the society benefits from the merger. The key argument which supports mergers hinges on the internal cost synergy from the merger. We show in this chapter that, even in the absence of cost reduction, mergers could be beneficial when congestion has the spillover effect.

Lastly, our work which measures the performance of an unregulated setting with respect to a centralized system is related to a stream of literature on *price of anarchy*, popularized by Koutsoupias and Papadimitriou (1999). It compares the performance of the worst-case Nash equilibrium with respect to the centralized system. The concept has been used in transportation networks (Roughgarden and Tardos 2002, Correa et al. 2004, 2007, Roughgarden 2005, Perakis 2007), network pricing (Acemoglu and Ozdaglar 2007, Weintraub et al. 2010), oligopolistic pricing games in a single tier (Farahat and Perakis 2010a,b), and supply chain games with exogenous pricing (Perakis and Roels 2007, Martinez-de Alberniz and Simchi-Levi 2009, Martinez-de Alberniz and Roels 2010).

The rest of the chapter is organized as follows. In Section 4.2, we introduce a three-level model with assumptions used in the paper. Section 4.3 performs the welfare analysis, comparing the total social welfare achieved in the unregulated setting to the maximum social welfare which could be achieved by implementing congestion pricing. In Section 4.4, we evaluate the tightness of the bounds via computational analysis. In Section 4.5, we discuss the potential merits of mergers. We address some issues which make congestion pricing unattractive in reality and discuss potential remedies that would make the scheme more appealing from an individual's perspective in Section 4.6. The conclusions can be found in Section 4.7.

4.2 Model

We consider a facility with n differentiated services, each offered by a provider. We denote q_i as the output level chosen by service provider $i = 1, \ldots, n$. The service providers compete in a quantity competition as he can control his service level q_i. The price for provider i's service is denoted by p_i. We use lower-case, boldface letters to represent column vectors, e.g., $\mathbf{p} = (p_1, ..., p_n)$ and $\mathbf{q} = (q_i, ..., q_n)$ represent the market prices and the output levels respectively.

Given an output level $\mathbf{q}$, we denote the marginal utility obtained from consuming an infinitesimal amount of service i as $u_i(\mathbf{q})$. As is traditional in the pricing literature (see Vives 1999), we consider the marginal utility function as a affine function of output levels:

$$\mathbf{u}(\mathbf{q}) = \bar{\mathbf{p}} - \mathbf{B}\mathbf{q} = \begin{bmatrix} \overline{p}_1 \\ \vdots \\ \overline{p}_n \end{bmatrix} - \begin{bmatrix} \beta_{11} & \ldots & \beta_{1n} \\ \vdots & \ddots & \vdots \\ \beta_{n1} & \ldots & \beta_{nn} \end{bmatrix} \begin{bmatrix} q_1 \\ \vdots \\ q_n \end{bmatrix},$$

where $\overline{\mathbf{p}} = (\overline{p}_1, \ldots, \overline{p}_n)$ represents the maximum prices that a user is willing to pay for the services. Different $\overline{p}_i$ captures the quality differences perceived by consumers, which could be affected by factors such as brand recognition, word-of-mouth effects, prior experience with the product, etc.

Assumption 4.2.1 *Matrix* $\mathbf{B}$ *is a symmetric and positive-definite matrix.*

$\beta_{ii} > 0$ means that each service sees a downward sloping demand resulted from users' diminishing return. $\beta_{ij} \geq 0$ captures the relationship that service i and j are substitutes, i.e., an increase in production of service j lowers the willingness of a user to pay for service i. Symmetry of the matrix implies that the cross-effects of any two service providers' output changes are equal. This is a natural consequence of maximizing a quasilinear utility function of a representative consumer.

In a congested facility, both the users and the service providers are affected congestion. For the case of airport congestion, airlines have to pay extra for crew, fuel, and maintenance costs while delayed travelers and their employers lose productivity, business opportunities and leisure activities. In this work, congestion effect is modeled in currency equivalent terms. Let $l_i^p(\mathbf{q})$ and $l_i^u(\mathbf{q})$ denote the congestion cost per service incurred by the service provider i and his users respectively. Denote $l_i(\mathbf{q}) = l_i^p(\mathbf{q}) + l_i^u(\mathbf{q})$ as the aggregate congestion cost associated with a service. Thus, when service provider i enrolls q_i users, the aggregate congestion cost associated with i and his users is $l_i(\mathbf{q})q_i$. Let $\mathbf{l}(\mathbf{q}) = (l_1(\mathbf{q}), \cdots, l_n(\mathbf{q}))$, then the total congestion cost in the system is given by $\mathbf{q}^T\mathbf{l}(\mathbf{q})$.

Assumption 4.2.2 *For every* i*, the cost function* $l_i(\mathbf{q})$ *is convex, continuous, non-decreasing, and continuously differentiable for every component of* $\mathbf{q}$.

Denote the Jacobian matrix of the cost function $\mathbf{l}(\mathbf{q})$ as $\mathbf{R}(\mathbf{q})$, where

$$\mathbf{R}(\mathbf{q}) = \begin{bmatrix} \partial l_1/\partial q_1 & \cdots & \partial l_1/\partial q_n \\ \vdots & \ddots & \vdots \\ \partial l_n/\partial q_1 & \cdots & \partial l_n/\partial q_n \end{bmatrix}.$$

Except with affine costs, the Jacobian matrix $\mathbf{R}(\mathbf{q})$ depends on the value of $\mathbf{p}$. For the ease of notation, we will write $\mathbf{R}$ instead of $\mathbf{R}(\mathbf{q})$ for the rest of this chapter. We will use superscripts to differentiate the matrix evaluated at different values. This assumption states that $\mathbf{R}$ is a nonnegative (componentwise) and positive semi-definite matrix. Denote the diagonal part of matrix $\mathbf{R}$ and its off-diagonal part as $\mathbf{\Gamma_R}$ and $\mathbf{R_{off}}$ respectively. Denote the $(i,j)$the element of the matrix $\mathbf{R}$ as r_{ij}.

In this work, we distinguish two types of costs which are formally defined as follows.

Definition 4.2.3 ***Self-contained cost***, *$\partial l_i/\partial q_i > 0$.* ***Spillover cost***, *$\partial l_i/\partial q_j \geq 0$ for all $j \neq i$.*

When costs are fully self-contained, the cost associated with service i only depends on i's output level and independent of others. It implies that the Jacobian matrix is simply a diagonal matrix, i.e., $\mathbf{R} = \mathbf{\Gamma_R}$. On the other hand, when spillover cost is also present, one's cost increases whenever there is an increase in the system output.

Assumption 4.2.4 *The maximum reservation price must satisfy, $\bar{\mathbf{p}} - \mathbf{l}(\mathbf{0}) > \mathbf{0}$.*

The assumption states that the maximum profit per service must be positive. If this assumption is violated, it implies that no user is willing to pay for the service and this corresponding "inactive" service provider could be removed from the equilibrium.

When spillover cost exists, we make use of the following assumption in our analysis.

Assumption 4.2.5 *The Jacobian matrix of the cost function $\mathbf{R}$ is symmetric.*

This assumption requires $\partial l_i/\partial q_j = \partial l_j/\partial q_i$ for all $j \neq i$. One justification is that if all users are homogenous in the sense that enrolling an additional user creates the same amount of additional cost to others, then this assumption holds. The assumption simplifies the derivation in the analysis as $\mathbf{R} = \mathbf{R}^T$, or $r_{ij} = r_{ji}$. It can be relaxed by considering its "symmetrized" matrix with a norm which measures the asymmetry as discussed in Sun (2006).

To establish bounds when costs are nonlinear, it is key to introduce a constant that will measure the degree of nonlinearity of the cost function. As a result, we briefly introduce the concept of Jacobian similarity. We refer rthe eader to Perakis (2007) for more information on this concept.

Definition 4.2.6 ***The Jacobian similarity property.*** *A positive semidefinite matrix* $\mathbf{F}(\mathbf{q})$ *satisfies the matrix similarity property if there exists a constant* $\kappa \geq 1$ *such that for all* $\mathbf{w}, \mathbf{q}$ *and* $\mathbf{q}'$: $\kappa \mathbf{w}^T(\mathbf{F}(\mathbf{q}))\mathbf{w} \geq \mathbf{w}^T(\mathbf{F}(\mathbf{q}'))\mathbf{w} \geq \frac{1}{\kappa}\mathbf{w}^T(\mathbf{F}(\mathbf{q}))\mathbf{w}$.

Note if $\mathbf{F}(\mathbf{q})$ is the Jacobian matrix an affine function, then $\kappa = 1$. Notice that constant κ is easy to compute when matrix $\mathbf{F}(\mathbf{q})$ is positive definite for all $\mathbf{q}$. In that case, the constant $\kappa = \frac{\max_{\mathbf{q}} \max_i \lambda(\mathbf{F}(\mathbf{q}))}{min_{\mathbf{q}} \min_i \lambda(\mathbf{F}(\mathbf{q}))}$, that is, the ratio between the maximum and minimum eigenvalue of the matrix.

To interpret the bounds established in this chapter, we will introduce two notions. They measure the level of competition and the extent of spillover respectively.

Definition 4.2.7 ***Competition index adjusted with self-contained cost.*** *Given a price sensitivity matrix* $\mathbf{B}$ *and self-contained cost* $\mathbf{\Gamma_R}$*, the competition index for service provider* i *is defined as* $\gamma_i = \sum_{j \neq i} \beta_{ij}/\beta_{ii} + 2r_{ii}/\beta_{ii}$*, for all* i*. Let* $\bar{\gamma} = \max_i \gamma_i$.

The notion that $\sum_{j \neq i} \beta_{ij}/\beta_{ii}$ is used to measure the intensity of competition (see Sun 2006, Farahat and Perakis 2010a,b). Suppose every service provider in the market changes his output level by 1 unit, β_{ii} reflects the amount of price change which is solely contributed by i's own output change, while $\sum_{j \neq i} \beta_{ij}$ measures the price change contributed by i's competitors. A high value of $\sum_{j \neq i} \beta_{ij}/\beta_{ii}$ suggests that i's price is more susceptible to his competitors' output change than his own change, implying that service provider i faces a high level of competition. When $\sum_{j \neq i} \beta_{ij} = 0$ for all i, it implies that $\beta_{ij} = 0$ for all $j \neq i$. That is, each of service provider acts as a monopolist and does not face any competition.

When there is self-contained cost (i.e., $r_{ii} \geq 0$), the competition index also contains the term r_{ii}/β_{ii}. It compares the marginal decrease in the revenue per service due to one's own cost increase to the decrease due to diminishing returns of the demand. Thus, when r_{ii}/β_{ii} is large, it implies that the cost increase is rather steep.

Thus, γ_i meausres the level of competition faced by service provider i, taking into account the self-contained cost, i.e., comparing the aggregate price impact from i's competitors and the self-contained congestion to the price change solely contributed by i's output change. With asymmetric service providers, γ_i differs across i. We will use $\bar{\gamma} = \max_i \gamma_i$ to approximate the competition intensity in the market.

Definition 4.2.8 ***Marginal net externality.*** *Define* $\rho_i = \sum_{j \neq i} r_{ji}/\beta_{ii}$, *for all* i. *Let* $\bar{\rho} = \max_i \rho_i$.

This ratio is nonzero only when spillover cost is present. The numerator $\sum_{j \neq i} r_{ji}$ captures the marginal spillover cost created by service provider i, and the denominator reflects the additional welfare increase to users when service provider i increases his output. Thus, the term ρ_i could be interpreted as the *external cost* (spillover congestion cost) versus the *external benefit* (additional consumer surplus) which service provider i brings to the society. $\bar{\rho}$ represents the value of the "worst offender" among all the service providers by taking the highest value of ρ_i. When $\bar{\rho} \leq 1$, it implies net positive externality, that is, after taking into account of spillover cost, every service provider is contributing more welfare to the society than the cost. When $\bar{\rho} > 1$, there exists some service provider whose spillover cost outweighs the welfare he brings to the society, thus, the net externality is negative.

4.2.1 User behavior

Given n differentiated services available, a representative user derives a different marginal utility $u_i(\mathbf{q})$ for services with provider i. A user's total disutility is the sum of the price he is charged for the service and the congestion cost he experiences, i.e., $p_i + l_i^u(\mathbf{q})$. This term is known as *full price* or *effective price* in the literature (e.g., Weintraub et al. 2010, Acemoglu and Ozdaglar 2007).

We assume that each user is "small" compared to the total traffic volume in the sense that when he switches from a service provider to another, there is no considerable change in the congestion cost. We use the General Wardrop equilibrium principle with multiple commodity flows to model user behavior in the presence of differentiated services. For a given price vector $\mathbf{p}$, a vector of service level $\mathbf{q}$ is a *General Wardrop equilibrium* (GWE) if

$$\begin{aligned} p_i + l_i^u(\mathbf{q}) &= u_i(\mathbf{q}), \quad \text{for all } i \text{ if } q_i > 0; \\ p_i + l_i^u(\mathbf{q}) &\geq u_i(\mathbf{q}), \quad \text{for all } i \text{ if } q_i = 0. \end{aligned}$$

The equilibrium condition states that for any price vector $\mathbf{p}$, the full price of a user with any active service provider, must be equal to the corresponding marginal utility obtained in equilibrium. Without loss of generality, we only restrict our attention to these active service providers with

$$p_i = u_i(\mathbf{q}) - l_i^u(\mathbf{q}), \quad \text{for all } i \text{ if } q_i > 0. \tag{4.1}$$

When there is no congestion, the market clearing price of any service is simply its marginal utility, i.e., $p_i = u_i(\mathbf{q})$. The presence of congestion directly lowers a user's willingness to pay for the service, and consequently reduces the profitability of the service providers.

Remark In a setting with a single type of service or symmetric service providers, GWE implies that a *single* full price prevails in an equilibrium. With differentiated services, different full price values exist as the service providers leverage product differentiation to capture consumer surplus. An ubiquitous observation is how airfares for the same trip within the same class vary across airlines.

4.2.2 Service provider's profit maximization problem

We define the service provider i's profit function as $\pi_i(q_i, \mathbf{q_{-i}}) = q_i(p_i(q_i, \mathbf{q_{-i}}) - t_i - l_i^p(q_i, \mathbf{q_{-i}}))$, where $\mathbf{q_{-i}}$ are the service levels set by i's competitors and t_i is the access fee per service imposed by the facility manager. Without loss of generality, we assume the unit cost of providing a service is zero. By Equation (4.1) which describes the users' behavior, we obtain the following, $\pi_i(q_i, \mathbf{q_{-i}}) = (u_i(q_i, \mathbf{q_{-i}}) - t_i - l_i(q_i, \mathbf{q_{-i}}))q_i$.

The dynamics between the facility manager and the service providers are modeled as a Stackelberg game. The facility manager, announces the access fee per service, t_i. The service provider then determines his appropriate output level. We assume service providers behave according to a *Subgame Perfect Equilibrium*, that is, for every access fee t_i, every service provider i determines his service level to maximize his profit, given the service level set by his competitors. Under Assumptions 4.2.1 to 4.2.5, the profit function of all service providers are diagonally strictly concave over the strategy space, thus, the existence and uniqueness of the equilibrium are guaranteed.

4.2.3 Facility manager's welfare maximization problem

The goal of the facility manager is to maximize the total social welfare (W), which is defined as the sum of consumer surplus (CS), producer surplus (PS) and revenue collected from acess fee (TR), i.e., $W = CS + PS + TR$. Consumer surplus is defined as the difference between the total utility derived from consuming $\mathbf{q}$ units of services and the total cost incurred by users. With affine marginal utility function, the total utility is given by $\int_{\mathbf{q}} \mathbf{u(x)dx} = \mathbf{q}^T(\bar{\mathbf{p}} - \frac{1}{2}\mathbf{Bq})$. The full price is given by $(\mathbf{p} + \mathbf{l}^u(\mathbf{q}))^T\mathbf{q}$. By Equation (4.1), full price in equilibrium is simply $\mathbf{u}(\mathbf{q})^T\mathbf{q}$. Thus, $CS = \mathbf{q}^T(\bar{\mathbf{p}} - \frac{1}{2}\mathbf{Bq}) - \mathbf{q}^T(\bar{\mathbf{p}} - \mathbf{Bq}) = \frac{1}{2}\mathbf{q^TBq}$. The producer surplus is the total profit generated by all service providers, i.e., $PS = \sum_i \pi_i(\mathbf{q}) = \mathbf{q}^T(\mathbf{u}(\mathbf{q}) - \mathbf{l}(\mathbf{q}) - \mathbf{t}) = \mathbf{q}^T(\bar{\mathbf{p}} - \mathbf{Bq} - \mathbf{l}(\mathbf{q}) - \mathbf{t})$. The revenue collected from congestion pricing is captured by $TR = \mathbf{q}^T\mathbf{t}$. Combine all three terms, the total welfare is given by the following:

$$W(\mathbf{q}) = \mathbf{q}^T(\bar{\mathbf{p}} - \frac{1}{2}\mathbf{Bq} - \mathbf{l}(\mathbf{q})). \tag{4.2}$$

The welfare maximization problem (4.2) is a strictly concave optimization problem in terms of the output level $\mathbf{q}$, where $\mathbf{q}^* = arg\max_{\mathbf{q}} W(\mathbf{q})$. To achieve the maximum welfare $W(\mathbf{q}^*)$ in the three-level model, the facility manager can use the access fees such that the desirable service level $\mathbf{q}^*$ is achieved, i.e., $\mathbf{q}(\mathbf{t}^*) = \mathbf{q}^*$. Therefore, the access fee could be viewed as a "coordinating contract" that aligns the profit-maximizing objective of service providers to one that maximizes the societal welfare.

Proposition 4.2.9 *For a given output level* $\mathbf{q}$ *in the facility, the access fee per service* i *is given by* $t_i(\mathbf{q}) = -\beta_{ii}q_i + \sum_{j\neq i} \partial l_j/\partial q_i$.

In the absence of spillover costs, the second term vanishes, i.e., $t_i(\mathbf{q}) = -\beta_{ii}q_i$. Since $\beta_{ii} > 0$, the access fee t_i is negative in this case. It implies that the facility manager has to give a subsidy to every service provider i in order to ensure the maximum societal welfare. In the unregulated setting (i.e., $\mathbf{t} = 0$), one can show that the output level generated by the profit-maximizing service providers is *always below* the socially optimal level. With the the subsidy, lower prices induce more users to acquire the services, leading to higher welfare.

With spillover cost, the second term in the access fee holds service providers accountable for the "spillover" they have imposed onto others. Therefore, the access fee has to balance the size of the subsidy which encourages production and the penalty to discourage "tragedy of the commons" phenomenon.

Proposition 4.2.9 captures a sharp contrast to the traditional road toll imposed on road users in two distinct aspects. Firstly, in the road traffic model, the congestion pricing is the full marginal cost, whereas an oligopolistic service provider is only penalized for the marginal cost which he has not internalized (the spillover). Secondly, the optimal access fee considered here also includes a subsidy which encourages production. As a result, it suggests that the optimal access fee charged to an oligopolist would be much smaller than the traditional "toll" with the same cost function.

Relate the result to the wireless service industry where costs are fully contained, Proposition 4.2.9 suggests that FCC should provide a subsidy so that more users will acquire the services. On the other hand, the FAA (or the airport regulator) should hold the airlines responsible for the additional delays which they impose onto others. Moreover, the access fees could provide some incentives to revert the trend of the decreasing aircraft size[1]. To model this, we incorporate a parameter μ_i in

[1]Whalen et al. (2008) have shown that from 1997 to 2007, the number of departures has skyrock-

our formulation which captures the size of an aircraft that airline i uses, and each airline adjusts its frequency of the flights, q_i/μ_i, to maximize the profit, where q_i is the total number of passengers to be transported[2]. The access fee per flight takes the form $t_i(\mathbf{q}) = -\beta_{ii}q_i + \sum_{j\neq i} \partial l_j(\mathbf{q})/\partial q_i/\mu_i$. If all other things being equal, an airline with a larger aircraft will pay a lower fee. For an industry with a measly profit margin of 0.7% in 2011 as predicted by the International Air Transport Association, a landing fee based on the aircraft's size would provide airlines with more incentives to use large aircrafts, which in turn ensures a more efficient use of the limited runways and alleviates congestion. Comparing this scheme to the current practice where the landing fees are based on the weight of an aircraft, the latter gives exactly the wrong incentives for smaller aircrafts, which exacerbates the congestion problem in many airports.

4.3 Efficiency Analysis

In this section, we will compare the social welfare achieved in the unregulated setting with the social optimum. By doing so, it enables us to address a question that is important to both theory and practice. That is, what is the maximum welfare loss due to the lack of regulation in an oligopoly? The answer to this question helps policy makers to gauge the need for regulation. While almost all policy changes have to deal with huge political and financial challenges, if one can show that the loss of welfare is significant in the unregulated setting, it provides one concrete evidence to concrete support the implementation of new policies such as the access fees discussed in the previous section.

Moreover, the comparative analysis also helps us isolate some key factors which affect the efficiency of such an oligopoly. In many scenarios, "optimum" might be infeasible to attain and it is just as important to determine what kind of actions that

eted by 35% while the total number of seats has risen by less than 6%, implying that a dramatic decrease in the number of seats per aircraft.

[2]To reduce notation and enhance the transparency of the model, we only discuss the case where each airline only uses a single type of aircraft. The case with multiple sizes of aircrafts could be easily incorporated.

will improve the efficiency.

Denote $\mathbf{q^N}$ and $\mathbf{q}^*$ as the output level in the unregulated setting and that in a social optimum respectively. Let $W(\mathbf{q^N})$ and $W(\mathbf{q}^*)$ be the the total welfare attained in these two settings. The quantity of interest is $W(\mathbf{q^N})/W(\mathbf{q}^*)$. We will address the settings with only self-contained cost and with spillover separably. We will first establish a lower bound on this quantity which gives the worst performance guarantee, followed by a upper bound.

With nonlinear convex costs in a model, it is generally hard to obtain closed-form solutions. Nonetheless, we can derive optimality conditions that can be used to quantify the total welfare. Since the Jacobian matrix on cost function depends on the output level, we will use $\mathbf{R}^*$ and $\mathbf{R^N}$ to distinguish the matrix evaluated at $\mathbf{q}^*$ and $\mathbf{q^N}$ respectively.

Lemma 4.3.1 *Under Assumptions 4.2.1 to 4.2.5, there exists a unique solution to both the coordinated and unregulated problems. In particular, the welfare generated in the social optimal setting and the unregulated setting are given by*

$$W(\mathbf{q}^*) = (\mathbf{q}^*)^T(\frac{1}{2}\mathbf{B} + \mathbf{R}^*)\mathbf{q}^*, \quad \textit{and}$$
$$W(\mathbf{q^N}) = (\mathbf{q^N})^T(\frac{1}{2}\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^N})\mathbf{q^N}.$$

4.3.1 Lower bound on $W(\mathbf{q^N})/W(\mathbf{q}^*)$

We begin with a lemma that compares the output level in the unregulated setting and the social optimum by making use of the convexity of the cost function and the optimality conditions derived in Lemma 4.3.1.

Lemma 4.3.2 *When cost* $\mathbf{l}(\mathbf{q}) = (l_1(\mathbf{q}), \cdots, l_n(\mathbf{q}))$ *is a convex function (componentwise),*

$$(\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^N} + \mathbf{R^N})\mathbf{q^N} \geq (\mathbf{B} + \mathbf{R}^* + \mathbf{R^N})\mathbf{q}^*.$$

Remark When the cost function $\mathbf{l}(\mathbf{q})$ is an affine function of $\mathbf{q}$, then $\mathbf{R}$ is independent from the output level $\mathbf{q}$ and Lemma 4.3.2 becomes an equality.

Now we are presenting the first main result in this chapter. We establish a constant lower bound on the efficiency of an unregulated oligopoly with self-contained cost.

Theorem 4.3.3 *When there is only self-contained congestion, under Assumptions 4.2.1 to 4.2.4, the welfare in the unregulated setting with nonlinear convex cost is at least 3/4 of the optimal welfare, i.e.,* $\frac{W(\mathbf{q}^{\mathrm{N}})}{W(\mathbf{q}^*)} \geq \frac{3}{4}$. *The bound is tight when service providers are noncompeting and have a constant marginal cost.*

Proof of Theorem 4.3.3. With Lemma B.5.1 and Lemma 4.3.1, we have a lower bound on $W(\mathbf{q}^{\mathrm{N}})/W(\mathbf{q}^*)$:

$$\frac{W(\mathbf{q}^{\mathrm{N}})}{W(\mathbf{q}^*).} \geq \frac{(\mathbf{q}^{\mathrm{N}})^T(\mathbf{B}+2\mathbf{\Gamma_B}+2\mathbf{\Gamma_R^N})\mathbf{q}^{\mathrm{N}}}{(\mathbf{q}^{\mathrm{N}})^T(\mathbf{B}+\mathbf{\Gamma_B}+\mathbf{\Gamma_R^N}+\mathbf{R}^{\mathrm{N}})(\mathbf{B}+2\mathbf{R}^{\mathrm{N}})^{-1}(\mathbf{B}+\mathbf{\Gamma_B}+\mathbf{\Gamma_R^N}+\mathbf{R}^{\mathrm{N}})\mathbf{q}^{\mathrm{N}}}. \tag{4.3}$$

In the absence of spillover, $\mathbf{R}^{\mathrm{N}} = \mathbf{\Gamma_R^N}$. Since all quantities are the Nash equilibrium quantities in this proof, we drop the superscript on matrices for simplicity:

$$\frac{W(\mathbf{q}^{\mathrm{N}})}{W(\mathbf{q}^*)} \geq \frac{(\mathbf{q}^{\mathrm{N}})^T(\mathbf{B}+2\mathbf{\Gamma_R}+2\mathbf{\Gamma_B})\mathbf{q}^{\mathrm{N}}}{(\mathbf{q}^{\mathrm{N}})^T(\mathbf{B}+\mathbf{2\Gamma_R}+\mathbf{\Gamma_B})(\mathbf{B}+2\mathbf{\Gamma_R})^{-1}(\mathbf{B}+\mathbf{2\Gamma_R}+\mathbf{\Gamma_B})\mathbf{q}^{\mathrm{N}}}.$$

Denote $\mathbf{G} = \mathbf{\Gamma_B^{-0.5}}\tilde{\mathbf{B}}\mathbf{\Gamma_B^{-0.5}} = \mathbf{\Gamma_B^{-0.5}}(\mathbf{B}+\mathbf{2\Gamma_R})\mathbf{\Gamma_B^{-0.5}}$. This is a symmetric, nonnegative positive definite matrix. We will first rewrite the right hand side of the inequality in terms of matrix $\mathbf{G}$ and identity matrix $\mathbf{I}$:

$$\begin{aligned}
&\frac{(\mathbf{q}^{\mathrm{N}})^T(\tilde{\mathbf{B}}+2\mathbf{\Gamma_B})\mathbf{q}^{\mathrm{N}}}{(\mathbf{q}^{\mathrm{N}})^T(\tilde{\mathbf{B}}+\mathbf{\Gamma_B})\tilde{\mathbf{B}}^{-1}(\tilde{\mathbf{B}}+\mathbf{\Gamma_B})\mathbf{q}^{\mathrm{N}}}\\
=&\frac{(\mathbf{q}^{\mathrm{N}})^T\mathbf{\Gamma_B^{0.5}}\mathbf{\Gamma_B^{-0.5}}(\tilde{\mathbf{B}}+2\mathbf{\Gamma_B})\mathbf{\Gamma_B^{-0.5}}\mathbf{\Gamma_B^{0.5}}\mathbf{q}^{\mathrm{N}}}{(\mathbf{q}^{\mathrm{N}})^T\mathbf{\Gamma_B^{0.5}}\mathbf{\Gamma_B^{-0.5}}(\tilde{\mathbf{B}}+\mathbf{\Gamma_B})\tilde{\mathbf{B}}^{-1}(\tilde{\mathbf{B}}+\mathbf{\Gamma_B})\mathbf{\Gamma_B^{-0.5}}\mathbf{\Gamma_B^{0.5}}\mathbf{q}^{\mathrm{N}}}\\
=&\frac{(\mathbf{q}^{\mathrm{N}})^T\mathbf{\Gamma_B^{0.5}}(\mathbf{G}+2\mathbf{I})\mathbf{\Gamma_B^{0.5}}\mathbf{q}^{\mathrm{N}}}{(\mathbf{q}^{\mathrm{N}})^T\mathbf{\Gamma_B^{0.5}}(\mathbf{G}+\mathbf{I})\mathbf{G}^{-1}(\mathbf{G}+\mathbf{I})\mathbf{\Gamma_B^{0.5}}\mathbf{q}^{\mathrm{N}}}\\
=&\frac{(\mathbf{q}^{\mathrm{N}})^T\mathbf{\Gamma_B^{0.5}}\mathbf{G}^{-0.5}\mathbf{G}^{0.5}(\mathbf{G}+2\mathbf{I})\mathbf{G}^{0.5}\mathbf{G}^{-0.5}\mathbf{\Gamma_B^{0.5}}\mathbf{q}^{\mathrm{N}}}{(\mathbf{q}^{\mathrm{N}})^T\mathbf{\Gamma_B^{0.5}}\mathbf{G}^{-0.5}\mathbf{G}^{0.5}(\mathbf{G}+\mathbf{I})\mathbf{G}^{-1}(\mathbf{G}+\mathbf{I})\mathbf{G}^{0.5}\mathbf{G}^{-0.5}\mathbf{\Gamma_B^{0.5}}\mathbf{q}^{\mathrm{N}}}
\end{aligned}$$

$$=\frac{\mathbf{w}^T\mathbf{G}^{0.5}(\mathbf{G}+2\mathbf{I})\mathbf{G}^{0.5}\mathbf{w}}{\mathbf{w}^T\mathbf{G}^{0.5}(\mathbf{G}+\mathbf{I})\mathbf{G}^{-1}(\mathbf{G}+\mathbf{I})\mathbf{G}^{0.5}\mathbf{w}}, \qquad \text{where } \mathbf{w}=\mathbf{G}^{-0.5}\mathbf{\Gamma}_{\mathbf{B}}^{0.5}\mathbf{q}^{\mathbf{N}}.$$

If this ratio is at least 3/4, it implies that

$$\begin{aligned} 4\mathbf{w}^T\mathbf{G}^{0.5}(\mathbf{G}+2\mathbf{I})\mathbf{G}^{0.5}\mathbf{w}-3\mathbf{w}^T\mathbf{G}^{0.5}(\mathbf{G}+\mathbf{I})\mathbf{G}^{-1}(\mathbf{G}+\mathbf{I})\mathbf{G}^{0.5}\mathbf{w} &\geq 0, \text{ or equivalently,} \\ \mathbf{w}^T\left(4\mathbf{G}^{0.5}(\mathbf{G}+2\mathbf{I})\mathbf{G}^{0.5}-3\mathbf{G}^{0.5}(\mathbf{G}+\mathbf{I})\mathbf{G}^{-1}(\mathbf{G}+\mathbf{I})\mathbf{G}^{0.5}\right)\mathbf{w} &\geq 0. \end{aligned} \tag{4.4}$$

To prove this statement, we will use the fact that for any given vector $\mathbf{x}$, if the matrix $\mathbf{A}$ is nonnegative component-wise, then $\mathbf{x}^T\mathbf{A}\mathbf{x}\geq 0$. To show the composite matrix in Equation (4.4) is nonnegative, we express it as follows,

$$\begin{aligned} & 4\mathbf{G}^{0.5}(\mathbf{G}+2\mathbf{I})\mathbf{G}^{0.5}-3\mathbf{G}^{0.5}(\mathbf{G}+\mathbf{I})\mathbf{G}^{-1}(\mathbf{G}+\mathbf{I})\mathbf{G}^{0.5} \\ = {} & 4(\mathbf{G^2}+\mathbf{2G})-3(\mathbf{G}-\mathbf{I})^2 \\ = {} & \mathbf{4G^2}+\mathbf{8G}-\mathbf{3G^2}-\mathbf{6G}-\mathbf{3I} \\ = {} & \mathbf{G^2}+\mathbf{2G}-\mathbf{3I} \\ = {} & (\mathbf{G}-\mathbf{I})(\mathbf{G}+\mathbf{3I}). \end{aligned} \tag{4.5}$$

Given matrix $\mathbf{G}$ is a nonnegative matrix, the second term which is a sum with an identity matrix is clearly nonnegative. Now consider the first term, $\mathbf{G}-\mathbf{I}$. The off-diagonal elements $\beta_{ij}/(\sqrt{\beta_{ii}\beta_{ij}})$ are nonnegative under Assumption 4.2.2, where β_{ij} is the ij-th element of matrix $\mathbf{B}$. Its diagonal elements are given by $(\beta_{ii}+2r_{ii})/\beta_{ii}=1+2r_{ii}/\beta_{ii}\geq 1$, where r_{ii} is the i-th diagonal element of the Jacobian matrix $\mathbf{R}$. Therefore, $\mathbf{G}-\mathbf{I}$ must also be nonnegative. We have shown that the composite matrix in Equation (4.4) could be expressed a product of two nonnegative matrices. Therefore, this composite matrix must also be nonnegative and we have shown that the statement in Equation (4.4) holds true and the ratio is bounded from below by 3/4.

Lastly, to show that the bound is tight, note that with noncompeting service providers, $\mathbf{B}=\mathbf{\Gamma_B}$. Since $\mathbf{l}(\mathbf{q})$ is a vector independent of $\mathbf{q}$, by Equation (B.1) and (B.3), we see that $\mathbf{q}^{\mathbf{N}}=\frac{1}{2}\mathbf{q}^*$. Substituting this condition into the welfare function derived in Lemma 4.3.1, it is straightforward to show that the ratio is exactly 3/4. □

This result is surprising, especially when it is compared to other works which have analyzed performance degradation caused by selfish behavior. Roughgarden and Tardos (2002) and Roughgarden (2005) consider the selsh behavior of noncooperative network users. The authors prove that if the latency of each edge is an affine function of its traffic, then the total latency of the routes chosen by selsh network users is at most 4/3 times the minimum possible total latency. They also show that performance degrades as the degree of the latency functions increases. We can cast our model to show the "similarities" between our model and those works: a network with n routes, each owned by a profit-maximizing provider and each offers a different marginal utility. Users choose which route to take based on a full price (the price paid to travel and disutility $l_i(q_i)$, where q_i is the traffic on route i). Theorem 4.3.3 quantifies the welfare degradation (producer surplus and consumer surplus) caused by the selfish behavior of noncooperative providers and users to the maximum possible welfare. In our setting, we have shown that regardless of the nonlinearity on the latency function, a constant bound of 3/4 is achieved.

In the next theorem, we present the efficiency analysis for an unregulated oligopoly in the presence of spillover cost. Let $\bar{\rho}$ be the cost-to-benefit ratio evaluated at the social optimal output level, i.e., $\bar{\rho} = \max_i \sum_{j \neq i} r^*_{ji}/\beta_{ii}$, where $r^*_{ji} = \partial l_j / \partial q_i|_{\mathbf{q}=\mathbf{q}^*}$. We are going to show in the next theorem that in contrast to the case where costs are fully internalized, performance degradation with spillover costs can be unbounded.

Theorem 4.3.4 *With spillover congestion cost, Under Assumption 4.2.1 to 4.2.5, the welfare in the unregulated setting depends on the the maximum spillover cost-to-benefit ratio $\bar{\rho}$ and can be bounded by:*
When $\bar{\rho} \leq 1$, $\frac{W(\mathbf{q^N})}{W(\mathbf{q^*})} \geq \frac{3}{4\kappa'}$;
When $\bar{\rho} > 1$, $\frac{W(\mathbf{q^N})}{W(\mathbf{q^*})} \geq \frac{1}{\kappa'}\left(1 - (\frac{\bar{\rho}-1}{\bar{\rho}+1})^2\right)$, *where κ' is the Jacobian similarity factor.*

Proof of Theorem 4.3.4. From Lemma 4.3.2, we have $(\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma^N_R} + \mathbf{R^N})\mathbf{q^N} \geq (\mathbf{B} + \mathbf{R^*} + \mathbf{R^N})\mathbf{q^*}$. Denote $\mathbf{\Sigma} = \mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma^N_R} + \mathbf{R^N}$ and $\mathbf{\Phi} = \mathbf{B} + \mathbf{R^*} + \mathbf{R^N}$, thus,

$\mathbf{\Sigma q^N} \geq \mathbf{\Phi q^*}$. Note the vectors on both sides of the inequality are nonnegative. By Lemma 4.3.1, we obtain the following,

$$W(\mathbf{q^N}) \ = (\mathbf{q^N})^T(\tfrac{1}{2}\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^N})\mathbf{q^N} = (\mathbf{q^N})^T\mathbf{\Sigma\Sigma}^{-1}(\tfrac{1}{2}\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^N})\mathbf{\Sigma}^{-1}\mathbf{\Sigma q^N}.$$

Replacing $\mathbf{\Sigma q^N}$ with $\mathbf{\Phi q^*}$, we obtain a lower bound on $W(\mathbf{q^N})$, i.e.,

$$\begin{aligned}
&W(\mathbf{q^N})\\
\geq&(\mathbf{q^*})^T\mathbf{\Phi\Sigma}^{-1}(\frac{1}{2}\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^N})\mathbf{\Sigma}^{-1}\mathbf{\Phi q^*}\\
=&(\mathbf{q^*})^T(\mathbf{B} + \mathbf{R^*} + \mathbf{R^N})(\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^N} + \mathbf{R^N})^{-1}(\frac{1}{2}\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^N})\\
&(\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^N} + \mathbf{R^N})^{-1}(\mathbf{B} + \mathbf{R^*} + \mathbf{R^N})\mathbf{q^*}.
\end{aligned}$$

By making using of the Jacobian similarity properties on matrices$\mathbf{\Gamma_R^N}$ and $\mathbf{R^N}$, there exists κ' such that

$$\begin{aligned}
&W(\mathbf{q^N})\\
\geq&\frac{1}{\kappa'}(\mathbf{q^*})^T(\mathbf{B} + \mathbf{R^*} + \mathbf{R^*})(\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^*} + \mathbf{R^*})^{-1}(\frac{1}{2}\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^*})\\
&(\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^*} + \mathbf{R^*})^{-1}(\mathbf{B} + \mathbf{R^*} + \mathbf{R^*})\mathbf{q^*}\\
=&\frac{1}{\kappa'}(\mathbf{q^*})^T(\mathbf{B} + 2\mathbf{R^*})(\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^*} + \mathbf{R^*})^{-1}(\frac{1}{2}\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^*})\\
&(\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^*} + \mathbf{R^*})^{-1}(\mathbf{B} + 2\mathbf{R^*})\mathbf{q^*}.
\end{aligned}$$

Note that by using the definition of the minimum eigenvalue of a positive semidefinite matrix, $1/\kappa'$ can be bounded as follows,

$$\begin{aligned}
\frac{1}{\kappa'} \geq&\lambda_{\min}\{(\mathbf{B} + \mathbf{R^*} + \mathbf{R^N})^2(\mathbf{B} + 2\mathbf{R^*})^{-2}(\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^N} + \mathbf{R^N})^{-2}(\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^*} + \mathbf{R^*})^2\\
&(\frac{1}{2}\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^N})(\frac{1}{2}\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^*})^{-1}\}.
\end{aligned}$$

If costs are affine, i.e, $\mathbf{R}^N = \mathbf{R^*}$, then $\kappa' = 1$. Combine the result with $W(\mathbf{q^*})$, notice that it is all in the optimal $\mathbf{q^*}$ space, therefore we will drop the superscript in this proof. Denote $\tilde{\mathbf{B}} = \mathbf{B} + 2\mathbf{\Gamma_R}$,

$$\begin{aligned}
&\frac{W(\mathbf{q^N})}{W(\mathbf{q^*})} \\
\geq&\frac{1}{\kappa'}\frac{(\mathbf{q}^*)^T(\tilde{\mathbf{B}}+2\mathbf{R_{off}})(\tilde{\mathbf{B}}+\mathbf{\Gamma_B}+\mathbf{R_{off}})^{-1}(\tilde{\mathbf{B}}+2\mathbf{\Gamma_B})(\tilde{\mathbf{B}}+\mathbf{\Gamma_B}+\mathbf{R_{off}})^{-1}(\tilde{\mathbf{B}}+2\mathbf{R_{off}})\mathbf{q}^*}{(\mathbf{q}^*)^T(\tilde{\mathbf{B}}+2\mathbf{R_{off}})\mathbf{q}^*}.
\end{aligned}$$

Denote $\mathbf{G} = \mathbf{\Gamma_B^{-0.5}}\tilde{\mathbf{B}}\mathbf{\Gamma_B^{-0.5}}$ and $\mathbf{\Xi} = \mathbf{\Gamma_B^{-0.5}}\mathbf{R_{off}}\mathbf{\Gamma_B^{-0.5}}$, the expression becomes

$$\begin{aligned}
&\frac{W(\mathbf{q^N})}{W(\mathbf{q^*})} \\
\geq&\frac{1}{\kappa'}\frac{(\mathbf{q}^*)^T\mathbf{G}^{0.5}(\mathbf{G}+2\mathbf{\Xi})(\mathbf{G}+\mathbf{I}+\mathbf{\Xi})^{-1}(\mathbf{G}+2\mathbf{I})(\mathbf{G}+\mathbf{I}+\mathbf{\Xi})^{-1}(\mathbf{G}+2\mathbf{\Xi})\mathbf{G}^{0.5}\mathbf{q}^*}{(\mathbf{q}^*)^T\mathbf{G}^{0.5}(\mathbf{G}+2\mathbf{\Xi})\mathbf{G}^{0.5}\mathbf{q}^*}.
\end{aligned}$$

Using the Rayleigh-Ritz Theorem, the lower bound is given by the minimum eigenvalue of the following composite matrix,

$$\begin{aligned}
&\frac{W(\mathbf{q^N})}{W(\mathbf{q^*})} \\
\geq&\frac{1}{\kappa'}\lambda_{\min}\{(\mathbf{G}+2\mathbf{I})(\mathbf{G}+\mathbf{I}+\mathbf{\Xi})^{-1}(\mathbf{G}+2\mathbf{\Xi})(\mathbf{G}+\mathbf{I}+\mathbf{\Xi})^{-1}\} \\
=&\frac{1}{\kappa'}\lambda_{\min}\{(\mathbf{G}+\mathbf{I}+\mathbf{\Xi}+\mathbf{I}-\mathbf{\Xi})(\mathbf{G}+\mathbf{I}+\mathbf{\Xi})^{-1}(\mathbf{G}+\mathbf{\Xi}+\mathbf{I}-\mathbf{I}+\mathbf{\Xi})(\mathbf{G}+\mathbf{I}+\mathbf{\Xi})^{-1}\} \\
=&\frac{1}{\kappa'}\lambda_{\min}\{(\mathbf{I}+(\mathbf{I}-\mathbf{\Xi})(\mathbf{G}+\mathbf{I}+\mathbf{\Xi})^{-1})(\mathbf{I}-(\mathbf{I}-\mathbf{\Xi})(\mathbf{G}+\mathbf{I}+\mathbf{\Xi})^{-1})\} \\
=&\frac{1}{\kappa'}\lambda_{\min}\{\mathbf{I}-((\mathbf{I}-\mathbf{\Xi})(\mathbf{G}+\mathbf{I}+\mathbf{\Xi})^{-1}))^2\} \\
=&\frac{1}{\kappa'}\left(1-\lambda_{\max}\{((\mathbf{I}-\mathbf{\Xi})(\mathbf{G}+\mathbf{I}+\mathbf{\Xi})^{-1}))^2\}\right).
\end{aligned}$$

Since matrix $\mathbf{G}$ is a nonnegative and positive definite matrix,

$$\lambda_{\max}\{((\mathbf{I}-\mathbf{\Xi})(\mathbf{G}+\mathbf{I}+\mathbf{\Xi})^{-1}))^2\} \leq \lambda_{\max}\{((\mathbf{I}-\mathbf{\Xi})(\mathbf{I}+\mathbf{\Xi})^{-1}))^2\}.$$

To understand the lower bound, we have to consider two cases based on how "big" the matrix $\mathbf{\Xi}$ is. Consider a function $f(x) = (1-x)^2/(1+x)^2$ with $x \in [\underline{x}, \bar{x}]$. When $x \geq 1$, $f(x)$ increases in x. Thus, $\max f(x) = f(\bar{x})$. Thus, when all eigenvalues of the matrix Ξ exceeds 1, or, equivalently, when $\lambda_{\max}\{\Xi\} \geq 1$,

$$\lambda_{\max}\{((\mathbf{I}-\Xi)(\mathbf{G}+\mathbf{I}+\Xi)^{-1}))^2\} \leq \left(\frac{\lambda_{\max}\{\Xi\}-1}{\lambda_{\max}\{\Xi\}+1}\right)^2. \tag{4.6}$$

Note that the matrix $\Xi = \mathbf{\Gamma_B^{-0.5} R_{off} \Gamma_B^{-0.5}}$ is a symmetric matrix and it is similar to a matrix $\mathbf{\Gamma_B^{-1} R_{off}}$, which shares the same set of eigenvalues. By Gergsgorin Disc Theorem, the maximum eigenvalue of a symmetric matrix could be bounded by its radius, i.e.,

$$\lambda_{\max}\{\Xi\} = \lambda_{\max}\{\mathbf{\Gamma_B^{-1} R_{off}}\} \leq \max_i \frac{\sum_{j\neq i} r_{ji}}{\beta_{ii}} = \max_i \rho_i = \bar{\rho}. \tag{4.7}$$

Substitute this to Equation (4.6), we have obtained the desired lower bound for the case when $\bar{\rho} \geq 1$: $\frac{W(\mathbf{q^N})}{W(\mathbf{q^*})} \geq \frac{1}{\kappa'}\left(1-(\frac{\rho-1}{\rho+1})^2\right)$.

Next, let's focus on the lower bound when $\bar{\rho} \leq 1$. In fact, $f(x) = (1-x)^2/(1+x)^2$ decreases in x when $x \leq 1$, which implies that $\max f(x) = f(\underline{x})$. However, it simply suggests that when all eigenvalues of matrix Ξ are smaller than 1, an upper bound on $\lambda_{\max}\{((\mathbf{I}-\Xi)(\mathbf{G}+\mathbf{I}+\Xi)^{-1}))^2\}$ is 1 (i.e., when ($\lambda_{\max}\{\Xi\} = 0$, i.e., no spillover), or equivalently, $W(\mathbf{q^N})/W(\mathbf{q^*}) \geq 0$, which is unfortunately, not a very useful bound. In fact, to prove that $W(\mathbf{q^N}) \geq (3/4\kappa')W(\mathbf{q^*})$, when $\bar{\rho} \leq 1$, we have to utilize a similar approach to prove the constant lower bound in Theorem 4.3.3 when costs are fully self-contained: When we expand the composite matrix, we will have 2 parts instead, one part is shown in Equation (4.5) and another part term in terms of $\mathbf{G}$ and Ξ. This is guaranteed to be nonnegative when $\bar{\rho} \geq 1$. □

The main difference between Theorem 4.3.3 and 4.3.4 is that the performance degradation in the presence of spillover could potentially be unbounded. The loss of efficiency depends on the maximum spillover cost-to-benefit ratio $\bar{\rho}$ and increases for all $\bar{\rho} > 1$. To interpret this result, note that $\bar{\rho} > 1$ is a sufficient condition which states that congestion occurs in the unregulated setting, i.e., $\mathbf{q^N} > \mathbf{q^*}$. It implies that when an additional user is enrolled, the increase in spillover cost outweighs the welfare gain, resulting in a net negative welfare change. Therefore, as $\bar{\rho} > 1$ increases, the gap between the unregulated setting and the social optimum widens.

When $\bar{\rho} \leq 1$, the output level in the unrelated setting is below the social optimal level. The welfare improvement brought by each additional user offsets the slipover cost, inducing a net positive welfare gain. The worst case happens when the output level is at its lowest compared to the optimal level, which happens with noncompeting service providers.

4.3.2 Upper bound on $W(\mathbf{q}^{\mathbf{N}})/W(\mathbf{q}^*)$

In this section, we focus on finding the best performance that can be achieved in an unregulated setting. In the rest of this section, we will use the competition index and maximum cost-to-benefit ratio evaluated in the social optimum, i.e., $\bar{\gamma} = \bar{\gamma}(\mathbf{q}^*)$, and $\bar{\rho} = \bar{\rho}(\mathbf{q}^*)$ and derive an upper bound on the efficiency of the unregulated oligopoly in terms of these quantities.

Lemma 4.3.5 *When cost* $\mathbf{l}(\mathbf{q}) = (l_1(\mathbf{q}), \cdots, l_n(\mathbf{q}))$ *is a convex function (componentwise),*

$$(\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^N} + \mathbf{R}^*)\mathbf{q}^{\mathbf{N}} \leq (\mathbf{B} + \mathbf{2R}^*)\mathbf{q}^*.$$

Proof. By Assumption 4.2.2, function $\mathbf{l}(\mathbf{q}) = (l_1(\mathbf{q}), \ldots, l_n(\mathbf{q}))$ is convex componentwise with $\mathbf{q}$.

$$\begin{aligned} \mathbf{l}(\mathbf{q}^{\mathbf{N}}) - \mathbf{l}(\mathbf{q}^*) &\geq \mathbf{R}^*(\mathbf{q}^{\mathbf{N}} - \mathbf{q}^*) \\ \Rightarrow \mathbf{l}(\mathbf{q}^{\mathbf{N}}) - \mathbf{R}^*\mathbf{q}^{\mathbf{N}} &\geq \mathbf{l}(\mathbf{q}^*) - \mathbf{R}^*\mathbf{q}^* \\ \Rightarrow -\mathbf{l}(\mathbf{q}^{\mathbf{N}}) + \mathbf{R}^*\mathbf{q}^{\mathbf{N}} &\leq -\mathbf{l}(\mathbf{q}^*) + \mathbf{R}^*\mathbf{q}^* \\ \Rightarrow \bar{\mathbf{p}} - \mathbf{l}(\mathbf{q}^{\mathbf{N}}) + \mathbf{R}^*\mathbf{q}^{\mathbf{N}} &\leq \bar{\mathbf{p}} - \mathbf{l}(\mathbf{q}^*) + \mathbf{R}^*\mathbf{q}^*. \end{aligned}$$

After substituting the optimality conditions from Equation (B.1) and (B.3), we obtain the desired result. □

To obtain the upper bound on $W(\mathbf{q}^{\mathbf{N}})$, we use convexity to express it in terms of $\mathbf{q}^*$ and $\mathbf{R}^*$ with the help of the Jacobian similarity factor.

Lemma 4.3.6 $\frac{W(\mathbf{q^N})}{W(\mathbf{q}^*)} \leq \kappa(1 - \lambda_{\min}\{((\mathbf{I} - \boldsymbol{\Xi})(\mathbf{G} + \mathbf{I} + \boldsymbol{\Xi})^{-1}))^2\})$*, where* $\kappa \geq 1$ *is the Jacobian similarity factor,* $\mathbf{G} = \boldsymbol{\Gamma}_\mathbf{B}^{-0.5}(\mathbf{B} + 2\boldsymbol{\Gamma}_\mathbf{R}^*)\boldsymbol{\Gamma}_\mathbf{B}^{-0.5}$ *and* $\boldsymbol{\Xi} = \boldsymbol{\Gamma}_\mathbf{B}^{-0.5}\mathbf{R}_{\mathbf{off}}^*\boldsymbol{\Gamma}_\mathbf{B}^{-0.5}$.

Proof of Lemma B.6.1. From Lemma 4.3.1, we obtain that

$$W(\mathbf{q}^N) = (\mathbf{q^N})^T(\frac{1}{2}\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{R}^\mathbf{N})\mathbf{q^N} = (\mathbf{q^N})^T\boldsymbol{\Psi}\boldsymbol{\Psi}^{-1}(\frac{1}{2}\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{R}^\mathbf{N})\boldsymbol{\Psi}^{-1}\boldsymbol{\Psi}\mathbf{q^N},$$

where $\boldsymbol{\Psi} = \mathbf{B} + \boldsymbol{\Gamma}_\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{R}^\mathbf{N} + \mathbf{R}^*$. Making use of Lemma 4.3.5 which shows that $\boldsymbol{\Psi}\mathbf{q^N} \leq (\mathbf{B} + 2\mathbf{R}^*)\mathbf{q}^*$, it follows that

$$\begin{aligned}
&W(\mathbf{q}^N) \\
\leq&(\mathbf{q}^*)^T(\mathbf{B} + 2\mathbf{R}^*)(\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{R}^\mathbf{N} + \mathbf{R^N})^{-1}(\frac{1}{2}\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{R}^\mathbf{N})(\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{R}^\mathbf{N} + \mathbf{R^N})^{-1} \\
&(\mathbf{B} + 2\mathbf{R}^*)\mathbf{q}^* \\
\leq&\kappa(\mathbf{q}^*)^T(\mathbf{B} + 2\mathbf{R}^*)(\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{R}^* + \mathbf{R}^*)^{-1}(\frac{1}{2}\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{R}^*)(\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{R}^* + \mathbf{R}^*)^{-1} \\
&(\mathbf{B} + 2\mathbf{R}^*)\mathbf{q}^*,
\end{aligned}$$

where we obtain the last inequality by using the Jacobian similarity property that we discussed for matrix $\mathbf{F}(\mathbf{q}) = \boldsymbol{\Gamma}_\mathbf{R}^\mathbf{N}$ and $\kappa \geq 1$. In particular, an upper bound on κ is given as follows, based on the definition of the maximum eigenvalue of a positive semidefinite matrix,

$$\begin{aligned}
\kappa \leq&\lambda_{\max}\{(\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{R}^\mathbf{N} + \mathbf{R^N})^{-2}(\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{R}^* + \mathbf{R}^*)^2(\frac{1}{2}\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{R}^\mathbf{N}) \\
&(\frac{1}{2}\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{R}^*)^{-1}\}. \qquad (4.8)
\end{aligned}$$

Combine this result with $W(\mathbf{q}^*)$ shown in Lemma 4.3.1, we obtain an upper bound

$$\frac{W(\mathbf{q^N})}{W(\mathbf{q}^*)} \leq \kappa\frac{(\mathbf{q}^*)^T(\mathbf{B} + 2\mathbf{R}^*)(\frac{1}{2}\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{R}^*)(\mathbf{B} + 2\mathbf{R}^*)\mathbf{q}^*}{(\mathbf{q}^*)^T(\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{R}^* + \mathbf{R}^*)(\frac{1}{2}\mathbf{B} + \mathbf{R}^*)(\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{B} + \boldsymbol{\Gamma}_\mathbf{R}^* + \mathbf{R}^*)\mathbf{q}^*}.$$

As all quantities are in the social optimum setting, we will skip the superscript on matrices. Denote $\mathbf{G} = \boldsymbol{\Gamma}_\mathbf{B}^{-0.5}(\mathbf{B} + 2\boldsymbol{\Gamma}_\mathbf{R})\boldsymbol{\Gamma}_\mathbf{B}^{-0.5}$ and $\boldsymbol{\Xi} = \boldsymbol{\Gamma}_\mathbf{B}^{-0.5}\mathbf{R}_{\mathbf{off}}\boldsymbol{\Gamma}_\mathbf{B}^{-0.5}$, the expression becomes

$$\frac{W(\mathbf{q}^{\mathbf{N}})}{W(\mathbf{q}^*)} \le \kappa \frac{(\mathbf{q}^*)^T \boldsymbol{\Gamma}_{\mathbf{B}}^{0.5}(\mathbf{G}+2\mathbf{I})\boldsymbol{\Gamma}_{\mathbf{B}}^{0.5}\mathbf{q}^*}{(\mathbf{q}^*)^T \boldsymbol{\Gamma}_{\mathbf{B}}^{0.5}(\mathbf{G}+\mathbf{I}+\Xi)(\mathbf{G}+2\Xi)^{-1}(\mathbf{G}+\mathbf{I}+\Xi)\boldsymbol{\Gamma}_{\mathbf{B}}^{0.5}\mathbf{q}^*}.$$

Using the Rayleigh-Ritz Theorem, the upper bound can be simplified as the follows. We skip some derivations as they follow exactly the same steps in the proof for Theorem 4.3.4.

$$\begin{aligned}\frac{W(\mathbf{q}^{\mathbf{N}})}{W(\mathbf{q}^*)} &\le \kappa \lambda_{\max}\{(\mathbf{G}+2\mathbf{I})(\mathbf{G}+\mathbf{I}+\Xi)^{-1}(\mathbf{G}+2\Xi)(\mathbf{G}+\mathbf{I}+\Xi)^{-1}\} \\ &\le \kappa(1-\lambda_{\min}\{((\mathbf{I}-\Xi)(\mathbf{G}+\mathbf{I}+\Xi)^{-1}))^2\}).\end{aligned}$$

□

Theorem 4.3.7 *When costs are fully self-contained, under Assumptions 4.2.1 to 4.2.4, the welfare in an unregulated oligopoly is bound above by*

$$\frac{W(\mathbf{q}^{\mathbf{N}})}{W(\mathbf{q}^*)} \le \kappa \left(1 - \frac{1}{(2+\bar{\gamma})^2}\right),$$

where κ is the Jacobian nonlinearity factor. The tight with symmetric service providers with affine costs (i.e., $\kappa = 1$).

Proof of Theorem 4.3.7. When there is no spillover, $\mathbf{R_{off}} = \mathbf{0}$ or $\Xi = \mathbf{0}$ and the upper bound in Lemma B.6.1 can be simplified to

$$\begin{aligned}\frac{W(\mathbf{q}^{\mathbf{N}})}{W(\mathbf{q}^*)} &\le \kappa \left(1 - \lambda_{\min}\{(\mathbf{G}+\mathbf{I})^{-2}\}\right) \\ &= \kappa \left(1 - \frac{1}{\lambda_{\max}(\mathbf{G}+\mathbf{I})^2}\right) \\ &\le \kappa \left(1 - \frac{1}{(1+\lambda_{\max}\{\mathbf{G}\})^2}\right), \qquad (4.9)\end{aligned}$$

where an upper bound on κ is given in Equation (B.6). Note $\mathbf{G} = \boldsymbol{\Gamma}_{\mathbf{B}}^{-0.5}(\mathbf{B}+2\boldsymbol{\Gamma}_{\mathbf{R}})\boldsymbol{\Gamma}_{\mathbf{B}}^{-0.5}$ is a symmetric, positive definite matrix. By the property of similar matrices, this matrix is similar to $\boldsymbol{\Gamma}_{\mathbf{B}}^{-1}(\mathbf{B}+2\boldsymbol{\Gamma}_{\mathbf{R}})$. Thus, they share the same set of eigenvalues. Using the Gershgorin Disc Theorem, we can upper bound the maximum eigenvalue of a matrix, i.e.,

$$\begin{aligned}\lambda_{\max}\{\mathbf{G}\} = \lambda_{\max}\{\mathbf{\Gamma_B}^{-1}(\mathbf{B}+2\mathbf{\Gamma_R})\} &\leq 1+\max_i \frac{\sum_{j\neq i}\beta_{ij}+2r_{ii}}{\beta_{ii}} \\ &\leq 1+\max_i \gamma_i \\ &= 1+\bar{\gamma}.\end{aligned}$$

Substituting these two inequalities into Equation (4.9), we obtain the the desired bound.

In order to show that the bound is tight, note that with affine cost, $\kappa = 1$. In addition, the inequalities in Lemma 4.3.5 and Lemma B.6.1 become equalities. Moreover, with symmetric service providers $\gamma_i = \bar{\gamma}$ for all i, one can show that the maximum eigenvalue of matrix $\mathbf{G}$ is exactly equal to $1+\bar{\gamma}$. □

Theorem 4.3.7 shows that the efficiency of the unregulated setting increases as the intensity of competition in the facility increases. The worst case happens when the service providers are independent with monopolistic power, i.e., $\bar{\gamma} = 0$. As the competition among the service providers intensifies, the efficiency gap between the unregulated setting and the social optimum diminishes. For example, with affine cost ($\kappa = 1$), when $\bar{\gamma} = 1$, the gap is reduced to 11.1%; as $\bar{\gamma}$ increases to 2, the gap becomes only 6.25%. One could argue that the unregulated setting is quite efficient when there is a fair amount of competition among the service providers.

Based on the analysis with only self-contained cost (Theorem 4.3.3 and 4.3.7), the results are somewhat comforting as the worst welfare loss is bounded and this number could be considerably smaller when there is competition among the service providers. An explanation is that when only self-contained cost is present, the only distortion in the unregulated setting is due to the oligopolistic power of the service providers, which leads to higher prices and lower output level as compared to the optimum. Competition reduces the market power of the service providers, thus, diminishing the distortion and closing the efficiency gap. Consider the extreme example of perfect competition, where $\bar{\gamma} \to \infty$, the total welfare obtained under the two settings converges, i.e., $W(\mathbf{q^N}) \to W(\mathbf{q}^*)$.

Theorem 4.3.8 *With the presence of spillover costs, under Assumptions 4.2.1 to 4.2.5, when the maximum cost-to-benefit ratio $\bar{\rho} \leq 1$, the welfare in the unregulated setting can be upper bounded by*

$$\frac{W(\mathbf{q^N})}{W(\mathbf{q}^*)} \leq \kappa \left(1 - \left(\frac{1-\bar{\rho}}{2+\bar{\gamma}+\bar{\rho}}\right)^2\right),$$

where κ is the Jacobian nonlinearity factor. The bound is tight with symmetric service providers with affine costs (i.e., $\kappa = 1$).

Remark When there is no spillover, i.e., $\bar{\rho} = 0$, Theorem 4.3.8 coincides with the upper bound in Theorem 4.3.7 with only self-contained cost.

Proof of Theorem 4.3.8. When there is spillover, an upper bound on the comparison of welfare achieved in the two settings are given by Lemma B.6.1, i.e.,

$$\frac{W(\mathbf{q^N})}{W(\mathbf{q}^*)} \leq \kappa(1 - \lambda_{\min}\{((\mathbf{I} - \Xi)(\mathbf{G} + \mathbf{I} + \Xi)^{-1}))^2\}).$$

Note matrix $\mathbf{G}$ is positive definite and the minimum eigenvalue of the composite matrix is bounded below by

$$\lambda_{\min}\{((\mathbf{I} - \Xi)(\mathbf{G} + \mathbf{I} + \Xi)^{-1}))^2 \geq \min_{\lambda\{\Xi\}} \left(\frac{\lambda\{\Xi\} - 1}{\lambda_{\max}\{\mathbf{G}\} + 1 + \lambda\{\Xi\}}\right)^2.$$

Consider a function $g(x, y) = ((x-1)(x+y+1))^2$ with $x \in [\underline{x}, \bar{x}]$, where $\underline{x} \geq 0$. If $\bar{x} \leq 1$, the function decreases in x. Thus, for a fixed y, the minimum of the function is achieved with $\bar{x}$. Therefore, we obtain the following:

When $\lambda_{\max}\{\Xi\} \leq 1$,

$$\lambda_{\min}\{((\mathbf{I} - \Xi)(\mathbf{G} + \mathbf{I} + \Xi)^{-1}))^2 \geq \left(\frac{\lambda_{\max}\{\Xi\} - 1}{\lambda_{\max}\{\mathbf{G}\} + 1 + \lambda_{\max}\{\Xi\}}\right)^2.$$

Since $\lambda_{\max}\{\Xi\} \leq \bar{\rho}$ and $\lambda_{\max}\{\mathbf{G}\} \leq 1 + \bar{\gamma}$, we can further lower bound the eigenvalue

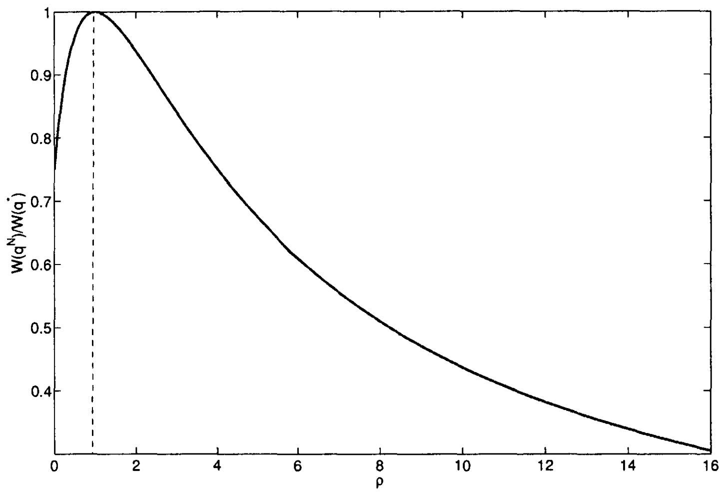

Figure 4-1: $W(\mathbf{q}^{\mathrm{N}})/W(\mathbf{q}^*)$ with respect to spllover cost-to-benefit ratio $\bar{\rho}$ for symmetric service providers with affine cost, where $\bar{\gamma}$ is set to zero (non-competing service providers).

of the composite matrix by

$$\lambda_{\min}\{((\mathbf{I}-\Xi)(\mathbf{G}+\mathbf{I}+\Xi)^{-1}))^2 \geq \left(\frac{\bar{\rho}-1}{\bar{\gamma}+2+\bar{\rho}}\right)^2.$$

To obtain the tightness result, note that with symmetric service providers, $\lambda_{\max}\{\Xi\} = \bar{\rho}$ and use the same argument in the proof for Theorem 4.3.7. □

Theorem 4.3.8 states that the efficiency of the unregulated setting increases with competition level $\bar{\gamma}$ and the maximum spillover cost-to-benefit ratio when $\bar{\rho} \leq 1$. Competition is beneficial to an unregulated setting as it reduces the distortion created by the oligopolistic power. When $\bar{\rho} \leq 1$, it implies that enrolling an additional user bring a net positive welfare to the society (additional consumer surplus is larger than the spillover cost). As a result, the efficiency of the unregulated setting also improves with this ratio.

Figure 4-1 shows the $W(\mathbf{q}^{\mathrm{N}})/W(\mathbf{q}^*)$ for symmetric noncompeting service providers with affine costs (i.e., $\kappa = 1$ and $\bar{\gamma} = 0$). The plot shows the effect of $\bar{\rho}$ on the efficiency of the unregulated setting. It starts at 75% because of noncompeting service providers. Then it first increases with $\bar{\rho}$ albeit at a decreasing rate and decreases as

$\bar{\rho}$ becomes large.

We would like to note that the upper bounds use $\bar{\rho}$ and $\bar{\gamma}$ evaluated at the optimal output level $\mathbf{q}^*$. The system welfare maximization problem is a strictly concave maximization problem, which is easier to solve compared to obtain a Nash equilibrium in the decentralized problem.

4.4 Simulation Experiments

So far, we have developed several bounds in terms of the efficiency analysis. We show that the bounds are tight for the special cases such as noncompeting or symmetric service providers and with affine costs. This section addresses a natural question, how "good" are our bounds? In this subsection, we will evaluate the performance of the bounds with nonlinear costs and asymmetric service providers via simulations.

4.4.1 Effect of nonlinearity and competition

In Theorem 4.3.3 and 4.3.7, we have shown that when costs are fully self-contained, the welfare in the unregulated setting is bounded between $\frac{3}{4} \leq \frac{W(\mathbf{q}^{\mathbf{N}})}{W(\mathbf{q}^*)} \leq \kappa(1 - \frac{1}{(1+\bar{\gamma})})^2$, where κ is the Jacobian nonlinearity factor with an upper bound as shown in Equation B.6 and $\bar{\gamma}$ is the competition index with the component on the self-contained cost evaluated at the $\mathbf{q}^*$.

To isolate the impact of nonlinearity on the cost function, we restrict to symmetric service providers with marginal utility function given by $p_i(q) = \bar{p}_i - \beta_{ii}q_i + \sum_{j \neq i} \alpha_{ij} q_j$ (and we will discuss the impact of asymmetry in the next subsection). Because of symmetric service providers, we will drop the subscript. For this experiment, the number of downstream retailers is fixed to 5 and $\beta = 5$. We consider a cost function of the following form, $l(q) = cq^k$, where $c > 0$ is the cost coefficient and k is the degree of this monomial. We run the experiment with the degree of the cost function ranging from 0 to 10 for both noncompeting ($\alpha = 0$) and competing service providers ($\alpha = 1$). The results are shown as the two curves labeled as "Exact" in Figure 4-2. It is clearly shown that the efficiency increases as service providers compete for

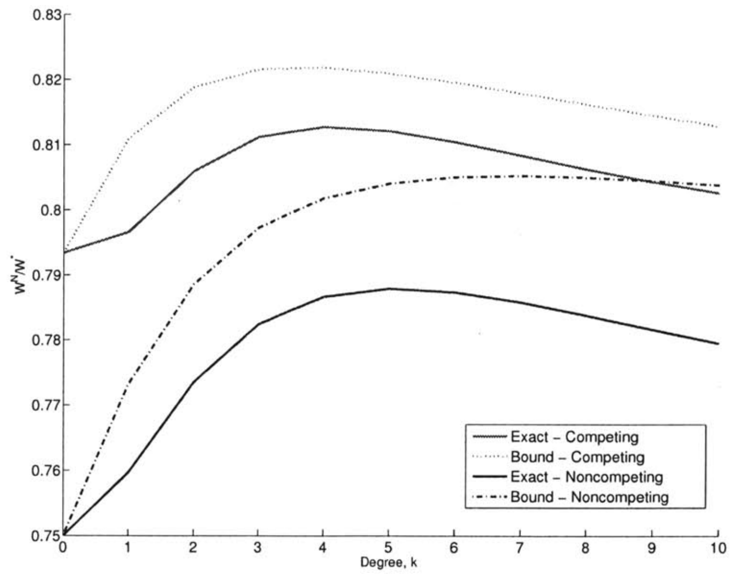

Figure 4-2: A simulation experiment to illustrate the impact of nonlinearity in the cost functions.

all degrees k. Moreover, Figure 4-2 also shows that for all $k \geq 0$, the efficiency of a decentralized system is strictly greater than 75%, which occurs with a constant marginal cost and no competition. The two dotted curves labeled as "Bound" in Figure 4-2 are computed by using the upper bound derived in Theorem 4.3.7. The differences between the exact values and the bounds are quite small. To compute the bound, we need to find out the maximum marginal self-contained cost evaluated in the system solution $\mathbf{q}^*$ which is easier to compute given the welfare maximization problem is standard concave maximization problem.

Figure 4-3 records the result of another experiment which evaluates the bound in Theorem 4.3.7 with competing service providers. We choose a cost function modeled after a standard M/M/1 queue where $l(q) = \frac{c}{\mu - q}$, and $\mu > q$. We use the same marginal utility function as described in the previous experiment with $\beta = 5$. Competition among service providers increases with the number of service providers, n. Meanwhile, for a fixed number n, competition also increases with α. Figure 4-3 depicts the actual efficiency ratio in the solid line and the upper bound in dotted line by adjusting the number of service providers from 1 to 20, and by varying $\alpha = 0.1, 1, 2$

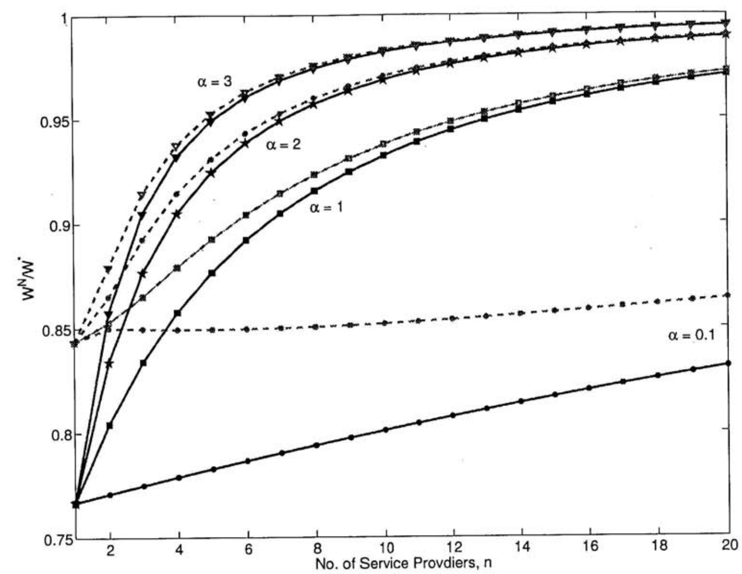

Figure 4-3: A simulation experiment to illustrate the impact of competition with self-contained cost $l(q) = \frac{c}{\mu - q}$.

and 3 respectively. This experiment shows that efficiency of an unregulated setting clearly improves with competition. Moreover, the upper bound becomes tighter as α and n increase.

4.4.2 Effect of asymmetry

In this experiment, we focus on a setting with linear cost but asymmetric service providers. Figure 4-4 reports the result of a simulation experiment with 500,000 instances. For each instance, a random number between 2 to 25 is first drawn to represent the size of the market. Next, various inputs (such as matrices $\mathbf{B}$ and $\mathbf{R}$) are generated to represent the *asymmetric* service providers (Jacobian $\mathbf{R}$ is independent of $\mathbf{q}$ since costs are affine). For this experiment, we restrict $\bar{\rho}$ (see Definition 4.2.8) to be equal or less than 1 so as to assess the quality of the upper bound we have developed in Theorem 4.3.8 (similar behavior is observed for the comparison with the upper bound when $\bar{\rho} > 1$). We first solve the problem in the unregulated and the social optimal settings respectively and compute the exact quantity. Next, we determine the corresponding upper bound in terms of $\bar{\gamma}$ and $\bar{\rho}$. We summarize the differences between the exact quantity and the lower bound in the histogram as shown in Figure 4-4. The experiment suggests that the upper bound provides a fairly accurate estimate of the exact quantity. For most of the 500,000 instances, the differences between the two quantities are within 0.05. Similar observations are also obtained for the other bounds developed in the previous two subsections.

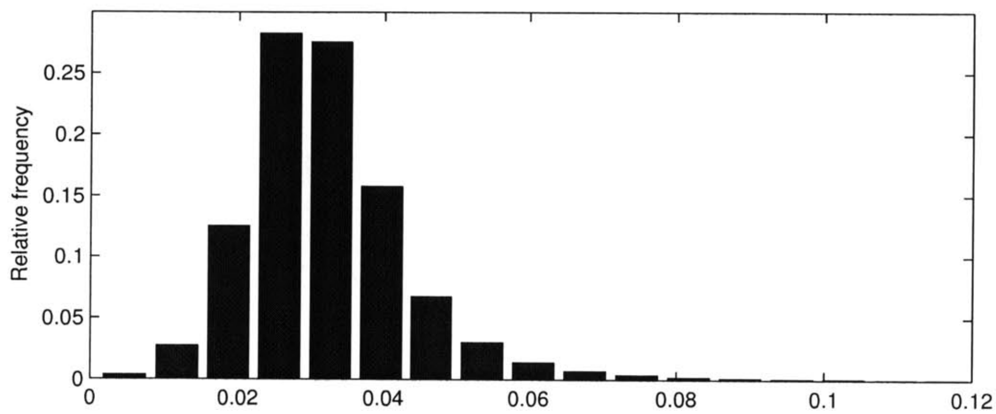

Figure 4-4: A simulation experiment with 500,000 samples to illustrate the strength of the upper bound in Theorem 4.3.8 where $\bar{\rho} \leq 1$. The x-axis represents the differences between the exact value of $W(\mathbf{q}^{\mathrm{N}})/W(\mathbf{q}^{*})$ and the upper bound grouped in bins. The y-axis represents the relative frequency of the instances within the bin size.

4.5 Impact of Mergers Between Service Providers

Recently, talks on several mega-mergers have created a stir in the media, e.g., AT&T with T-Mobile in the mobile marketplace, as well as United with Continental in the airline industry. The common argument against mergers claims that consumers could be at a disadvantage as the reduced competition leads to price increases. Views which support merges argue that cost synergy derived from a merger could lead to an improvement in the total social welfare. In this section, we will evaluate the impact of mergers on the total welfare in the absence of cost synergy. For this part of the analysis, an additional assumption of affine costs (i.e., $l_i(\mathbf{q}) = l_i(0) + \sum_j l_{ij}q_j$) is imposed so as to isolate the impact from mergers on the welfare. When service providers are also symmetric, then we will drop the subscripts, i.e., $l_i(0) = l(0)$ for all i, and $l_{ij} = l$ for all i and j,

Consider a pre-merger setting with $n > 2$ firms. Every firm competes against each other by offering a single service. The firm determines its equilibrium service level which maximizes the profit. We denote the two indices for all firms, namely, the competition index adjusted with internalized congestion effect and the external-cost-to-benefit ratio, as follow, $\gamma^{pre} = (\gamma_1^{pre}, \ldots, \gamma_n^{pre})$ and $\rho^{pre} = (\rho_1^{pre}, \ldots, \rho_n^{pre})$. Suppose m service providers decide to merge, where $n > m \geq 2$. Thus, in the post-merger setting, there is a total of $n - m + 1$ firms in the market, i.e., the merged firm with m types of services, together with $n - m$ other firms, each providing a single service. Without the loss of generality, let firms $\{1, \ldots, m\}$ be the merged firm. Denote the two indices in the post-merger setting as γ^{post} and ρ^{post} respectively. Note that the two indices in the post-merger setting are modified to capture the fact that the merged firm has to determine the service levels which maximize the total profit from all m types of services. The derivations can be found in the Appendix B.7.1. We will use W^{pre} and W^{post} to denote the total welfare achieved in the pre-merger and post-merger settings respectively.

Proposition 4.5.1 *With symmetric service providers who offer homogeneous services, in the pre-merger setting:* $\gamma_i^{pre} = (n - 1 + 2l)/\beta$ *and* $\rho_i^{pre} = (n - 1)l/\beta$ *for all* $i \in \{1, \ldots, n\}$. *In the post-merger setting:*

- *For the firms which are not involved in the merger, the competition index and the congestion cost-to-benefit ratio remain the same, i.e.,* $\gamma_j^{post} = \gamma_j^{pre}$, $\rho_j^{post} = \rho_j^{pre}$ *for all* $j \in \{m + 1, \ldots, n\}$;
- *For the merged firm, for all* $i \in \{1, \ldots, m\}$,
  - *when congestion effect is self-contained,* $\gamma_i^{post} = (n - m + 2l)/(\beta - 1 + m) < \gamma_i^{pre}$;
  - *for the merged firm, when congestion has the spilling effect,* $\rho_i^{post} = (n - m)l/(\beta - 1 + m) < \rho_i^{pre}$ *and* $\gamma_i^{post} = (n - m + 2ml)/(\beta - 1 + m)$.

For the firms which are not involved in the merger, since there are still $n-1$ homogenous services in the market, the relative price change with respect to quantity change stays the same, i.e., the competition index remains the same. The same argument also applies to the external-cost-to-benefit ratio for those firms. For the merged firm, with the self-contained congestion, the competition index of any of its services strictly decreases, because there are fewer "competitors" (numerator decreases), and his impact on prices increases (denominator increases). Similarly, consider its external-cost-to-benefit ratio, the uninternalized congestion is reduced from $(n-1)l$ to $(n-m)l$ as the merged firm internalizes more congestion. Moreover, being a larger firm, it also offers more marginal benefit to the society, i.e., it increases from β to $\beta-1+m$. Combining both effects, the external-cost-to-benefit ratio for the merged firm strictly decreases. However, with the spilling congestion effect, "competition" faced by the merged firm could potentially increase after it is adjusted by the self-contained congestion effect. As shown in Proposition 4.5.1, when the marginal congestion cost l is high enough (i.e., for $l > 1/2(1+n/(\beta-1))$, it is possible that $r_i^{post} > r_i^{pre}$, for $i \in \{1,\ldots,m\}$. After the merger, although the merged firm enjoys higher pricing power, at the same time, it also has to bear more congestion cost, which increases from $2l$ to $2ml$ due to its larger size. Thus, if we include the additional cost which the merged firm has to internalize, it is plausible that it could be at a disadvantage after the merger.

Proposition 4.5.2 *When costs are fully self-contained, when service providers are symmetric,* $W^{post} < W^{pre}$*, for all* $n > m \geq 2$.

The result coincides with the conventional view on mergers, that is, without considering cost synergies, mergers reduces competition in the market and it leads to leads to a decreases in the total social welfare. As shown in Theorems 4.3.3 and 4.3.7, the main distortion in the unregulated setting is the oligopolistic power which leads to "underproduction" compared to the socially optimal level. When congestion is self-contained, a merger strictly increases the power of the merged firm, resulted in more welfare loss in the unregulated setting.

Proposition 4.5.3 *With spilling congestion effect, when* $\rho^{pre} > 1$*, in the absence of* *cost synergies, mergers lead to an increase in the total welfare when service providers are symmetric. When $\rho^{pre} < 1$, depending on how much congestion cost the merged firm internalizes, the total welfare after the merge could either increase or decrease.*

The result suggests that without cost synergies, it is still possible for the society to benefit from mergers. It happens when the unregulated setting is the "bad" regime before the merger, where the external uninternalized congestion exceeds the external benefit of enrolling an additional user. We have shown in Proposition 4.5.1 that one of the effects of the merger is that the decrease in external-cost-to-ratios. Thus, the society benefits from the decrease in the external cost as a result of a merger. The impact on the total welfare when $\rho^{pre} < 1$ is less conclusive. From Theorem 4.3.8, when the competition index stays fixed, decreasing ρ^{post} leads to a welfare loss. However, if the merged firm internalizes a substantial amount of congestion effect such that the competition intensity actually increases after it is adjusted for the congestion cost, the society could still benefit from the merger.

The results seem to provide some (at least partial) explanations to the decisions made for the two merger proposals recently. On September 1st, 2011, the U.S. Justice Department filed a law suit in the federal court to block the merger between the country's second- and fourth-largest wireless carriers, AT&T and T-Mobile. In a statement released by the Justice Department, the combination would reduce wireless communication competition in the U.S., driving prices higher, making service worse and offering fewer products for U.S. consumers. Our result agrees with the decision as we have shown in the wireless market where congestion effect is self-contained, mergers between the carriers reduce the total welfare in the absence of cost synergy.

On the other hand, the merger between United Airlines and Continental has received its clearance from the Justice Department since August 2010. The department said the airlines would combine "largely complementary networks, which would result in overlap on a limited number of routes where United and Continental offer competing nonstop service." While our analysis does not consider the benefit stemmed from network effect of the merger, the results based on our model suggest that the merger could benefit passengers and airlines (the merged airline and all other airlines operating in the same airport) in the regions with congested airports.

4.6 Effectiveness, Attractiveness and an Alternative Implementation

In this section, we address some of the issues which traditionally have made the adoption of congestion pricing a challenging task in practice. We focus on the spilling congestion effect as we have shown in the earlier section that the potential welfare loss in the unregulated setting could be huge (an alternative interpretation is that the potential gains from implementing congestion pricing are high). As opposed to focusing on the aggregate impact such as the total social welfare, the main emphasis of this section is on the individual impact of congestion pricing on service providers and users respectively. We demonstrate that the surplus of the service providers and the users almost always decreases after implementing congestion pricing, unless the revenue collected from congestion pricing is utilized efficiently. Thus, it puts a lot of emphasis on how to use the revenue from congestion pricing. Moreover, in order to attract support for congestion pricing in reality, we attempt to design the scheme which appeals to the self-interest of individual participants.

4.6.1 Using the revenue from congestion pricing

The economic theory behind congestion pricing relies on using the revenue collected to improve the efficiency of the facility (recall the total welfare is the sum of consumer surplus, producer surplus and the revenue collected from congestion pricing). In the context of an airport, the revenue collected can be invested in infrastructure such as constructing more runways and/or technological improvement in air control which would increase the capacity of the airspace and reduce the impact due to bad weather. In reality, the facility manager might only intend to use part of the revenue for that purpose. It could also happen that the revenue collected is spent unwisely such that only part of the expected gain is realized. To model this, we introduce a parameter

$\phi \in [0, 1]$, called the revenue utilization rate, to denote the fraction of the revenue which ultimately benefits the society. The total welfare with congestion pricing would depend on ϕ, $W(\mathbf{q}, \phi) = CS(\mathbf{q}) + PS(\mathbf{q}) + \phi TR(\mathbf{q})$.

In the earlier section, we have quantified the welfare impact from congestion pricing by focusing on the ideal scenario with $\phi = 1$, that is, all the congestion pricing proceeds have been used to boost the social welfare. In the next two propositions, we take another look at this impact without being overly optimistic on the revenue utilization.

Proposition 4.6.1 *The optimal congestion pricing charged to the service provider, $t^*(\phi)$, increase in ϕ.*

The result spells out the relationship between the size of congestion pricing and the facility manager's ability to utilize the tax revenue to improve the social welfare. It is intuitive as the amount that a facility manager is allowed to charge should be directly related to his ability or commitment to utilize the revenue. Higher charge is only justified when the facility manager is capable of utilizing a significant amount of the revenue proceedings.

Proposition 4.6.2 *There exists a $\bar{\phi}$ such that $W^*(\phi) \geq W^N$, for all $\phi \geq \bar{\phi}$.*

The result imposes a criterion on a facility manager's ability in order to justify congestion pricing. While it is straightforward to see that the total social welfare increases with the revenue utilization rate ϕ, Proposition 4.6.2 suggests that the intervention from the manager could possibly lead to a welfare decrease when the utilization rate is is low. In particular, if at most $\bar{\phi}$ of the revenue is planned to benefit the society (airlines and/or passengers in the context of airports), congestion pricing should not be implemented. With symmetric service providers, the condition boils down to having $\bar{\phi} > (3 + r_{\max})/(2 + r_{\max} + \rho_{\max})$. It increases with the competition among the service providers and decreases when the external congestion cost outweighs the external benefit. $\bar{\phi}$ is below 1 only when $\rho_{\max} > 1$, implying a rather stringent requirement on the regulator's ability of utilizing the proceeds in order to justify imposing congestion pricing.

4.6.2 Individual impact of congestion pricing and welfare sharing

As most individuals are somewhat concerned with aggregate performance metrics (such as total welfare) and are more interested in their individual impact, we investigate the impact of congestion pricing on the individual service providers and the users respectively in this subsection. Let us denote the producer surplus and consumer surplus in the unregulated setting as PS^N and CS^N and the corresponding quantities and the revenue collected with congestion pricing as PS^*, CS^* and TR^* respectively.

Figure 4-5 shows producer surplus and consumer surplus with respect to the external-cost-to-benefit ratio (ρ) for the symmetric service providers for both the unregulated and regulated settings. When $\rho \leq 1$, $PS^* \geq PS^N$ and $CS^* \geq CS^N$. Note that when $\rho \leq 1$, the optimal congestion pricing is negative. With the subsidy from the facility manager, it increases the profit margins of the service provider who sees an increase in his profit. Meanwhile, lower prices also encourage more users from enrolling which leads to a higher consumer surplus. The reverse happens when $\rho > 1$ as PS^* and CS^* are strictly less than their counterparts in the unregulated setting as the facility manager imposes a positive charge on every service. Intuitively, PS^* decreases because the profit-maximizing service providers experience lower profit margin due to the positive access fee. A lower consumption level results in a lower CS^*. The classic definition of total social welfare in economics with taxation is defined as the sum of producer surplus, consumer surplus and the tax revenue ($W^* = PS^* + CS^* + TR^*$). Therefore, in the settings with a positive congestion pricing ($\rho > 1$), the welfare improvement is accrued as the "revenue" collected by the facility manager, at the expense of both the service providers and the users. It is not surprising that when all participants experience a surplus decrease, there is little desire for them to adopt the scheme.

Suppose the facility manager would like to share the revenue between the service providers and the users with the goal that every participants can do better than

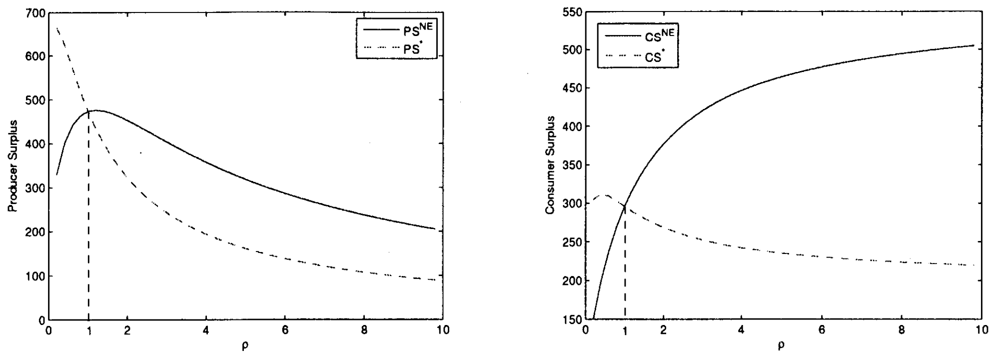

Figure 4-5: Producer surplus and consumer surplus with respect to the marginal spillover congestion cost-to-benefit ratio, ρ for the identical service providers, where $\beta = 10$, $l = 2$, $\bar{p} - \bar{l} = 10$.

in the unregulated setting. In particular, suppose α_i is the portion of the revenue TR^* that is given to service provider i and β_i portion of the revenue that is given to the users enrolling under service provider i, where $\sum_i(\alpha_i + \beta_i) = 1$. We denote the new producer surplus and consumer surplus after the welfare redistribution as $\overline{PS_i}$ and $\overline{CS_i}$ respectively, whereby $\overline{PS_i} = PS_i^* + \alpha_i TR^*$, and $\overline{CS_i} = CS_i^* + \beta_i TR^*$. While every participant would like to obtain a larger share of the pie, we define the following welfare-sharing rule such that everyone can benefit.

Proposition 4.6.3 *Suppose $W^* = \psi W^N$, where ψ denotes the total welfare increase after the implementation of congestion pricing. If the revenue TR^* is shared according to α_i and β_i, where*

$$\alpha_i = \frac{\psi PS_i^N - PS_i^*}{TR^*}, \quad \text{and } \beta_i = \frac{\psi CS_i^N - CS_i^*}{TR^*}$$

then $\overline{PS_i} \geq PS_i^N$ and $\overline{CS_i} \geq CS^N$ for all i.

The welfare-sharing rule ensures a *Pareto-efficient* outcome, i.e., $\sum_i(\overline{PS_i}+\overline{CS_i}) = W^*$. Moreover, it also guarantees *collective individual rationality* as both producer surplus and consumer surplus experience an improvement compared to their counterparts in the unregulated setting. We would like to point out the surplus obtained under the welfare-sharing rule, i.e., $\overline{PS_i} = \frac{PS_i^N}{W^N}W^*$ and $\overline{CS_i} = \frac{CS_i^N}{W^N}W^*$ is equivalent to using the proportional value as an allocation rule in a *n-person bargaining game* with their outside option as their gain in the unregulated setting. The welfare-sharing rule also captures a notion of "fairness": It scales each entity's surplus in the unregulated setting by a same factor, ψ, which represents how much benefit the society as a whole reaps from congestion pricing. It then compares that quantity (ψPS_i^N or ψCS_i^N) with every entity's corresponding surplus after congestion pricing and compensates each participant accordingly.

In the context of airport congestion pricing, FAA has to compensate airlines and traveling passengers by giving back the revenue collected. To compensate airlines, FAA could give discounts on existing charges such as counter and baggage claim rental fees. To redistribute the revenue to passengers, one remedy is to reduce airport tax or passenger facility fees. Passenger facility charges (PFC) are special taxes which airports collect from individual airline passengers. A PFC is imposed each time a passenger passes through an airport eligible to collect a PFC (only a small number of airports in the United States do not collect a PFC). The PFC is actually added to the price of an airline ticket, and airlines are forced to collect the tax for airports at the time the passenger pays for their ticket. Nonetheless, we recognize the difficulties associated with implementing the proposed scheme due to the administrative burden on both FAA and airlines.

In the following proposition, we explore another implementation of congestion pricing which guarantees that each participant experience a welfare improvement without going through the hassle of collecting congestion pricing and later redistributing the revenue.

Proposition 4.6.4 *There exists a new marginal utility function $\tilde{u}_i(\mathbf{q}) = \theta_i - \sum_{j=1}^{n} \beta_{ij} q_j$, such that producer surplus and consumer surplus earned in the unregulated setting are the same as the target level, $\overline{PS_i}$ and $\overline{CS_i}$ respectively.*

Under this new utility function, without any intervention from the facility manager, the service providers and the users could achieve the target surplus level in an unregulated setting, and all participating entities experience a surplus increase. Compare this new demand function with the initial model (where $u_i(\mathbf{q}) = \bar{p}_i - \sum_{j=1}^{n} \beta_{ij} q_j$), the only difference is that the maximum willingness of a representative user to pay for i's service is changed from $\bar{p}_i$ to θ_i. It suggests an "easy fix" as an alternative implementation of congestion pricing and welfare redistribution: For every service, the facility manager mandates a surcharge $(\theta_i - \bar{p}_i)$ which alters its base price. Once the surcharge is imposed, the service providers determine their respective profit-maximizing service level and get to keep all the revenue earned. Because of the new higher base price, some users will choose to leave the facility and this relieves the congestion problem. For the service providers, despite serving fewer users, they benefit from a higher profit margin per user (due to the surcharge) and reap higher profit. Meanwhile, the remaining users derive higher utility from using the service, therefore, they also experience higher surplus. More importantly, this implementation bypasses the difficulties of actual implementation of congestion pricing and the subsequent welfare-redistribution.

4.7 Conclusions

In this work, we have considered a setting where several services providers compete for users who are sensitive to both prices and congestion by providing multiple differentiated services. We have provided tight parametric bounds on the welfare loss of the unregulated setting. We have shown that with self-contained costs, the unregulated setting could be quite efficient and the maximum welfare loss is capped at 25%, even with highly nonlinear convex costs. With spillover cost, the efficiency of the unregulated setting highly depends on the relative magnitude of the marginal external cost and the marginal benefit associated with enrolling an additional user. We have also shown that the congestion pricing charged to service providers with market power is lower than that of the traditional "road toll". Contrary to conventional wisdom that mergers reduce social welfare, we show that in a setting with severe spilling congestion effect, mergers could lead to a social welfare improvement, even in the absence of cost synergy. Lastly, based on a welfare-sharing scheme, we proposed an alternative implementation of congestion pricing such that both the users and service providers can enjoy higher surplus.

Chapter 5

Price of Anarchy in Joint Ventures

5.1 Introduction

A proliferation of joint ventures has been witnessed across the globe in the recent years. A joint venture takes place when two or more business partners pool their resources and expertise to achieve a particular goal for a contractual period of time. Joint ventures stand in the middle ground between non-cooperative competition and merging. They provide companies with the opportunities to gain new capacity and expertise, enter related businesses or new geographic markets, gain new technological knowledge access to greater resources, and share risks with other venture partners.

In this work, we consider a setting where multiple entities take part in a joint venture and each of them contributes one type of resources. We distinguish two types of resource pooling in joint ventures, depending on whether the resources are *heterogeneous* or *homogeneous*. When resources are heterogeneous, they are not fully substitutable. Thus, the effective capacity of a joint venture is limited to the the minimum level of an individual contribution. In other words, the lowest contribution by one partner becomes the bottleneck in planning the capacity for the joint venture. This is in contrast with homogeneous resource pooling, where the resources are perfectly substitutable and the overall capacity of a joint venture is determined by aggregating all individual contributions.

One example that demonstrates the success of a joint venture with heterogeneous resource sharing is Massachusetts Eye and Ear Infirmary (MEEI), a hospital specialized in providing patient care for disorders of the eye, ear, nose, throat, head and neck in downtown Boston. With the vast majority of its services is outpatient in nature, MEEI experiences lower profit margins than a regular hospital and has been pressured to increase its patient volume so as to strengthen its financial status. Since 2005, MEEI has established five satellite clinics through joint ventures by collaborating with community hospitals in the suburbs. A typical agreement specifies that MEEI provides expertise (physicians and nurses) and its brand name[1] while the community hospital is responsible for providing facility and other necessary hardware. The two types of resources, i.e., expertise and facility, are not interchangeable. The maximum number of services that can be supported in such a satellite clinics is limited by MEEI's input as well as the space constraints such as the number of operating rooms available in the new location.

In 2003, US-based car rental firm Avis has set up a joint venture in Shanghai, China. The new company named Anji Car Rental and Leasing, 50-50 owned by Avis Europe and Shanghai Automotive Industry Sales Corporation, takes over the existing fleet of 1,000 vehicles from Shanghai Anji Car Rental and operates it under the Avis brand name. The venture expects to establish more than 70 outlets nationwide. This is a typical joint venture with homogeneous resource sharing, where the capacity in the new company is supported by aggregating the number of vehicles from the two companies.

Besides the healthcare industry and car-rental industry, another sector which has been a flurry to establish joint ventures is the airline industry. An airline alliance is an agreement between multiple independent partners to collaborate in various activities to streamline costs while expanding global reach and market penetration. The presence of alliances in the airline industry has followed an increasing trend since the first large airline alliance was formed in 1989 between Northwest and KLM. By March 2009, the three major alliances (Star, Sky Team and Oneworld) combined flew around 73% of all passengers worldwide (Hu et al. 2012). On the cost side, there are

[1] The satellite clinic located within the community hospital is labeled as a MEEI branch.

strong incentives for airlines to operate large networks as the evidences on economies of scale have been well documented (Caves et al. 1984, Brueckner and Spiller 1994, Keeler and Formby 1994, etc). On the revenue side, one of the fundamental attractions of an airline alliance is the ability to offer codeshare fights. Code sharing is an agreement between two carriers whereby one carrier allows a different carrier to market and sell seats on some of its flights. Based on empirical evidences, Brueckner (2003) conclude that codesharing among Star Alliance partners yielded an annual benefit of around \$20 million. Morever, the information comes with codesharing can be tremendously beneficial. Jain (2010) show that sharing information on bid prices yields higher revenues of the order of \$100 million for every big partnering carrier in the alliance.

5.1.1 Results and Contributions

In this work, we study both types of joint venture models and address some issues pertinent to the success of joint ventures. When several companies agree to a partnership, disparate interests often exist as each participant is more concerned with his or her own gain. Given the misalignment in incentives and uncertainties in demand, we are interested in measuring the performance of a joint venture by quantifying the difference in the investment level and the total profit attained with respect to a system optimal outcome.

We have shown that in a joint venture with heterogeneous resource pooling, despite the existence of asymmetric parties, the equilibrium induces an equal capacity contribution from every partner. Although multiple Nash equilibria could exist, we show that there always exists a unique strong Nash equilibrium. In addition, we show that there exists a fair and efficient way to share revenue such that an optimal investment decision could be reached in a Nash equilibrium. Next, we also consider Nash Bargaining model which is a natural framework to define and design fair assignment of the capacity investment levels between multiple players. We show that the same revenue sharing scheme could also be used in a Nash Bargaining model to induce the optimal decision. This optimal revenue sharing scheme indicates that the awarded each party receives must be proportional to his marginal cost.

For joint ventures with homogeneous resource pooling, we first prove some structural properties on the effective capacity under any demand distribution with convex costs. The analysis is challenging as the investment of each player could only be determined by solving a system of implicit equations. We show that joint venture always *underinvests* as the effective capacity is always lower than that of a coordinated setting.

We then focus on quadratic-linear cost functions and show that, through an intercept-argument, the effective capacity in a joint venture with respect to any revenue sharing ratio is *at least* $1/n$ of the optimal level, where n is the number of participants. Moreover, the ratio between the capacity level could be upper bounded in terms of the cost asymmetry between the two players and the revenue sharing ratio. While we show that there does not exist a fixed marginal revenue sharing contract which can coordinate the players, we propose an interval for the revenue sharing ratio which induces an outcome that is guaranteed to achieve *at least 50%* the optimal profit for a 2-player model. This interval depends on the cost asymmetry between the two players and the demand concentration.

Lastly, we consider general convex cost in the homogeneous resource pooling model with an arbitrary number of asymmetric players. We show that a lower bound to the efficiency of the original setting with the nonlinear convex costs is that of a modified setting with linear costs, where the coefficients are equal to the marginal cost of each player evaluated in the Nash equilibrium of the original problem. As a result, we show that the comparative analysis on profit can be reduced to analyze the joint investment level made in the Nash and the system in the setting with the linear costs.

The rest of the chapter is organized as follows. We begin with a review on related literature in Section 5.1.2. Section 5.2 describes the two models and assumptions. We analyze and present the main results on capacity sharing and substitution model in Section 5.3 and Section 5.4 respectively.

5.1.2 Related Literature

This paper studies strategic capacity management under uncertainty. In the operations management literature, there is a vast body of work using the classic newsvendor model or some variations to capture uncertainties. Federgruen and Zipkin (1986) is the classic reference for capacitated inventory management. Papers including Kapuscinski and Tayur (1998), Angelus and Porteus (2002), Bradley and Glynn (2002), Van Mieghem and Rudi (2002) consider capacity investment decisions in capacitated Newsvendor networks. Van Mieghem and Dada (1999) take a different approach at capacity management and address how the relative timing of the decisions on capacity, inventory, and price impact the sensitivity and profitability. We refer readers to Van Mieghem (2003) for an excellent survey paper on the recent development on capacity management. In this work, the capacity of a joint venture depends on the contribution of multiple participants. Depending on the nature of the resources, the effective capacity can be the minimum or the sum of individual contributions.

In many settings, capacity-investment decisions are the results after interacting with other economic agents. Thus, it seems natural for capacity investment models to incorporate the strategic behavior of self-interested agents. Cachon and Lariviere (1999) consider the manufacturer's capacity investment and allocation decisions to several downstream retailers that have private information. Caldentey and Wein (2003) present contracts that are linear in backorder, inventory, and capacity levels to coordinate a manufacturer and retailer production-inventory system, including the capacity decision. Examples on single-resource, multiple-agent also include Carr and Lovejoy (2000), Porteus and Whang (1991), Kouvelis and Lariviere (2000), etc. Bassok et al. (1999) and Netessine and Rudi (2003) explore the impact of substitution in an inventory context, and its effects are likely to be similar in capacity problems.

In this work, the strategic behavior of participants involving in a joint ventures is captured in a noncooperative game, as each entity determines his level of contribution with the goal to maximize his profit. While the revenue each party receives depends on the effective capacity of the joint venture, the incentive of each entity might not be correctly aligned to one which maximizes the collective return. We consider a fixed rate revenue sharing contract described in Cachon and Lariviere (2005) to split revenue among the participants. To capture the high capital investment incurred in joint ventures in the healthcare industry, we consider general convex cost function so as to capture the diminishing returns, in contrast to linear cost function which is common in the operations management literature (e.g., Bernstein and Federgruen 2007, Cachon 2003, Corbett et al. 2005, Martinez-de Alberniz and Simchi-Levi 2009). In this setting, we show that an "optimal" coordinating contract which enables the parties with self-interests to behave as a coordinated entity does not necessarily exist with homogeneous resources. We then propose a range for fixed revenue sharing ratio which induces reasonably good outcomes.

Standing in the middle ground between non-cooperative competition and merging, one of the most fundamental building blocks of joint ventures is negotiation. Empirical studies suggest that "the power of a joint venture is only as strong as the negotiation behind it" (Y. and O. 2002, Lin and Germain 1998). The topic on negotiation has gained a lot of attraction in the economics literature since Nash (1950) (e.g., Myerson 1979, Binmore et al. 1986, Rubinstein 1982, etc). In the fast few years, more results on negotiation have become known in the field of operations management (see for example, Reyniers and Tapiero 1995, Miller 1992, Chod and Rudi 2006, etc). Nagarajan and Sosic (2008) present an excellent survey paper on cooperative game theory in the field of supply chain management. In this work, utilizing the bargaining model, we propose a revenue sharing scheme which induces an outcome which is coincides with the system optimum.

Lastly, our work which measures the performance of an unregulated setting with respect to a centralized system is related to a stream of literature on *price of anarchy*, popularized by Koutsoupias and Papadimitriou (1999). It compares the performance of the worst-case Nash equilibrium with respect to the centralized system. The concept has been used in transportation networks (Roughgarden and Tardos 2002, Correa et al. 2004, 2007, Roughgarden 2005), network pricing (Acemoglu and Ozdaglar 2007, Weintraub et al. 2010), oligopolistic pricing games in a single tier (Farahat and Perakis 2010a,b), and supply chain games with exogenous pricing (Perakis and Roels 2007, Martinez-de Alberniz and Simchi-Levi 2009, Martinez-de Alberniz and Roels 2010).

5.2 Model Formulation

In this section, we first present the model for a joint venture with n players as an uncoordinated game. As a benchmark, we also present the model in the system setting, i.e., n entities were merged and coordinated as a single entity with the goal to maximize the total return.

5.2.1 Joint-venture: an uncoordinated game

Consider a joint venture with n profit-maximizing players with asymmetric cost functions. The joint venture generates a joint revenue $R(p, \mathbf{K})$ where p is the fixed price and $\mathbf{K} = (K_1, \ldots, K_n)$ captures the resources contributed by each player. A revenue-sharing contract dictates that player i receives revenue $\beta_i R(p, \mathbf{K})$. Let $f_i(K_i)$ be the convex cost associated with investing K_i resources by player i. Based on a pre-negotiated revenue-sharing ratio $\beta = (\beta_1, \ldots, \beta_n)$, player i tries to maximize her profit $\pi_i(\beta) \triangleq \beta_i R(p, \mathbf{K}) - f_i(K_i)$ by choosing her own investment level K_i, which leads to a Nash equilibrium (NE).

5.2.2 Merger: the system optimum

Consider the centralized system in which n players are merged and coordinated as a single player. The merger generates the highest possible profit $\pi_T^* \triangleq R(p, \mathbf{K}) - \sum_{i=1}^n f_i(K_i)$ by collectively choosing the resource investment $\mathbf{K}$. This yields the first-best or system optimal solution.

5.2.3 Resource-sharing models

We consider two types of resource-sharing models depending on the nature of the resources pooled from different players. The nature of the resources determines the effective capacity in a joint venture, which in turn affects the revenue function $R(p, \mathbf{K})$. We formally define them as follows:

Definition 5.2.1 ***Heterogeneous resource-sharing.*** *The aggregate revenue generated by the joint venture is given by* $R(p, \mathbf{K}) = p\mathbb{E}(\min(D, \min_i(K_i)))$.

The type of resource provided by each player is heterogeneous and not fully substitutable. A service can only be performed with a complete portfolio of resource types. The effective capacity supported by the joint venture is therefore limited to the minimum capacity level invested by the players.

Definition 5.2.2 ***Homogeneous resource-sharing.*** *The aggregate revenue generated by the joint venture is given by* $R(p, \mathbf{K}) = p\mathbb{E}(\min(D, \sum_{i=1}^{n}(K_i)))$.

The type of resource provided by each player is homogeneous to each other and hence fully substitutable. A service can be performed by using the resource contributed by any (possibly single) player. The effective capacity supported by the joint venture is therefore the sum of capacity level invested by each player.

In the next two sections, we will study both types of resource-sharing models and present the differences in the capacity investment and the total profit generated in a joint venture to those in a system optimum.

5.3 Heterogeneous Resource-sharing Models

With heterogeneous resources, the effective capacity is limited by the minimum capacity invested among all players, which becomes the *bottleneck capacity.* Consider the merger setting, the central planner tries to maximize the aggregate revenue by collectively choosing the capacity investment $\mathbf{K}$, i.e.,

$$\pi_T^* \triangleq \max_{K,K_i} p\mathbb{E}[\min(K,D)] - \sum_{i=1}^{n} f_i(K_i), \text{ s.t. } K \leq K_i,\ i = 1,\ldots,n. \tag{5.1}$$

Let K^* and $K_1^*,\ldots,K_n^*$ be the system optimal solution.

Lemma 5.3.1 *At system optimality, the capacity invested by each player is the same, i.e., $K^* = K_i^*$ for all $i = 1,\ldots,n$, where K^* solves $\mathbb{P}(D \leq K^*) = 1 - \sum_{i=1}^{n} f_i'(K^*)/p$.*

Proof of Lemma 5.3.1. Without loss of generality, if there exists a pair of players i and j such that $K_i^* < K_j^*$, we can decrease the capacity invested by player j from K_j^* to K_i^*. By doing so, the profit increases by reducing the cost while maintaining the same revenue. Hence, we reach a contradiction. At system optimality, $K^* = K_i^*$ for all $i = 1,\ldots,n$, and (5.1) reduces to a single variable optimization in which K^* can be obtained by the first-order condition. □

In the system optimum, each individual capacity investment K_i^* must be reduced to the bottleneck capacity K^* when resource-sharing is heterogeneous, since any further investment beyond the bottleneck capacity only increases the total cost and decreases the total profit.

In a joint-venture with a pre-negotiated revenue-sharing contract β, player i tries to maximize her profit by choosing her profit-maximizing capacity investment level K_i based on other players' strategies K_{-i}, which leads to a Nash equilibrium, i.e.,

$$\pi_i^N(\beta) \triangleq \max_{K,K_i;K_{-i}} \beta_i p\mathbb{E}[\min(K,D)] - f_i(K_i), \text{ s.t. } K \leq K_j,\ j = 1,\ldots,n,$$

Now, let K^N and $K_1^N,\ldots,K_n^N$ be the Nash equilibrium solutions.

Lemma 5.3.2 *In joint-ventures, any $K^N(\beta) = K_1^N(\beta) = \ldots = K_n^N(\beta) \leq \min_{1\leq k\leq n}(A_k)$ are Nash Equilibria, where A_k solves*

$$\mathbb{P}(D \leq A_k) = 1 - \frac{f_k'(A_k)}{\beta_k p}.$$

In particular, $K^{SN}(\beta) = K_1^{SN}(\beta) = \ldots = K_n^{SN}(\beta) = \min_{1\leq k\leq n}(A_k)$ *is a unique Strong Nash equilibrium.*

Proof of Lemma 5.3.2. Without loss of generality, if there exists a pair of players i and j such that $K_i^N(\beta) < K_j^N(\beta)$, player j can decrease its capacity investment from $K_j^N(\beta)$ to $K_i^N(\beta)$ lowering her cost and improving her profit. Thus, at Nash equilibrium, all players must have the same capacity investment level, i.e., $K^N(\beta) = K_i^N(\beta)$ for all $i = 1, \ldots, n$.

Now assume that $\min_{1\leq k\leq n}(A_k) = A_m$. Now if $A_m < K^N(\beta) = K_m^N(\beta)$, player m always has incentives to unilaterally lower her investment level to A_m since A_m is her profit-maximizer. This forces all players to invest at A_m. Any capacity investment level $\tilde{A}_m$ such that $0 \leq \tilde{A}_m \leq A_m$ is also a Nash equilibrium since no player has incentives to unilaterally deviate from $\tilde{A}_m$. In particular, $K^{SN}(\beta) = K_1^{SN}(\beta) = \ldots = K_n^{SN}(\beta) = A_m$ is a unique *Strong Nash equilibrium* in which no coalition, taking the actions of its complements as given, can cooperatively deviate in a way that benefits all of its members. □

Lemma 5.3.2 indicates that the capacity invested by each player must be the same in a joint venture. Since the revenue received by player i depends solely on the bottleneck capacity $K^N(\beta)$ when resource-sharing is heterogeneous, any further investment beyond the bottleneck capacity only increases her cost and decreases her profit. Lemma 5.3.2 also implies that A_k is the profit-maximizing capacity for player k. Since the resource-sharing is heterogeneous, the player m with the lowest profit-maximizing capacity (i.e., $A_m = \min_{1\leq k\leq n}(A_k)$) can unilaterally choose to invest at her profit maximizing capacity, forcing all other players to invest at the same capacity level. Note that any capacity investment level no greater than A_m is a Nash equilibrium whereas any capacity investment level above A_m is not. As a result, it is easy to see that with the existence of multiple Nash equilibria, it is possible for a joint venture to achieve an arbitrarily bad outcome compared to the system optimum.

So far, we have modeled the decision making process in a joint venture as a Nash Equilibrium. Next, we will propose an alternative model where the players participate a Nash bargaining game to determine their respective investment decisions for a given revenue sharing ratio β.

Nash Bargaining Solution (NBS). The Nash Bargaining Solution (see Appendix C) is a natural framework that allows us to define and design fair assignment of the capacity investment levels between n players, which can derive desirable properties such as Pareto efficiency and proportional fairness. Based on a particular revenue sharing contract β, n players choose their capacity investment levels according to a Nash Bargaining game, i.e.,

$$\max_{K,K_i} \prod_{i=1}^{n} \pi_i^B(\beta), \text{ s.t. } K \le K_j, \; j = 1, \dots, n,$$

which is equivalent to solving

$$\max_{K,K_i} \log \sum_{i=1}^{n} \pi_i^B(\beta), \text{ s.t. } K \le K_j, \; j = 1, \dots, n. \tag{5.2}$$

Let K^B and $K_1^B, \dots, K_n^B$ be the Nash Bargaining Solution from solving (5.2).

Theorem 5.3.3 *There exists a unique revenue sharing contract,*

$$\beta_i^* = \frac{f_i'(K^*)}{\sum_{j=1}^{n} f_j'(K^*)}, \qquad i = 1, \dots, n,$$

such that the Nash Bargaining Solution, the unique Strong Nash equilibrium, and the system optimal solution coincide, i.e., $K^B(\beta^*) = K^{SN}(\beta^*) = K^*$.

Proof of Theorem 5.3.3. Observe that (5.2) is equivalent to a single variable optimization,

$$\max_{K} \log \sum_{i=1}^{N} \left(\beta_i p \mathbb{E}[\min(K, D)] - f_i(K)\right). \tag{5.3}$$

The first-order condition gives us

$$\sum_{i=1}^{N} \frac{\beta_i p \mathbb{P}(D \ge K^B) - f_i'(K^B)}{(\beta_i p \mathbb{E}[\min(K^B, D)] - f_i(K^B))} = 0. \tag{5.4}$$

By Lemma 5.3.2, at Nash Equilibrium, $K^N \leq \min_{1\leq k\leq n}(A_k)$, where A_k solves

$$\mathbb{P}(D \geq A_k) = \frac{f_k'(A_k)}{\beta_k p}.$$

This implies that

$$\beta_i p \mathbb{P}(D \geq K^N) - f_i'(K^N) \geq 0, \text{ for all } i = 1, \ldots, n. \tag{5.5}$$

Suppose that there exists a solution γ to both the Nash Bargaining game and the Nash equilibrium, i.e., $\gamma = K^N(\beta) = K^B(\beta)$. Then γ must satisfy (5.4) and (5.5) simultaneously, implying that

$$\beta_i p \mathbb{P}(D \geq \gamma) - f_i'(\gamma) = 0 \text{ for all } i = 1, \ldots, n. \tag{5.6}$$

If such γ exists, $\gamma = K^{SN}(\beta)$, i.e. γ is the unique Strong Nash equilibrium since $\gamma = A_1 = \ldots = A_n = \min_{1\leq k\leq n}(A_k)$ by (5.6).

Now summing (5.6) over all players and $\sum_{i=1}^n \beta_i = 1$, we have

$$p\mathbb{P}(D \geq \gamma) - \sum_{i=1}^n f_i'(\gamma) = 0. \tag{5.7}$$

By (5.6) and (5.7), we know that β must be of the following form,

$$\beta_i = \frac{f_i'(\gamma)}{\sum_{j=1}^n f_j'(\gamma)}, \qquad i = 1, \ldots, n,$$

Moreover, note that by Lemma 5.3.1, (5.7) implies that $\gamma = K^*$. Since K^* is the unique system optimal solution, there exists a unique revenue sharing contract

$$\beta_i^* = \frac{f_i'(K^*)}{\sum_{j=1}^n f_j'(K^*)}, \qquad i = 1, \ldots, n,$$

such that $\gamma = K^* = K^{SN}(\beta^*) = K^B(\beta^*)$. $\square$

Theorem 5.3.3 shows that when resources are heterogeneous, there is a way to rely on the revenue sharing contract to eliminate the incentive misalignment among the players and induce the system optimal outcome. In addition, the way to do so is the same when the players' behavior is predicted by a Nash equilibrium as well as the Nash bargaining solution.

In addition, besides inducing the efficient decision, the optimal revenue sharing contract in Theorem 5.3.3 also embodies the notion of *proportional fairness.* For an investment level K^*, player i bears a marginal cost $f_i'(K^*)$ and the aggregate marginal cost is given by summing up the marginal cost of every player participating in the joint venture, $\sum_j f_j'(K^*)$. Theorem 5.3.3 specifics that the marginal revenue ratio which player i is entitled to receive (β_i) should be equal to the proportion of his marginal cost to the aggregate marginal cost ($f_i'(K^*)/\sum_j f_j'(K^*)$). In simple words, "fairness" in this context suggests that every participant in a joint venture should be awarded "proportionally" to the risk (cost) she has to undertake.

5.3.1 Numerical Examples

We conduct numerical studies to compare our approach with the existing approach adpoted by some joint-ventures (such as MEEI). In the existing model, joint-ventures set their capacity investment level according to the long-run average demand, i.e. $K^{EX} = \mathbb{E}[D]$. In addition, they split the revenue based on how much each party invests in *total* capacity. More specifically, they set the revenue sharing parameter to be

$$\beta_i^{EX} = \frac{f_i(K^{EX})}{\sum_{j=1}^n f_j(K^{EX})}, \qquad i = 1, \ldots, n.$$

We consider a 2-player game with unit service price $p = 1200$. Assume that the demand follows a normal distribution, and the cost functions to be quadratic, i.e. $f_i(K_i) = a_i K_1^2/2 + b_i K_i + c_i$ for $i = 1, 2$. Without loss of generality, we let $a_1 = 1$, $a_2 = 0.5$, $b_1 = b_2 = 100$ and $c_1 = c_2 = 0$. Table 5.1 shows the simulation results.

The simulation results show that our approach outperforms the existing approach

Demand	Player 1					Player 2					Total		
	Share		Profit ($\times 10^5$)			Share		Profit			Profit ($\times 10^5$)		
	RS	EX	RS	EX	%	RS	EX	RS	EX	%	RS	EX	%
N(800,100)	63.8%	62.5%	2.19	1.70	29%	36.2	37.5	1.11	1.01	8.8%	3.30	2.72	21%
N(800,200)	63.5%	62.5%	2.06	1.40	47%	36.5	37.5	1.06	0.84	26%	3.12	2.24	39%
N(800,300)	63.3%	62.5%	1.86	1.10	69%	36.7	37.5	0.96	0.66	46%	2.83	1.77	60%
N(700,100)	63.5%	62.1%	2.12	1.77	20%	36.5	37.9	1.09	1.08	1.2%	3.21	2.84	13%
N(700,200)	63.3%	62.1%	1.92	1.47	31%	36.7	37.9	1.01	0.90	12%	2.93	2.37	23%
N(700,300)	63.0%	62.1%	1.69	1.17	45%	37.0	37.9	0.89	0.72	24%	2.58	1.89	37%

Table 5.1: Numerical results comparing the revenue-sharing contract (RS) with the existing contract (EX).

by increasing the profit of both players. The profit increases in the variability of the demand distribution. Moreover, we observe that the proportional sharing scheme based on marginal costs (our approach) gives slightly more weight to the less cost-effective player as compared to the proportional sharing scheme based on total costs (the existing approach).

5.4 Homogeneous Resource-sharing Models

When resources are homogeneous, they are completely substitutable for one another. The effective capacity is therefore the sum of the individual capacity invested by each player. The alliances among airlines and car rental companies are some of the applications of this model.

In a merger (system), the central planner tries to maximize the aggregate revenue by collectively choosing the capacity investment $\mathbf{K}$, i.e.,

$$\pi_T^* \triangleq \max_{K_i} p\mathbb{E}[\min(L, D)] - \sum_{i=1}^{n} f_i(K_i).$$

where the total capacity investment L is the sum of all K_i's, i.e., $L \triangleq \mathbf{Ke}$ with $\mathbf{e}$ being the column vector with all one's.

Lemma 5.4.1 *Define an auxiliary function*

$$g(\hat{L}) \triangleq \max_{K_i} p\mathbb{E}[\min(L, D)] - \sum_{i=1}^{n} f_i(K_i), \quad s.t. \;\; L \leq \hat{L}.$$

Then $g(\hat{L})$ is concave in $\hat{L}$ where $\hat{L}$ is the budget on total capacity investment.

Proof of Lemma 5.4.1. Suppose L^* is the optimal solution to the system problem. It is easy to see that for all $\hat{L} \geq L^*$, $g(\hat{L}) = \pi_T^*$. For all $L < L^*$, the budget constraint becomes tight. It suffices to show that

$$h(\hat{L}) = \min_{K_i} \sum_{i=1}^{n} f_i(K_i), \text{ s.t. } \sum_{i=1}^{n} K_i = \hat{L}$$

is convex in $\hat{L}$. For any $\lambda \in [0,1]$,

$$h(\lambda \hat{L}_1 + (1-\lambda)\hat{L}_2) = \min_{K_i, K_i'} \sum_{i=1}^{n} f_i(\lambda K_i + (1-\lambda)K_i') \text{ s.t. } \sum_{i=1}^{n} K_i = \hat{L}_1, \ \sum_{i=1}^{n} K_i' = \hat{L}_2,$$

and

$$\begin{aligned} \lambda h(\hat{L}_1) + (1-\lambda)h(\hat{L}_2) = \quad & \min_{K_i, K_i'} \lambda \sum_{i=1}^{n} f_i(K_i) + (1-\lambda) \sum_{i=1}^{n} f_i(K_i') \\ & \text{s.t. } \sum_{i=1}^{n} K_i = \hat{L}_1, \ \sum_{i=1}^{n} K_i' = \hat{L}_2. \end{aligned}$$

By convexity of function f_i for $i = 1, \dots, n$, for any K_i, we know that

$$f_i(\lambda K_i + (1-\lambda)K_i') \geq \lambda f_i(K_i) + (1-\lambda) f_i(K_i').$$

Taking the minimum with respect to the same constraints preserves the inequality, we have

$$h(\lambda \hat{L}_1 + (1-\lambda)\hat{L}_2) \geq \lambda h(\hat{L}_1) + (1-\lambda) h(\hat{L}_2).$$

This completes the proof. □

In a joint-venture with a pre-negotiated revenue-sharing contract β, player i tries to maximize her profit by choosing her profit-maximizing capacity investment level K_i based on other players' strategy K_{-i}, i.e.,

$$\pi_i^N(\beta) \triangleq \max_{K, K_i; K_{-i}} \beta_i p \mathbb{E}[\min(L, D)] - f_i(K_i),$$

which leads to a Nash equilibrium.

Lemma 5.4.2 *The total capacity investment level in a joint-venture is no greater than that in a merger (system), i.e.,* $\sum_{i=1}^{n} K_i^N \leq \sum_{i=1}^{n} K_i^*$.

Proof of Lemma 5.4.2. Suppose that, without loss of generality, $K_1^N \geq K_1^*$. Then we have

$$\frac{\beta p - f_1'(K_1^N)}{\beta p} \leq \frac{\beta p - f_1'(K_1^*)}{\beta p} \leq \frac{p - f_1'(K_1^*)}{p}.$$

Take F^{-1} on both sides (F^{-1} is monotonely increasing, so the sign does not change), then we have $\sum_{i=1}^n K_i^N \leq \sum_{i=1}^n K_i^*$. □

The result in Lemma 5.4.2 does not depend on demand distribution or symmetry among the players. It shows that the effective capacity in a joint venture is always lower compared to a system optimum. However, when the players have asymmetric costs, it is likely that some players over-invest as compared to their counterparts in the optimal setting. In particular, the individual contribution depend on the revenue sharing ratio β.

In contrast to the heterogeneous resource sharing case where an optimal revenue sharing method exists, one can show that there does not exist a fixed revenue sharing method which will induce the system optimal actions in the Nash equilibrium. In other words, there does not exist β such that $\pi_T^N(\beta) = \pi_T^*$.

In the rest of the section, we will investigate the following questions: (1) For a fixed revenue sharing ratio β, how is performance in a joint venture compared to the optimum. (2) How to choose β such that we can have some performance guarantee. We will first restrict ourselves to linear quadratic costs. We begin with a 2-player game and extend our results to a n-player setting. In the end of this section, we will consider n-player setting with general convex costs.

5.4.1 2-player game with linear-quadratic cost functions

Assume that the cost functions are linear-quadratic, i.e.,

$$f_1(K_1) = \frac{a_1(K_1 + b_1)^2}{2} + c_1, \qquad f_2(K_2) = \frac{a_2(K_2 + b_2)^2}{2} + c_2.$$

Without loss of generality, assume that $a_1 \geq a_2$. Now define $\bar{K}_1 = K_1 + b_1$ and $\bar{K}_2 = K_2 + b_2$, and their corresponding modified total capacity investment levels,

$$\bar{L}^N = L^N + b_1 + b_2, \qquad \bar{L}^* = L^* + b_1 + b_2.$$

Lemma 5.4.3 *For a 2-player game with any demand distribution D and linear-quadratic cost functions, for all $\beta_1 \leq 0.5$, the ratio of the total capacity investment level in the system to that in the joint-venture is upper and lower bounded by*

$$1 \geq \frac{\bar{L}^N}{\bar{L}^*} \geq \frac{\beta_1 a_2 + \beta_2 a_1}{a_1 + a_2} \geq \frac{1}{2}.$$

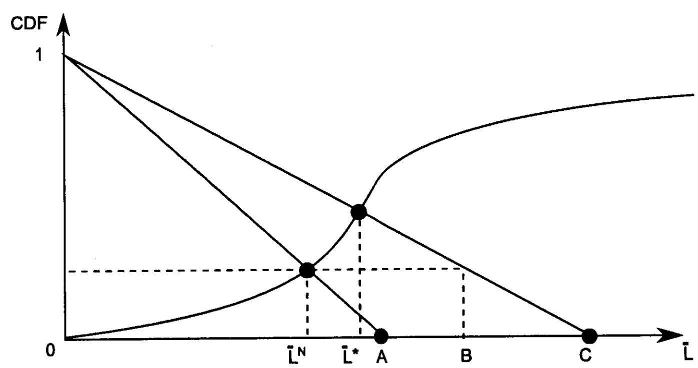

Figure 5-1: A graphical proof for Lemma 5.4.3.

Proof of Lemma 5.4.3. The lower bound is proven by Lemma 5.4.2. Now we show how to obtain an upper bound by utilizing an intercept argument. By optimality conditions, we have

$$\mathbb{P}(D \leq K_1^N + K_2^N) = \frac{\beta_1 p - a_1(K_1^N - b_1)}{\beta_1 p} = \frac{\beta_2 p - a_2(K_2^N - b_2)}{\beta_2 p}.$$

By changing of variables,

$$\mathbb{P}(D + b_1 + b_2 \leq \bar{K}_1^N + \bar{K}_2^N) = \frac{\beta_1 p - a_1 \bar{K}_1^N}{\beta_1 p} = \frac{\beta_2 p - a_2 \bar{K}_2^N}{\beta_2 p}.$$

Then $\beta_2 a_1 \bar{K}_1^N = \beta_1 a_2 \bar{K}_2^N$ and we have

$$\bar{L}^N = \frac{\beta_1 a_2 + \beta_2 a_1}{\beta_1 a_2} \bar{K}_1^N, \qquad \text{or} \qquad \bar{L}^N = \frac{\beta_1 a_2 + \beta_2 a_1}{\beta_2 a_1} \bar{K}_2^N.$$

Thus, we have

$$\mathbb{P}(D + b_1 + b_2 \leq \bar{L}^N) = 1 - \frac{1}{p} \left(\frac{a_1 a_2}{\beta_1 a_2 + \beta_2 a_1} \right) \bar{L}^N. \tag{5.8}$$

By the similar transformation of the first-order condition in the system optimal, we have

$$\mathbb{P}(D + b_1 + b_2 \leq \bar{L}^*) = 1 - \frac{1}{p} \left(\frac{a_1 a_2}{a_1 + a_2} \right) \bar{L}^* \tag{5.9}$$

As shown in Figure 5-1, the horizontal axis is the modified total capacity investment level and the vertical axis is the cumulative distribution function of the demand. The upward sloping curve (cumulative distribution function) represents the left hand sides of (5.8) and (5.9), and the two downward sloping lines represent the right hand sides of (5.8) and (5.9). Thus, $\bar{L}^N$ and $\bar{L}^*$ can be solved graphically. We also observe that

$$\frac{\bar{L}^*}{\bar{L}^N} \leq \frac{B}{\bar{L}^N} = \frac{C}{A} = \frac{a_1 + a_2}{\beta_1 a_2 + \beta_2 a_1}.$$

where the points C and A are the x-intercepts which can be evaluated from (5.8) and (5.9). □

Lemma 5.4.3 shows that for a 2-player game with linear-quadratic costs, the effective modified capacity in a joint venture depends on both the cost asymmetry as well as the revenue sharing ratio. However, the worst case, $\bar{L}^* = 2\bar{L}^N$, can happen under two circumstances: (1) equal revenue sharing ($\beta_1 = \beta_2$) and independent of cost asymmetry, and/or (2) with symmetric players ($a_1 = a_2$) and independent of revenue sharing contracts (with the assumption that $\beta_1 \leq 0.5$. Intuitively, dividing revenue equally among asymmetric entities sounds like a bad idea. It is surprising to see that having symmetric players in a joint venture could lead to the worst outcome, and having different revenue sharing contracts might not mitigate its impact. Note that when $\beta_1 > 0.5$, it is easy to construct examples that worst case becomes unbound.

Lemma 5.4.3 also highlights a notable difference between the homogeneous and the heterogeneous resource pooling. Note that in Theorem 5.3.3 for the heterogeneous resources, we have shown that the optimal revenue sharing rule suggests that every player should be compensated proportionally to his share of the marginal cost to the aggregate marginal cost. That if, if $a_1 \geq a_2$, the optimal way to share revenue must follow that $\beta_1 \geq \beta_2$. Lemma 5.4.3 implies the exact opposite, i.e., in order to have the worst case performance guarantee, given $a_1 \geq a_2$, then $\beta_1 \leq \beta_2$!

The intuition is that for heterogeneous resource pooling, the effective capacity of the entire system is constrained by a bottleneck capacity due to certain key players. To induce these players to produce at K^*, they have to be awarded such that they are willing to produce at K^* but not lower. Now consider homogeneous resource pooling, every player can contribute to the effective capacity, the only difference is the cost. Therefore, one should encourage the cost efficient player to produce more and discourage those with higher cost. It is captured by a lower revenue sharing ratio for the player with higher marginal cost.

This observation on a 2-player game can be generalized to a n-player game as shown in the following proposition.

Proposition 5.4.4 *Consider a n-player game with cost structure $a_1 \geq a_2 \cdots \geq a_n$ and revenue sharing contract $\beta_1 \leq \beta_2 \cdots \leq \beta_n$. Under any demand distribution D and any linear-quadratic cost functions, the ratio of the total capacity investment level in the system to that in the joint-venture is upper and lower bounded by*

$$1 \geq \frac{\bar{L}^N}{\bar{L}^*} \geq \frac{\sum_{i=1}^n \frac{\beta_i}{a_i}}{\sum_{i=1}^n \frac{1}{a_i}} \geq \frac{1}{n}.$$

With n-players, the worst case in terms of the effective capacity is $\bar{L}^* = n\bar{L}^N$, i.e., the worst case of a joint venture decreases as the number of participants increases. The result is intuitive as with more parties involved, it becomes increasingly challenging to coordinate the joint venture. Similar to the 2-player game studied earlier, the worst case occurs with symmetric players and/or equal sharing of the revenue when players are asymmetric.

In the next theorem, we will show that the profit generated in a joint venture can be bounded by the optimal profit.

Theorem 5.4.5 *For a 2-player game with any linear-quadratic cost functions and any demand distribution with mode* m*, we have*

$$\frac{\pi_T^N(\beta)}{\pi_T^*} \geq \frac{1}{2}, \qquad \text{for all } a_1 \geq a_2 \text{ and } \beta_1 \in \left[\frac{mp+1}{2mp+(a_1/a_2+1)}, \frac{1}{2}\right].$$

Moreover, the optimal β_1^* *that maximizes the total joint-venture profit falls in the following interval,*

$$\beta_1^* \in \left[\frac{1}{a_1/a_2+1}, \frac{mp+1}{2mp+(a_1/a_2+1)}\right].$$

Proof of Theorem 5.4.5. By Lemma 5.4.3, we know that

$$\bar{L}^N = \frac{\beta_1 a_2 + \beta_2 a_1}{\beta_1 a_2} \bar{K}_1^N, \qquad \text{or} \qquad \bar{L}^N = \frac{\beta_1 a_2 + \beta_2 a_1}{\beta_2 a_1} \bar{K}_2^N.$$

The Nash profit functions can be expressed as functions of $\bar{L}^N$ i.e.,

$$\pi_T^N(\beta) = p\mathbb{E}[\min(\bar{L}^N - b_1 - b_2, D)] - \left(\frac{a_1 a_2^2 \beta_1^2 + a_2 a_1^2 \beta_2^2}{2(a_2\beta_1 + a_1\beta_2)^2}\right) \bar{L}^{N2} - c_1 - c_2. \quad (5.10)$$

If we impose a budget constraint $L \leq \bar{L}^N$ on the system optimal, the budget-constrained system optimal profit can also be expressed as functions of $\bar{L}^N$ i.e.,

$$g(\bar{L}^N) = p\mathbb{E}[\min(\bar{L}^N - b_1 - b_2, D)] - \left(\frac{a_1 a_2}{2(a_2 + a_1)}\right) \bar{L}^{N2} - c_1 - c_2,$$

Observe that $g(\bar{L}^N) = \pi_T^N(\beta)$ when $\beta_1 = \frac{1}{2}$. By Lemma 5.4.3, we know that for all $\beta_1 \leq \frac{1}{2}$, $\frac{\bar{L}^*}{\bar{L}^N} \leq 2$. In addition, $g(\bar{L})$ is concave in $\bar{L}$ by Lemma 5.4.1. Thus, we have

$$\frac{\pi_T^N(\beta_1 = \frac{1}{2})}{\pi_T^*} = \frac{\pi_T^N(\beta_1 = \frac{1}{2})}{g(\bar{L}^*)} \geq \frac{\pi_T^N(\beta_1 = \frac{1}{2})}{2g(\bar{L}^*/2)} \geq \frac{\pi_T^N(\beta_1 = \frac{1}{2})}{2g(\bar{L}^N)} = \frac{1}{2}.$$

Now let $\bar{D} = D + b_1 + b_2$. Since

$$\mathbb{P}(\bar{D} \le \bar{L}^N) = 1 - \frac{1}{p}\left(\frac{a_1a_2}{\beta_1 a_2 + (1-\beta_1)a_1}\right)\bar{L}^N \Rightarrow \frac{\mathbf{d}\bar{L}^N}{\mathbf{d}\beta_1} = \frac{-\frac{\bar{L}^N}{p}\left(\frac{a_1a_2(a_1-a_2)}{(\beta_1a_2+(1-\beta_1)a_1)^2}\right)}{f_{\bar{D}}(\bar{L}^N) + \frac{1}{p}\left(\frac{a_1a_2}{\beta_1a_2+(1-\beta_1)a_1}\right)},$$

We have by (5.10),

$$\begin{aligned}
& \frac{\mathbf{d}\pi_T^N(\beta_1)}{\mathbf{d}\beta_1} \\
= \; & \frac{-(1-\mathbb{P}(\bar{D} \le \bar{L}^N))\bar{L}^N\left(\frac{a_1a_2(a_1-a_2)}{(\beta_1a_2+(1-\beta_1)a_1)^2}\right)}{f_{\bar{D}}(\bar{L}^N) + \frac{1}{p}\left(\frac{a_1a_2}{\beta_1a_2+(1-\beta_1)a_1}\right)} \\
& + \frac{a_1^2a_2^2(1-2\beta_1)}{(a_2\beta_1 + a_1(1-\beta_1))^3}\bar{L}^{N2} - \left(\frac{a_1a_2^2\beta_1^2 + a_2a_1^2(1-\beta_1)^2}{(\beta_1a_2+(1-\beta_1)a_1)^2}\right)\bar{L}^N\frac{\mathbf{d}\bar{L}^N}{\mathbf{d}\beta_1} \\
= \; & \frac{\bar{L}^{N2}}{(\beta_1a_2+(1-\beta_1)a_1)^3}\left(-\frac{a_1^2a_2^2(a_1-a_2)(a_1+a_2)\beta_1(1-\beta_1)}{p(\beta_1a_2+(1-\beta_1)a_1)f_{\bar{D}}(\bar{L}^N) + a_1a_2} + a_1^2a_2^2(1-2\beta_1)\right).
\end{aligned}$$

If the mode of D is m, then $\pi_T^N(\beta_1)$ is decreasing in β_1 for all

$$\beta \in \left[\frac{mp + a_2}{2mp + a_1 + a_2}, \frac{1}{2}\right],$$

and $\pi_T^N(\beta_1)$ is increasing in β_1 for all

$$\beta \in \left[0, \frac{a_2}{a_1 + a_2}\right].$$

Thus, the optimal β_1^* lies in the following interval

$$\beta^* \in \left[\frac{a_2}{a_1+a_2}, \frac{mp + a_2}{2mp + a_1 + a_2}\right].$$

This completes the proof. □

In Theorem 5.4.5, we propose an interval for which the aggregate Nash profit is guaranteed to achieve at least half of the optimal profit. The interval depends on the cost asymmetry between the two players and the mode of demand. In particular, the interval shrinks as the two players have more similar cost structure, i.e., with two fully symmetric players, the best revenue sharing ratio asks for an equal division of the revenue. On the other hand, the interval widens as the mode of demand increases, i.e., if the demand distribution is flatter, our proposed revenue sharing contracts have more rooms for error in capturing the peak demand.

For a n-player game, we show that an equal revenue sharing scheme could guarantee a worst case performance of at least $1/n$ of the optimal profit as shown in the following proposition.

Proposition 5.4.6 *For a n-player game with any linear-quadratic cost functions and any demand distribution, if we choose $\beta_i = 1/n$, i.e., dividing the aggregate revenue equally among all the players, we have*

$$\frac{\pi_T^N(\beta)}{\pi_T^*} \geq \frac{1}{n}.$$

Proof of Proposition 5.4.6. From a 2-player setting, one can see that the profit functions can be expressed as functions of $\bar{L}$,

$$\pi_T(\bar{L}) \quad = \quad p\mathbb{E}[\min(\bar{L} + b_1 + b_2, D)] - \left(\frac{a_1}{2(a_2 + a_1)}\right)\bar{L}^2 - c_1 - c_2.$$

Note that it is equivalent to $\pi_T^N(\beta_1, \beta_2)$ when $\beta_1 = \beta_2 = 0.5$, where

$$\pi_T^{\mathrm{N}}(\beta) \quad = \quad p\mathbb{E}[\min(\bar{L}^{\mathrm{N}} + b_1 + b_2, D)] - \left(\frac{a_1 a_2 \beta_1^2 + a_1^2 \beta_2^2}{2(a_2\beta_1 + a_1\beta_2)^2}\right)\bar{L}^{\mathrm{N}2} - c_1 - c_2.$$

In Lemma 5.4.1, we have shown the concavity of $\pi_T(\bar{L})$. Then by making use of the bound on investment level as shown in Proposition 5.4.4, we obtain the desired result. □

5.4.2 n-player game with general convex costs

We consider n-player games with asymmetric convex cost functions. Denote $f = (f_i(K_i))_{i=1}^n$ as general convex cost functions. Let $\pi^N(f)$ and $\pi^*(f)$ be the Nash and system profit of n players with respect to the general cost f, respectively. Define the

Price of Anarchy with respect to f as

$$POA(f) = \frac{\pi^N(f)}{\pi^*(f)}.$$

We first show that $POA(f)$ can be lower bounded by $POA(\bar{f})$ where $\bar{f}$ is a set of modified *linear* cost functions.

Proposition 5.4.7 *The price of anarchy on the total profit of a joint venture is lower bounded by*

$$POA(f) = \frac{\pi^N(f)}{\pi^*(f)} \geq \frac{\pi^N(\bar{f})}{\pi^*(\bar{f})} = POA(\bar{f}),$$

where $\bar{f} = (\bar{f}_1, \ldots, \bar{f}_n)$ *are linear cost functions such that* $\bar{f}_i = \alpha_i \cdot K_i$ *where* $\alpha_i = f_i'(K_i^N)$.

Proof of Proposition 5.4.7. By convexity of f_i for all $i = 1, \ldots, n$, we know that

$$f_i(K_i^*) \geq f_i(K_i^N) + f_i'(K_i^N)(K_i^* - K_i^N).$$

Therefore

$$\begin{aligned} POA(f) &= \frac{p\mathbb{E}[\min(L^N, D)] - \sum_{i=1}^n f_i(K_i^N)}{p\mathbb{E}[\min(L^*, D)] - \sum_{i=1}^n f_i(K_i^*)} \qquad (5.11) \\ &\geq \frac{p\mathbb{E}[\min(L^N, D)] - \sum_{i=1}^n f_i(K_i^N)}{p\mathbb{E}[\min(L^*, D)] - \sum_{i=1}^n (f_i(K_i^N) + f_i'(K_i^N)(K_i^* - K_i^N))}. \end{aligned}$$

Since

$$0 = f_i(0) \geq f_i(K_i^N) + f_i'(K_i^N)(-K_i^N) \Rightarrow f_i(K_i^N) - f_i'(K_i^N)(K_i^N) \leq 0, \qquad (5.12)$$

we add (5.12) onto both the numerator and denominator of (5.11),

$$\begin{aligned} &POA(f) \\ \geq &\frac{p\mathbb{E}[\min(L^N, D)] + \sum_{i=1}^n (-f_i(K_i^N) + f_i(K_i^N) - f_i'(K_i^N)(K_i^N))}{p\mathbb{E}[\min(L^*, D)] + \sum_{i=1}^n (-f_i(K_i^N) - f_i'(K_i^N)(K_i^* - K_i^N) + f_i(K_i^N) - f_i'(K_i^N)(K_i^N))} \end{aligned}$$

$$\geq \frac{p\mathbb{E}[\min(L^N, D)] - \sum_{i=1}^n f_i'(K_i^N)(K_i^N)}{p\mathbb{E}[\min(L^*, D)] - \sum_{i=1}^n f_i'(K_i^N)(K_i^*)}.$$

Now let $\tilde{K}_i^N$ and $\tilde{K}_i^*$ be the Nash Equilibrium solution and the system optimal solution with respect to the same problem but with the modified linear cost functions such that $\bar{f}_i = \alpha_i \cdot K_i$ where $\alpha_i = f_i'(K_i^N)$. Correspondingly, $\tilde{L}^N = \sum_{i=1}^n \tilde{K}_i^N$ and $\tilde{L}^* = \sum_{i=1}^n \tilde{K}_i^*$.

Since $\tilde{K}_i^N = K_i^N$ (having the same set of first-order conditions), we have

$$p\mathbb{E}[\min(L^N, D)] - \sum_{i=1}^n f_i'(K_i^N)(K_i^N) = p\mathbb{E}[\min(\tilde{L}^N, D)] - \sum_{i=1}^n \alpha_i \tilde{K}_i^N.$$

Because $\tilde{K}_i^*$ is the optimal capacity investment level for the modified problem, it implies that

$$p\mathbb{E}[\min(L^*, D)] - \sum_{i=1}^n f_i'(K_i^N)(K_i^*) \leq p\mathbb{E}[\min(\tilde{L}^*, D)] - \sum_{i=1}^n \alpha_i \tilde{K}_i^*.$$

Thus, we have

$$POA(f) \geq \frac{p\mathbb{E}[\min(\tilde{L}^N, D)] - \sum_{i=1}^n \alpha_i \tilde{K}_i^N}{p\mathbb{E}[\min(\tilde{L}^*, D)] - \sum_{i=1}^n \alpha_i \tilde{K}_i^*} \geq \frac{\pi^N(\bar{f})}{\pi^*(\bar{f})} = POA(\bar{f}).$$

This completes the proof. □

By making use of Proposition 5.4.7, we can obtain a lower bound on the profit by using the cost asymmetry factor and the ratio between the investment levels in the Nash and the system optimum.

Lemma 5.4.8 *Price of anarchy on the total profit of a joint venture is lowered bounded by*

$$POA(\bar{f}) = \frac{\pi^N(\bar{f})}{\pi^*(\bar{f})} \geq \tilde{\alpha} \frac{\tilde{L}^N}{\tilde{L}^*},$$

where the cost asymmetry factor is given by

$$\tilde{\alpha} = \frac{\min_i \alpha_i}{\max_i \alpha_i} \leq 1.$$

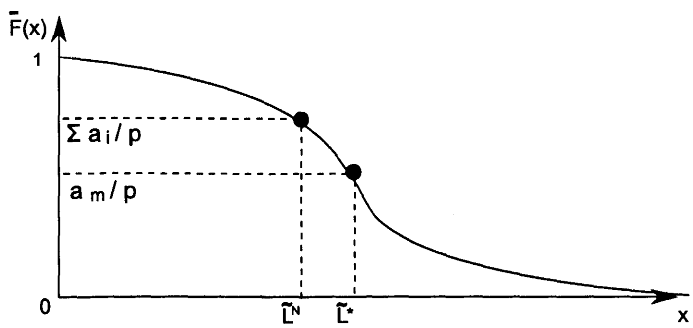

Figure 5-2: A graphical proof for Lemma 5.4.8.

Proof of Lemma 5.4.8. Assume that, without loss of generality, $\alpha_m = \alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_n = \alpha_M$. Define the set $P = \{j \mid \alpha_j = \alpha_m\}$. If $|P| = s$, s symmetric players invest in the system optimal solution and therefore $\tilde{L}^* = s\tilde{K}_j^*$ for $i \in P$.

$$
\begin{aligned}
POA(\bar{f}) &= \frac{p\int_0^{\tilde{L}^N} \bar{F}_D(x)dx - \sum_{i=1}^N \alpha_i \tilde{K}_i^N}{p\int_0^{\tilde{L}^N} \bar{F}_D(x)dx + p\int_{\tilde{L}^N}^{\tilde{L}^*} \bar{F}_D(x)dx - \alpha_m \tilde{L}^N} \\
&\geq \frac{\sum_{i=1}^N \alpha_i \tilde{L}^N - \sum_{i=1}^N \alpha_i \tilde{K}_i^N}{\sum_{i=1}^N \alpha_i \tilde{L}^N + \sum_{i=1}^N \alpha_i (\tilde{L}^* - \tilde{L}^N) - \alpha_m \tilde{L}^*} \\
&\geq \frac{\sum_{i=1}^N \alpha_i (\tilde{L}^N - \tilde{K}_i^N)}{\sum_{i=1}^N \alpha_i \tilde{L}^* - \alpha_m \tilde{L}^*} \\
&\geq \frac{\alpha_m (n-1)\tilde{L}^N}{\alpha_M (n-1)\tilde{L}^*} \geq \tilde{\alpha}\frac{\tilde{L}^N}{\tilde{L}^*}.
\end{aligned}
$$

where the cost asymmetry factor $\tilde{\alpha} = \alpha_m/\alpha_M \leq 1$. This completes the proof. □

Note that equal revenue sharing induces equal marginal costs for every player in a Nash equilibrium, since $\beta_i = \alpha_i / \sum_{j=1}^n \alpha_j$. Therefore, $\tilde{\alpha} = 1$, and the comparison between the profit can be reduced to a comparison between the total investment level, i.e, $\frac{\pi^N(\bar{f})}{\pi^*(\bar{f})} \geq \frac{\tilde{L}^N}{\tilde{L}^*}$

Next, we will present the how the profit in a joint venture can be bounded from below by the system optimum. Define the demand spread

$$\tilde{\theta} = \frac{\theta_m}{\theta_M} = \frac{\max f_D(x)}{\min f_D(y)},$$

where $x \leq \tilde{L}^N \leq y \leq \tilde{L}^*$.

Theorem 5.4.9

$$POA(f) \geq \tilde{\alpha}\frac{1-n\bar{r}}{1-n\bar{r}+(n-1)\bar{r}\tilde{\theta}},$$

where $\bar{r} = \max_i \alpha_i/p$, *and* $\tilde{\theta} \geq 1$ *measures the demand spread.*

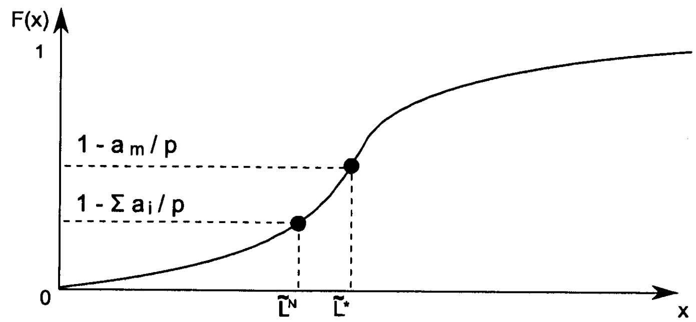

Figure 5-3: A graphical proof for Theorem 5.4.9.

Proof of Theorem 5.4.9. First we lower bound the ratio of $\tilde{L}^N$ to $\tilde{L}^*$.

$$\begin{aligned}
\frac{\tilde{L}^N}{\tilde{L}^*} &\geq \frac{\tilde{L}^N}{\tilde{L}^N + (\sum_{i=1}^n \alpha_i - \alpha_m)/(\theta_m p)} \\
&\geq \frac{(1-\sum_{i=1}^n \alpha_i/p)/\theta_M}{(1-\sum_{i=1}^n \alpha_i/p)/\theta_M + (\sum_{i=1}^n \alpha_i - \alpha_m)/(\theta_m p)} \\
&= \frac{p - \sum_{i=1}^n \alpha_i}{p - \sum_{i=1}^n \alpha_i + (\sum_{i=1}^n \alpha_i - \alpha_m)\tilde{\theta}} \\
&\geq \frac{p - n\alpha_M}{p - n\alpha_M + (n-1)\alpha_M\tilde{\theta}} \\
&= \frac{1-n\bar{r}}{1-n\bar{r}+(n-1)\bar{r}\tilde{\theta}},
\end{aligned}$$

where $\bar{r} = \alpha_M/p$. This result then follows from Lemme 5.4.8. □

Note that when D is uniform, the demand spread $\tilde{\theta} = 1$, we have

$$POA(f) \geq \tilde{\alpha}\frac{1-n\bar{r}}{1-\bar{r}}.$$

Figure 5-4, 5-5 and 5-6 show the lower bounds on POA with uniform demand, normal demand $N(400, 100)$ and exponential demand $\exp(400)$, respectively. The lower bound on POA decreases as the number of players increases or the marginal cost to price ratio increases. We also observe that the lower bound on POA has a steeper rate of decrease when the demand spead is higher. Note that in our simulation, the exponential demand has the highest demand spread ($\tilde{\theta} = 7.35$), followed by the normal demand ($\tilde{\theta} = 3.86$) and then the uniform demand ($\tilde{\theta} = 1$).

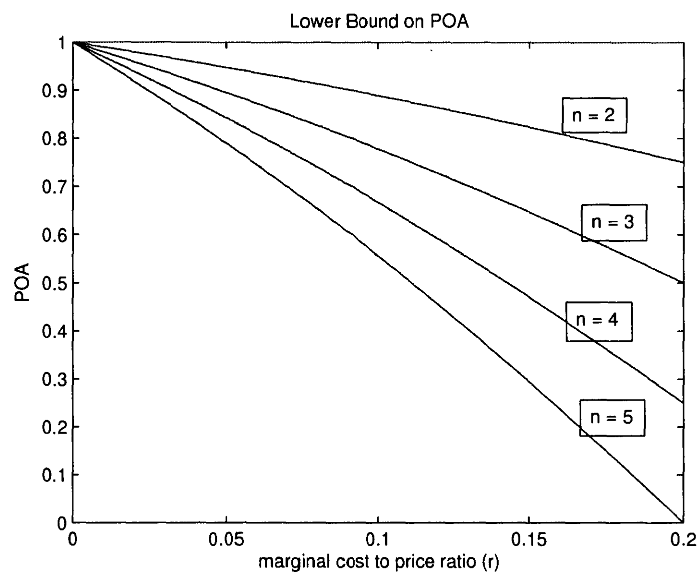

Figure 5-4: Lower bound on price of anarchy for uniform demand.

5.5 Conclusion

In this work, we study resource pooling and capacity planning in joint ventures under uncertainties. We distinguish two types of resources pooling, based on whether the resources are heterogeneous or homogeneous. When resources are heterogeneous, the effective capacity in a joint venture is constrained by the lowest level of contribution from one participant. We have shown that every participant is committed to make an equal contribution in a joint venture with heterogeneous resources. We have also shown that, there exists a same efficient and fair revenue sharing scheme in both Nash equilibrium and Nash Bargaining solution. The optimal scheme rewards every participant proportionally to his marginal cost. When resources are homogeneous, however, there does not exist a revenue sharing scheme which induces actions to

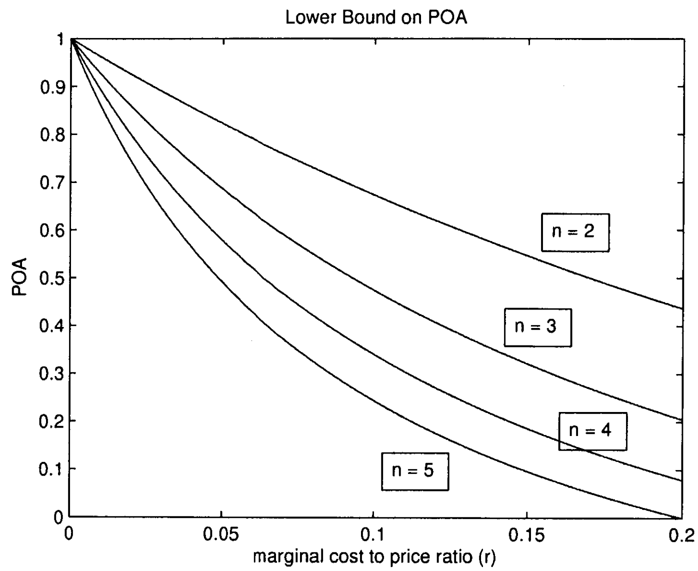

Figure 5-5: Lower bound on price of anarchy for normal demand.

achieve the optimum. Nonetheless, we propose some methods to share revenue with the worst case performance guarantee. The methods suggest that the reward should be inversely proportional to the marginal cost of each participant with homogeneous resources.

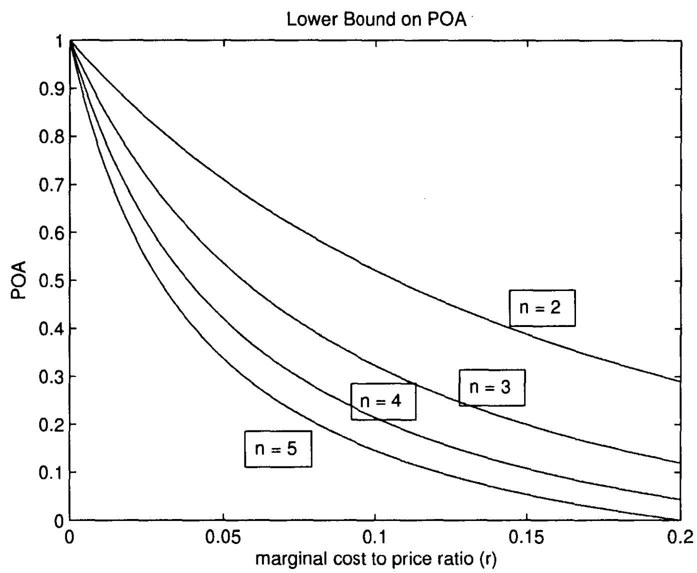

Figure 5-6: Lower bound on price of anarchy for exponential demand.

Chapter 6

Conclusions

Inefficiency caused by decentralization has been the subject of extensive research in operations management in recent years. In this thesis, we studied its impact on supply chains, congested systems and joint ventures.

The study on supply chains has revealed that when downstream retailers compete with substitutes, product substitutability promotes output levels, channel profit and social welfare. Although asymmetry deteriorates the performance, a decentralized supply chain with substitutes is fairly efficient. The opposite happens for complements. Although asymmetry has some countervailing effect, a decentralized chain with complements exhibits a significant loss of efficiency, which suggests that large potential gains could be achieved through coordination. In addition, a decentralized supply chain is relatively more efficient compared to its optimal counterpart when the demand is concave rather than convex.

An important take-away from our analysis is that for substitutes, price-only contracts are often "good enough" in the sense that there is limited room for potential gains from implementing other more complex contracts. In addition, as price-only contracts disproportionately favor the supplier, she has less incentive to adopt other contracts. The results provide some partial intuition that may help explain the popularity of price-only contracts in practice. Nevertheless, our results suggest that complex coordinating contracts should be applied to decentralized supply chains with complementary products as the loss due to lack of coordination could be huge. One could also argue that in this case, both channel members have an incentive to coordinate as they both could benefit from a profit increase.

The analysis on congested systems with profit-driven service providers shows that the impact of decentralization primarily depends on the nature of costs. When the costs are fully self-contained, maximum welfare loss is limited to 25% of the social optimum, even in the presence of highly nonlinear convex cost. The inefficiency decreases further with competition among the service providers. However, with the spillover cost, the potential welfare loss could be arbitrarily severe. The results provide some evidences on the usefulness of airport congestion pricing as potential gains from coordination could be huge.

The last topic in this thesis focuses on capacity planning and resource pooling in joint ventures under demand uncertainties. We have shown that the performance of a joint venture heavily depends on the nature of resources. In particular, when resources are heterogenous or not fully substitutable, the effective capacity in a joint venture is constrained by the lowest level of contribution from one participant. The optimal revenue sharing scheme rewards every participant proportionally to his marginal cost. When resources are homogeneous, however, there does not exist a revenue sharing scheme which induces actions to achieve the optimum. Nonetheless, we propose some methods to share revenue with the worst case performance guarantee. The methods suggest that the reward should be inversely proportional to the marginal cost of each participant with homogeneous resources.

The methodology that we utilize is a departure from traditional approaches and thus gives rise to new and interesting theoretical and computational challenges. To establish the analytical bounds in this thesis, we have utilized tools from matrix analysis such as Cassini ovals of eigenvalues, M-matrix and copositivity (Horn and Johnson 1985). We believe the methodology proposed in this thesis could potentially be used in other problems.

Appendix A

Appendix for Chapter 2

A.1 Useful Matrix Analysis Results

Proofs in the paper utilize a number of results from matrix analysis. For completeness, we state them in the following lemmas (see Horn and Johnson 1985 for more information) .

A.1.1 M-matrices

A square matrix whose off-diagonal elements are nonpositive is called a Z-matrix. A symmetric Z-matrix is a M-matrix if and only if it is positive definite. M-matrices enjoy several structural properties some of which are listed below.

Let $\mathbf{A}$ be an M-matrix and $\mathbf{B}$ be a Z-matrix such that $\mathbf{A} \leq \mathbf{B}$:

- $\mathbf{A}^{-1}$ exists and $\mathbf{A}^{-1} \geq \mathbf{0}$;
- $\mathbf{B}$ is an M-matrix and $\mathbf{B}^{-1} \leq \mathbf{A}^{-1}$;
- $\mathbf{AB}^{-1}$ and $\mathbf{B}^{-1}\mathbf{A}$ are M-matrices;
- Any sum of M-matrices is still an M-matrix;
- Any principle submatrix of $\mathbf{A}$ is an M-matrix.
- If $\mathbf{D}$ is a positive diagonal matrix, then $\mathbf{DA}$ and $\mathbf{AD}$ are M-matrices.

A.1.2 Inverse binomial theorem

If $\mathbf{A}$, $\mathbf{U}$, $\mathbf{B}$, $\mathbf{V}$ are matrices with appropriate dimensions, then$(\mathbf{A}+\mathbf{UBV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{UB}\left(\mathbf{B}+\mathbf{BVA}^{-1}\mathbf{UB}\right)^{-1}\mathbf{BVA}^{-1}$, provided $\mathbf{A}$ and $\mathbf{B}+\mathbf{BVA}^{-1}\mathbf{UB}$ are nonsingular.

Lemma A.1.1 ***(Inverse of M-matrices)*** *If* $\mathbf{A}$ *and* $\mathbf{B}$ *are M-matrices and* $\mathbf{B} \geq \mathbf{A}$, *then* $\mathbf{B}^{-1} \leq \mathbf{A}^{-1}$.

Lemma A.1.2 ***(Brauer's Ovals Theorem)*** *Let* $\mathbf{A}$ *be a square matrix,* $R_i(\mathbf{A}) = \sum_{j\neq i}|a_{ij}|$, *for* $1 \leq i \leq n$. *Cassini ovals are defined as,* $O_{ij} = \{|z-a_{ii}||z-a_{jj}| \leq R_i(\mathbf{A})R_j(\mathbf{A})\}$ *for all* $i \neq j$. *Then, all the eigenvalues of* $\mathbf{A}$ *lie inside the union of the Cassini ovals, i.e.,* $\lambda(\mathbf{A}) \in \cup_{i\neq j} O_{ij}$.

A.2 Preliminary Results: Bounds on the Minimum Eigenvalue

Lemma A.2.1 *The minimum eigenvalue of matrix* $\mathbf{G} = \mathbf{\Gamma}^{-1/2}\mathbf{B}\mathbf{\Gamma}^{-1/2}$ *is bounded by* $\lambda_{\min}(\mathbf{G}) \geq 1 - r_{(2)}(\mathbf{B}) \geq 1 - r_{(1)}(\mathbf{B})$.

Proof of Lemma A.2.1. The matrix $\mathbf{G}$ and $\mathbf{\Gamma}^{-1}\mathbf{B}^T$ are similar matrices, thus, $\lambda_{\min}(\mathbf{G}) = \lambda_{\min}(\mathbf{\Gamma}^{-1}\mathbf{B}^T)$. Notice $R_i(\mathbf{\Gamma}^{-1}\mathbf{B}^T) = \sum_{j\neq i}\left|\frac{\beta_{ji}}{\alpha_i}\right| = r_i(\mathbf{B})$. By Brauer's theorem (Lemma A.1.2), we can write down the following:

$$\begin{aligned}(1-\lambda_{\min}(\mathbf{\Gamma}^{-1}\mathbf{B}^T))^2 &\leq \max_{i,j|i\neq j} r_i(\mathbf{B})r_j(\mathbf{B}) \\ \Rightarrow \lambda_{\min}(\mathbf{\Gamma}^{-1}\mathbf{B}^T) &\geq 1 - r_{(2)}.(\mathbf{B})\end{aligned}$$

Since $r_{(2)}(\mathbf{B}) \leq r_{(1)}(\mathbf{B})$, it completes the proof. □

A.3 Proof for the Results in the Paper

A.3.1 Proof of Theorem 2.3.1 $\mathbf{e}^T\mathbf{q}_d/\mathbf{e}^T\mathbf{q}_c$

Upper bound

Let $\mathbf{X}$ be a diagonal matrix with positive diagonal elements such that $\mathbf{Xe} = \mathbf{p}_c - \mathbf{c}$.

$$\begin{aligned}
\frac{\mathbf{e}^T\mathbf{q}_d}{\mathbf{e}^T\mathbf{q}_c} &= \frac{\mathbf{e}^T\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma}\mathbf{X}\mathbf{e}}{\mathbf{e}^T\mathbf{B}\mathbf{X}\mathbf{e}} \\
&= \frac{\mathbf{w}^T(\mathbf{B})^{-1/2}\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma}\mathbf{X}(\mathbf{B})^{-1/2}\mathbf{w}}{\mathbf{w}^T\mathbf{w}}, \text{ where } \mathbf{w} = (\mathbf{B})^{1/2}\mathbf{e} \\
&\leq \lambda_{\max}\left((\mathbf{B})^{-1/2}\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma}\mathbf{X}(\mathbf{B})^{-1/2}\right) \\
&= \lambda_{\max}\left((\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma}\right) \quad \text{(By similar matrices)} \\
&= \frac{1}{\lambda_{\min}\left(\mathbf{\Gamma}^{-1}(\mathbf{B}+\mathbf{\Gamma})\right)} \\
&= \frac{1}{1+\lambda_{\min}(\mathbf{\Gamma}^{-1}\mathbf{B})}
\end{aligned}$$

Because the bound is decreasing in $\lambda_{\min}(\mathbf{\Gamma}^{-1}\mathbf{B})$, we can upper bound it by lower bound the eigenvalue. By Lemma A.2.1, it implies

$$\frac{\mathbf{e}^T\mathbf{q}_d}{\mathbf{e}^T\mathbf{q}_c} \leq \frac{1}{2-r_{(2)}(\mathbf{B})} \leq \frac{1}{2-r_{(1)}(\mathbf{B})}.$$

Lower bound

To prove the lower bound that $\frac{\mathbf{e}^T\mathbf{q}_d}{\mathbf{e}^T\mathbf{q}_c}$ is always larger than 1/2, we are trying to prove

$$\begin{aligned}
\frac{\mathbf{e}^T\mathbf{q}_d}{\mathbf{e}^T\mathbf{q}_c} &= \frac{\mathbf{e}^T\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma}\mathbf{X}\mathbf{e}}{\mathbf{e}^T\mathbf{B}\mathbf{X}\mathbf{e}} \geq \frac{1}{2} \\
&\Rightarrow 2\mathbf{e}^T\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma}\mathbf{X}\mathbf{e} - \mathbf{e}^T\mathbf{B}\mathbf{X}\mathbf{e} \geq 0 \\
&\Rightarrow (\mathbf{q_d})^T \underbrace{\left(2\mathbf{B}^{-1}(\mathbf{B}+\mathbf{\Gamma})\mathbf{\Gamma}^{-1}\mathbf{X}^{-1} - \mathbf{B}^{-1}(\mathbf{B}+\mathbf{\Gamma})\mathbf{\Gamma}^{-1}\mathbf{X}^{-1}\mathbf{B}\mathbf{\Gamma}^{-1}(\mathbf{B}+\mathbf{\Gamma})\mathbf{B}^{-1}\right)}_{\Phi(\mathbf{B})} \mathbf{q_d} \geq 0,
\end{aligned}$$

that is, it is equivalent to prove $\Phi(\mathbf{B})$ is nonnegative. To see this, rewrite $\Phi(\mathbf{B})$ as follows,

$$\begin{aligned}\Phi(\mathbf{B}) &= \mathbf{B}^{-1}(\mathbf{B}+\mathbf{\Gamma})\mathbf{\Gamma}^{-1}\mathbf{X}^{-1}\left(2\mathbf{I}-\mathbf{B}\mathbf{\Gamma}^{-1}(\mathbf{B}+\mathbf{\Gamma})\mathbf{B}^{-1}\right)\\ &= (\mathbf{\Gamma}^{-1}+\mathbf{B}^{-1})\mathbf{X}^{-1}\left(2\mathbf{I}-(\mathbf{M}\mathbf{\Gamma}^{-1}+\mathbf{I})\right)\\ &= (\mathbf{\Gamma}^{-1}+\mathbf{B}^{-1})\mathbf{X}^{-1}\left(\mathbf{\Gamma}-\mathbf{B}^{-1}\right)\mathbf{\Gamma}^{-1}\end{aligned}$$

$\Phi(\mathbf{B})$ is clearly nonnegative because it is a product of four positive/nonngative matrices. □

A.3.2 Proof of Theorem 2.3.1 CS_d/CS_c and TS_d/TS_c

Preliminary work

Given the demand function in terms of prices as $\mathbf{q}(\mathbf{p}) = \mathbf{q}(\mathbf{0}) - \mathbf{B}\mathbf{p}$, its corresponding demand function in quantities is given by $\mathbf{p}(\mathbf{q}) = \bar{\mathbf{p}} - \mathbf{B}^{-1}\mathbf{q}$, where $\bar{\mathbf{p}} = \mathbf{B}^{-1}\mathbf{p}(\mathbf{0})$. Consumer surplus (CS) is the difference between the utility which a representative consumer derives from consuming $\mathbf{q}$ units of products ($\mathbf{U}(\mathbf{q}) = \bar{\mathbf{p}}^T\mathbf{q} - \frac{1}{2}\mathbf{q}^{\mathbf{T}}\mathbf{B}^{-1}\mathbf{q}$) and the cost he spends on acquiring them ($\mathbf{q}^{\mathbf{T}}\mathbf{p}$).

$$CS = \mathbf{U}(\mathbf{q}) - \mathbf{p}^{\mathbf{T}}\mathbf{q}(\mathbf{p}) = \frac{1}{2}\mathbf{q}(\mathbf{p})^T\mathbf{B}^{-1}\mathbf{q}(\mathbf{p}). \tag{A.1}$$

The total surplus (TS) is defined as the sum of consumer surplus and producer surplus which is the channel profit in the model.

$$TS = PS(\mathbf{p}) + CS(\mathbf{p}) = (\mathbf{p}(\mathbf{q}) - \mathbf{c})^T\,\mathbf{q}(\mathbf{p}) + \frac{1}{2}\mathbf{q}(\mathbf{p})^T\mathbf{B}^{-1}\mathbf{q}(\mathbf{p}). \tag{A.2}$$

Upper bound

From Equation (A.1), the ratio between the consumer surplus is given by

$$\frac{CS_d}{CS_c} = \frac{1/2(\mathbf{q}_d)^T\mathbf{B}^{-1}\mathbf{q}_d}{1/2(\mathbf{q}_c)^T\mathbf{B}^{-1}\mathbf{q}_c}$$

$$
\begin{aligned}
&= \frac{\mathbf{w}^T\mathbf{B}^{-1/2}\boldsymbol{\Gamma}(\mathbf{B}+\boldsymbol{\Gamma})^{-1}\mathbf{B}(\mathbf{B}+\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}\mathbf{B}^{-1/2}\mathbf{w}}{\mathbf{w}^T\mathbf{w}}, \text{ where } \mathbf{w} = \mathbf{B}^{1/2}(\mathbf{p}_c - \mathbf{c}) \\
&\le \lambda_{\max}\left((\mathbf{B}^{1/2}(\mathbf{B}+\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}\mathbf{B}^{-1/2})^T(\mathbf{B}^{1/2}(\mathbf{B}+\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}\mathbf{B}^{-1/2})\right) \\
&= \lambda^2_{\max}(\mathbf{B}^{1/2}(\mathbf{B}+\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}\mathbf{B}^{-1/2}) \\
&= \lambda^2_{\max}((\mathbf{B}+\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}) \quad \text{(By similar matrices)} \\
&= \frac{1}{\lambda^2_{\min}(\boldsymbol{\Gamma}^{-1}(\mathbf{B}+\boldsymbol{\Gamma}))} \\
&= \frac{1}{(1+\lambda_{\min}(\boldsymbol{\Gamma}^{-1}\mathbf{B}))^2}
\end{aligned}
$$

. Now consider the total surplus defined in Equation (A.2),

$$
\frac{TS_d}{TS_c} = \frac{(\mathbf{p}_c - \mathbf{c})^T(4\mathbf{B}+3\boldsymbol{\Gamma})(\mathbf{B}+\boldsymbol{\Gamma})^{-1}\mathbf{B}(\mathbf{B}+\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}(\mathbf{p}_c - \mathbf{c})}{(\mathbf{p}_c - \mathbf{c})^T 3\mathbf{B}(\mathbf{p}_c - \mathbf{c})}.
$$

Denote $\mathbf{w} = \mathbf{B}^{1/2}(\mathbf{p}_c - \mathbf{c})$, then the expression becomes

$$
\begin{aligned}
\frac{TS_d}{TS_c} &= \frac{1}{3}\frac{\mathbf{w}^T \overbrace{\mathbf{B}^{-1/2}(4\mathbf{B}+3\boldsymbol{\Gamma})(\mathbf{B}+\boldsymbol{\Gamma})^{-1}\mathbf{B}(\mathbf{B}+\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}\mathbf{B}^{-1/2}}^{\mathbf{A}}\mathbf{w}}{\mathbf{w}^T\mathbf{w}} \\
&\le \frac{1}{3}\lambda_{\max}(\mathbf{A}) \\
&= \frac{1}{3}\lambda_{\max}\left((4\mathbf{B}+3\boldsymbol{\Gamma})(\mathbf{B}+\boldsymbol{\Gamma})^{-1}(\mathbf{B}+\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}\right) \\
&= \frac{1}{3}\lambda_{\max}\left((4\mathbf{I}-\boldsymbol{\Gamma}(\mathbf{B}+\boldsymbol{\Gamma})^{-1})(\mathbf{B}+\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}\right) \\
&= \frac{1}{3}\lambda_{\max}\left(4(\boldsymbol{\Gamma}^{-1}\mathbf{B}+\mathbf{I})^{-1} - (\boldsymbol{\Gamma}^{-1}\mathbf{B}+\mathbf{I})^{-2}\right).
\end{aligned}
$$

The function $4/x + 1/x^2$ decrease in x, thus the maximum eigenvalue of matrix $\mathbf{A}$ is obtained at the minimum eigenvalue of matrix $\boldsymbol{\Gamma}^{-1}\mathbf{B}+\mathbf{I}$.

$$
\begin{aligned}
\frac{TS_d}{TS_c} &\le \frac{1}{3}\left[\frac{4}{1+\lambda_{\min}(\boldsymbol{\Gamma}^{-1}\mathbf{B})} - \frac{1}{(1+\lambda_{\min}(\boldsymbol{\Gamma}^{-1}\mathbf{B}))^2}\right] \\
&= \frac{4\lambda_{\min}(\boldsymbol{\Gamma}^{-1}\mathbf{B})+3}{3(1+\lambda_{\min}(\boldsymbol{\Gamma}^{-1}\mathbf{B}))^2}.
\end{aligned}
$$

The ratio is decreasing in $\lambda_{\min}(\boldsymbol{\Gamma}^{-1}\mathbf{B})$, thus, we can upper bound the ratio by using Lemma A.2.1 to obtain the desired upper bound for CS and TS.

Lower bound

To prove the lower bound, i.e., $\frac{CS_d}{CS_c} \geq \frac{1}{4}$, it is equivalent to prove the following,

$$
\begin{aligned}
& \frac{\mathbf{q}_d^T \mathbf{B}^{-1} \mathbf{q}_d}{\mathbf{q}_c^T \mathbf{B}^{-1} \mathbf{q}_c} \geq \frac{1}{4} \\
\Leftrightarrow\ & 4\mathbf{q}_d^T \mathbf{B}^{-1} \mathbf{q}_d - \mathbf{q}_c^T \mathbf{B}^{-1} \mathbf{q}_c \geq 0 \\
\Leftrightarrow\ & \mathbf{q}_d^T \underbrace{\left(4\mathbf{B}^{-1} - \mathbf{B}^{-1}(\mathbf{B}+\mathbf{\Gamma})\mathbf{\Gamma}^{-1}\mathbf{B}\mathbf{\Gamma}^{-1}(\mathbf{B}+\mathbf{\Gamma})\mathbf{B}^{-1}\right)}_{\mathbf{\Phi}(\mathbf{B})} \mathbf{q}_d \geq 0
\end{aligned}
$$

To prove that matrix $\mathbf{\Phi}(\mathbf{B})$ is nonnegative, notice that

$$
\begin{aligned}
\mathbf{\Phi}(\mathbf{B}) &= 4\mathbf{B}^{-1} - (\mathbf{B}^{-1} + \mathbf{\Gamma}^{-1})\mathbf{B}(\mathbf{B}^{-1} + \mathbf{\Gamma}^{-1}) \\
&= 4\mathbf{B}^{-1} - \left(\mathbf{\Gamma}^{-1}\mathbf{B}\mathbf{\Gamma}^{-1} + 2\mathbf{\Gamma}^{-1} + \mathbf{B}^{-1}\right) \\
&= 3\mathbf{B}^{-1} - 2\mathbf{\Gamma}^{-1} - \mathbf{\Gamma}^{-1}\mathbf{B}\mathbf{\Gamma}^{-1} \\
&= (3\mathbf{B}^{-1} + \mathbf{\Gamma}^{-1})(\mathbf{I} - \mathbf{B}\mathbf{\Gamma}^{-1}).
\end{aligned}
$$

$\mathbf{\Phi}(\mathbf{B})$ is the product of two nonnegative matrices, thus, it is also nonnegative.

To prove the lower bound that $\frac{TS_d}{TS_c} \geq \frac{7}{12}$, it is equivalent to prove $\Phi(\mathbf{B}) = \mathbf{B}^{-1}(\mathbf{B}+\mathbf{\Gamma})\mathbf{\Gamma}^{-1}\left(4(4\mathbf{B}+3\mathbf{\Gamma})(\mathbf{B}+\mathbf{\Gamma})^{-1} - 7\mathbf{B}\mathbf{\Gamma}^{-1}(\mathbf{B}+\mathbf{\Gamma})\mathbf{B}^{-1}\right)$ is nonnegative.

$$
\begin{aligned}
& (\mathbf{q_d})^T \Phi(\mathbf{B}) \mathbf{q_d} \geq 0 \\
\Rightarrow\ & (\mathbf{p}_c - \mathbf{c})^T \left(4(4\mathbf{B}+3\mathbf{\Gamma})(\mathbf{B}+\mathbf{\Gamma})^{-1} - 7\mathbf{B}\mathbf{\Gamma}^{-1}(\mathbf{B}+\mathbf{\Gamma})\mathbf{B}^{-1}\right) \mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma}(\mathbf{p}_c - \mathbf{c}) \geq 0 \\
\Rightarrow\ & (\mathbf{p}_c - \mathbf{c})^T \left(4(4\mathbf{B}+3\mathbf{\Gamma})(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma} - 7\mathbf{B}\right)(\mathbf{p}_c - \mathbf{c}) \geq 0 \\
\Rightarrow\ & \frac{(\mathbf{p}_c - \mathbf{c})^T(4\mathbf{B}+3\mathbf{\Gamma})(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma}(\mathbf{p}_c - \mathbf{c})}{(\mathbf{p}_c - \mathbf{c})^T(3\mathbf{B})(\mathbf{p}_c - \mathbf{c})} \geq \frac{7}{12} \\
\Rightarrow\ & \frac{TS_d}{TS_c} \geq \frac{7}{12}
\end{aligned}
$$

Similarly, we express $\Phi(\mathbf{B})$ as $\mathbf{\Gamma}^{1/2}\Phi(\mathbf{G})\mathbf{\Gamma}^{1/2}$ and prove $\Phi(\mathbf{G})$ is copositive.

$$
\begin{aligned}
\Phi(\mathbf{G}) &= \mathbf{G}^{-1}(\mathbf{G}+\mathbf{I})(4(4\mathbf{G}+3\mathbf{I})(\mathbf{G}+\mathbf{I})^{-1} - 7(\mathbf{G}+\mathbf{I})) \\
&= 4\mathbf{G}^{-1}(4\mathbf{G}+3\mathbf{I}) - 7\mathbf{G}^{-1}(\mathbf{G}+\mathbf{I})^2
\end{aligned}
$$

$$
\begin{aligned}
&= 16\mathbf{I} + 12\mathbf{G}^{-1} - 7(\mathbf{G} + 2\mathbf{I} + \mathbf{G}^{-1}) \\
&= 2\mathbf{I} - 7\mathbf{G} + 5\mathbf{G}^{-1} \\
&= (\mathbf{I} - \mathbf{G})(5\mathbf{G}^{-1} + 7\mathbf{I})
\end{aligned}
$$

$\Phi(\mathbf{G})$ is a product of nonnegative matrices and that has established the bound. □

Proof of Proposition 2.3.3

Given the setting, $r_{(2)}(\mathbf{B}) = \frac{n-1}{n\sqrt{k}} = \delta/\sqrt{k}$. By Theorem 2.3.1, $\pi_d/\pi_c(\mathbf{B}) \leq \frac{3+2\delta/\sqrt{k}}{(2+\delta/\sqrt{k})^2}$, where the uppe bound decreases in k, i.e., when the asymmetry factor $k > 1$, $\pi_d/\pi_c(\mathbf{B}) \leq \frac{3+2\delta/\sqrt{k}}{(2+\delta/\sqrt{k})^2} < \frac{3+2\delta}{(2+\delta)^2} = \pi_d/\pi_c|_{k=1}$. The last equality holds as the bounds are tight for symmetric retailers where $k = 1$. □

Proof of Proposition 2.3.2

We first derive the profits earned by the retailers $(\pi_d)_R$ and the supplier $(\pi_d)_S$ respectively. For complements,

$$
\begin{aligned}
(\pi_d)_R &= \sum_{i=1}^{n} (\pi_d)_{r_i} \\
&= (\mathbf{p}_d - \mathbf{w}_d)^T \mathbf{q}_d \\
&= \frac{1}{2}\left((\mathbf{B}+\mathbf{\Gamma})^{-1}(2\mathbf{B}+\mathbf{\Gamma})(\mathbf{p}_c - \mathbf{c}) + 2\mathbf{c} - (\mathbf{p}_c + \mathbf{c})\right)^T \mathbf{q}_d \\
&= \frac{1}{2}\left((\mathbf{B}+\mathbf{\Gamma})^{-1}(2\mathbf{B}+\mathbf{\Gamma})(\mathbf{p}_c - \mathbf{c}) - (\mathbf{p}_c - \mathbf{c})\right)^T \mathbf{q}_d \\
&= \frac{1}{2}\left((\mathbf{B}+\mathbf{\Gamma})^{-1}((2\mathbf{B}+\mathbf{\Gamma}) - \mathbf{I})(\mathbf{p}_c - \mathbf{c})\right)^T \mathbf{q}_d \\
&= \frac{1}{2}\left((\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{B}(\mathbf{p}_c - \mathbf{c})\right)^T \mathbf{q}_d \\
&= \frac{1}{4}(\mathbf{p}_c - \mathbf{c})^T \mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma}(\mathbf{p}_c - \mathbf{c}) \\
(\pi_d)_S &= (\mathbf{w}_d - \mathbf{c})^T \mathbf{q}_d \\
&= \frac{1}{4}(\mathbf{p}_c - \mathbf{c})^T \mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma}(\mathbf{p}_c - \mathbf{c})
\end{aligned}
$$

Let $\mathbf{w} = \mathbf{B}^{1/2}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma}(\mathbf{p}_c - \mathbf{c})$, then we obtain

$$\frac{(\pi_d)_R}{(\pi_d)_S} = \frac{\mathbf{w}^T\mathbf{B}^{-1/2}(\mathbf{B}+\mathbf{\Gamma})\mathbf{\Gamma}^{-1}\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{B}^{1/2}\mathbf{w}}{\mathbf{w}^T\mathbf{B}^{-1/2}(\mathbf{B}+\mathbf{\Gamma})\mathbf{\Gamma}^{-1}(\mathbf{B}+\mathbf{\Gamma})(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{B}^{1/2}\mathbf{w}}.$$

The lower bound is obtained by noticing

$$\begin{aligned}
\frac{(\pi_d)_R}{(\pi_d)_S} &= \frac{\mathbf{w}^T\overbrace{\mathbf{A}\mathbf{G}\mathbf{A}^{-1}}^{\mathbf{Q}}\mathbf{w}}{\mathbf{w}^T\mathbf{A}(\mathbf{G}+\mathbf{I})\mathbf{A}^{-1}\mathbf{w}}, \text{ where } \mathbf{A} = \mathbf{B}^{-1/2}(\mathbf{B}+\mathbf{\Gamma}) \text{ and } \mathbf{G} = \mathbf{\Gamma}^{-1}\mathbf{B} \\
&= \frac{\mathbf{z}^T\mathbf{z}}{\mathbf{z}^T\left(\mathbf{I}+\mathbf{Q}^{-1}\right)\mathbf{z}}, \text{ where } \mathbf{z} = \mathbf{Q}^{1/2}\mathbf{w} \\
&\geq \frac{1}{\lambda_{\max}\left(\mathbf{I}+\mathbf{Q}^{-1}\right)} \\
&= \frac{1}{1+\frac{1}{\lambda_{\min}(\mathbf{Q})}} \\
&= \frac{\lambda_{\min}(\mathbf{Q})}{\lambda_{\min}(\mathbf{Q})+1} \\
&= \frac{\lambda_{\min}(\mathbf{G})}{\lambda_{\min}(\mathbf{G})+1}, \text{ since } \lambda_{\min}(\mathbf{Q}) = \lambda_{\min}(\mathbf{A}\mathbf{G}\mathbf{A}^{-1}) = \lambda_{\min}(\mathbf{G})
\end{aligned}$$

The bound is increasing in $\lambda_{\max}(\mathbf{G})$, thus, it is can be lower bounded by lower bounding $\lambda_{\min}(\mathbf{G})$ by using Lemma A.2.1, we obtain the two lower bounds, i.e.,

$$\frac{(\pi_d)_R}{(\pi_d)_S} \geq \frac{1-r_{(2)}(\mathbf{B})}{2-r_{(2)}(\mathbf{B})} \geq \frac{1-r_{(1)}(\mathbf{B})}{2-r_{(1)}(\mathbf{B})}.$$

To prove the upper bound that $\frac{(\pi_d)_R}{(\pi_d)_S} \leq \frac{1}{2}$, it is equivalent to prove that matrix $\Phi(\mathbf{B})$ is copositive, where $\Phi(\mathbf{B}) = \mathbf{B}^{-1}(\mathbf{B}+\mathbf{\Gamma})\mathbf{\Gamma}^{-1}(\mathbf{I}-2\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1})$. To see this, by definition of copositivity,

$$\begin{aligned}
& (\mathbf{q}_d)^T\Phi(\mathbf{B})\mathbf{q}_d \geq 0 \\
\Rightarrow\ & \left(\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma}(\mathbf{p}_c-\mathbf{c})\right)^T\Phi(\mathbf{B})\left(\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma}(\mathbf{p}_c-\mathbf{c})\right) \geq 0 \\
\Rightarrow\ & (\mathbf{p}_c-\mathbf{c})^T\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma}(\mathbf{p}_c-\mathbf{c}) \geq (\mathbf{p}_c-\mathbf{c})^T2\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma}(\mathbf{p}_c-\mathbf{c}) \\
\Rightarrow\ & \frac{(\mathbf{p}_c-\mathbf{c})^T\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma}(\mathbf{p}_c-\mathbf{c})}{(\mathbf{p}_c-\mathbf{c})^T\mathbf{B}(\mathbf{B}+\mathbf{\Gamma})^{-1}\mathbf{\Gamma}(\mathbf{p}_c-\mathbf{c})} \leq \frac{1}{2}
\end{aligned}$$

$$\Rightarrow \quad \frac{(\pi_d)_R}{(\pi_d)_S} \leq \frac{1}{2}$$

$$\begin{aligned}
\Phi(\mathbf{B}) &= (\mathbf{B}^{-1} + \mathbf{\Gamma}^{-1})(\mathbf{I} - 2\mathbf{B}(\mathbf{B} + \mathbf{\Gamma})^{-1}) \\
&= (\mathbf{B}^{-1} + \mathbf{\Gamma}^{-1})(\mathbf{\Gamma} - \mathbf{B})(\mathbf{B} + \mathbf{\Gamma})^{-1} \\
&= (\mathbf{I} - \mathbf{\Gamma}^{-1}\mathbf{B} + \mathbf{B}^{-1}\mathbf{\Gamma} - \mathbf{I})(\mathbf{B} + \mathbf{\Gamma})^{-1} \\
&= (\mathbf{I} - \mathbf{\Gamma}^{-1}\mathbf{B}\mathbf{\Gamma}^{-1}\mathbf{B})\mathbf{B}^{-1}\mathbf{\Gamma}(\mathbf{B} + \mathbf{\Gamma})^{-1} \\
&= (\mathbf{I} - \mathbf{\Gamma}^{-1}\mathbf{B})(\mathbf{I} + \mathbf{\Gamma}^{-1}\mathbf{B})\mathbf{B}^{-1}\mathbf{\Gamma}(\mathbf{B} + \mathbf{\Gamma})^{-1} \\
&= (\mathbf{I} - \mathbf{\Gamma}^{-1}\mathbf{B})(\mathbf{B}^{-1}\mathbf{\Gamma} + \mathbf{I})(\mathbf{B} + \mathbf{\Gamma})^{-1} \\
&= (\mathbf{\Gamma} - \mathbf{B})(\mathbf{B}^{-1} + \mathbf{\Gamma}^{-1})(\mathbf{B} + \mathbf{\Gamma})^{-1}
\end{aligned}$$

We have shown that $\Phi(\mathbf{B})$ is a nonnegative matrix by expressing it as a product of three nonnegative matrices, thus, the upper bound of $\frac{1}{2}$ holds for substitutes. □

Appendix B

Appendix for Chapter 4

B.1 Proof of Proposition 4.2.9

Proof. For every access fee per service, $\mathbf{t} = (t_1, ..., t_n)$, the corresponding output level in a subgame perfect equilibrium, $\mathbf{q^{SPE}(t)}$, must satisfy the following profit-maximizing condition for every service provider i for all i, $\bar{p}_i - t_i - \sum_j \beta_{ij} q_j^{SPE}(t_j) - \beta_{ii} q_i^{SPE}(t_i) - \partial c_i(\mathbf{q})/\partial q_i|_{\mathbf{q}=\mathbf{q^{SPE}(t)}} = 0$.

The total social welfare is given by

$$W(\mathbf{q})CS(\mathbf{q}) + PS(\mathbf{q}) + TR(\mathbf{q}) \quad = \quad \sum_i q_i(\bar{p}_i - \sum_j \beta_{ij} q_j) - \sum_i c_i(\mathbf{q}).$$

One way to find out the optimal congestion pricing is to substitute $\mathbf{q^{SPE}(t)}$ into the profit-maximizing condition, i.e., $W(\mathbf{q^{SPE}(t)})$, and maximize it with respect to $\mathbf{t}$. Alternatively, we can first determine the optimal service level, $\mathbf{q}^*$, that should be maintained in this facility so as to maximize the total welfare. The optimal service level, $\mathbf{q}^*$, must satisfy, $\bar{p}_i - \sum_j \beta_{ij} q_j^* - \partial c_i(\mathbf{q})/\partial q_i = 0|_{\mathbf{q}=\mathbf{q}^*}$. By comparing it with equilibrium condition, we obtain the desired result on the access fee as a function of output level $\mathbf{q}$ and satisfies $\mathbf{q^{SPE}(t^*)} = \mathbf{q}^*$. □

B.2 Proof of Lemma 4.3.1.

Proof of Lemma 4.3.1. The welfare objective is given by

$$W(\mathbf{q}) = \mathbf{q}^T(\bar{\mathbf{p}} - \frac{1}{2}\mathbf{B}\mathbf{q} - \mathbf{l}(\mathbf{q})).$$

The optimality condition can be written as $\nabla W(\mathbf{q}) = \bar{\mathbf{p}} - \mathbf{B}\mathbf{q} - \mathbf{l}(\mathbf{q}) - \mathbf{R}\mathbf{q} = \mathbf{0}$, where matrix $\mathbf{R}$ denotes the Jacobian matrix of function $\mathbf{l}(\mathbf{q})$. It is important to note that the matrix $\mathbf{R}$ depends on the output level $\mathbf{q}$. Since $\nabla W(\mathbf{q}^*) = 0$, we get

$$\bar{\mathbf{p}} - \mathbf{l}(\mathbf{q}^*) = (\mathbf{B} + \mathbf{R}^*)\mathbf{q}^*, \text{ or} \tag{B.1}$$

$$\mathbf{q}^* = (\mathbf{B} + \mathbf{R}^*)^{-1}(\bar{\mathbf{p}} - \mathbf{l}(\mathbf{q}^*)). \tag{B.2}$$

Substitute Equation (B.2) into the welfare objective, we obtain the following:

$$\begin{aligned} W(\mathbf{q}^*) &= (\mathbf{q}^*)^T \left(\bar{\mathbf{p}} - \mathbf{l}(\mathbf{q}^*) - \frac{1}{2}\mathbf{B}(\mathbf{B} + \mathbf{R}^*)^{-1}(\bar{\mathbf{p}} - \mathbf{l}(\mathbf{q}^*))\right) \\ &= (\mathbf{q}^*)^T(\frac{1}{2}\mathbf{B} + \mathbf{R}^*)(\mathbf{B} + \mathbf{R}^*)^{-1}(\bar{\mathbf{p}} - \mathbf{l}(\mathbf{q}^*)) \\ &= (\mathbf{q}^*)^T(\frac{1}{2}\mathbf{B} + \mathbf{R}^*)\mathbf{q}^*. \end{aligned}$$

The profit function of service provider i is given by $\pi_i = q_i(\bar{p}_i - \sum_j \beta_{ij}q_j - l_i(\mathbf{q}))$. The equilibrium condition for all service providers can be written in the matrix form: $\bar{\mathbf{p}} - \mathbf{B}\mathbf{q} - \mathbf{l}(\mathbf{q}) - \mathbf{\Gamma_B}\mathbf{q} - \mathbf{\Gamma_R}\mathbf{q}$, where $\mathbf{\Gamma_B}$ and $\mathbf{\Gamma_R}$ represent the diagonal matrix of $\mathbf{B}$ and $\mathbf{R}$ respectively. Since $\mathbf{q}^{\mathbf{N}}$ satisfies the equilibrium condition, we obtain

$$\bar{\mathbf{p}} - \mathbf{l}(\mathbf{q}^{\mathbf{N}}) = (\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^N})\mathbf{q}^{\mathbf{N}}, \text{ or} \tag{B.3}$$

$$\mathbf{q}^{\mathbf{N}} = (\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^N})^{-1}(\bar{\mathbf{p}} - \mathbf{l}(\mathbf{q}^{\mathbf{N}})). \tag{B.4}$$

Substituting Equation (B.4) into the welfare objective gives rise to the welfare achieved in the unregulated setting:

$$W(\mathbf{q}^N) = (\mathbf{q}^{\mathbf{N}})^T \left(\bar{\mathbf{p}} - \mathbf{l}(\mathbf{q}^{\mathbf{N}}) - \frac{1}{2}\mathbf{B}(\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^N})^{-1}(\bar{\mathbf{p}} - \mathbf{l}(\mathbf{q}^{\mathbf{N}}))\right)$$

$$= (\mathbf{q^N})^T(\frac{1}{2}\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^N})(\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^N})^{-1}(\bar{\mathbf{p}} - \mathbf{l}(\mathbf{q^N}))$$
$$= (\mathbf{q^N})^T(\frac{1}{2}\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^N})\mathbf{q^N}.$$

□

B.3 Proof of Lemma 4.3.2

Proof of Lemma 4.3.2. By convexity on cost function $\mathbf{l}(\mathbf{q}) = (l_1(\mathbf{q}), \cdots, l_n(\mathbf{q}))$ (Assumption 4.2.2),

$$\mathbf{l}(\mathbf{q}^*) - \mathbf{l}(\mathbf{q^N}) \geq \mathbf{R^N}(\mathbf{q}^* - \mathbf{q^N})$$
$$\Rightarrow \mathbf{l}(\mathbf{q}^*) - \mathbf{R^N}\mathbf{q}^* \geq \mathbf{l}(\mathbf{q^N}) - \mathbf{R^N}\mathbf{q^N}$$
$$\Rightarrow -\mathbf{l}(\mathbf{q}^*) + \mathbf{R^N}\mathbf{q}^* \leq -\mathbf{l}(\mathbf{q^N}) + \mathbf{R^N}\mathbf{q^N}.$$

Adding a positive vector $\bar{\mathbf{p}}$ to both sides maintains the inequality, i.e,

$$\bar{\mathbf{p}} - \mathbf{l}(\mathbf{q}^*) + \mathbf{R^N}\mathbf{q}^* \leq \bar{\mathbf{p}} - \mathbf{l}(\mathbf{q^N}) + \mathbf{R^N}\mathbf{q^N}.$$

After substituting the optimality conditions derived Equation (B.1) and (B.3), we obtain the desired result: $(\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^N} + \mathbf{R^N})\mathbf{q^N} \geq (\mathbf{B} + \mathbf{R}^* + \mathbf{R^N})\mathbf{q}^*$. □

B.4 Proof of Lemma 4.3.5

Proof. By Assumption 4.2.2, function $\mathbf{l}(\mathbf{q}) = (l_1(\mathbf{q}), \ldots, l_n(\mathbf{q}))$ is convex componentwise with $\mathbf{q}$.

$$\mathbf{l}(\mathbf{q^N}) - \mathbf{l}(\mathbf{q}^*) \geq \mathbf{R}^*(\mathbf{q^N} - \mathbf{q}^*)$$
$$\Rightarrow \mathbf{l}(\mathbf{q^N}) - \mathbf{R}^*\mathbf{q^N} \geq \mathbf{l}(\mathbf{q}^*) - \mathbf{R}^*\mathbf{q}^*$$
$$\Rightarrow -\mathbf{l}(\mathbf{q^N}) + \mathbf{R}^*\mathbf{q^N} \leq -\mathbf{l}(\mathbf{q}^*) + \mathbf{R}^*\mathbf{q}^*$$
$$\Rightarrow \bar{\mathbf{p}} - \mathbf{l}(\mathbf{q^N}) + \mathbf{R}^*\mathbf{q^N} \leq \bar{\mathbf{p}} - \mathbf{l}(\mathbf{q}^*) + \mathbf{R}^*\mathbf{q}^*.$$

After substituting the optimality conditions from Equation (B.1) and (B.3), we obtain the desired result. □

B.5 Lemma B.5.1 and its proof

Lemma B.5.1 *The optimal societal welfare is bounded by the following,*

$$W(\mathbf{q}^*) \leq \frac{1}{2}(\mathbf{q}^{\mathbf{N}})^T(\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^N} + \mathbf{R^N})(\mathbf{B} + 2\mathbf{R^N})^{-1}(\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^N} + \mathbf{R^N})\mathbf{q^N}.$$

Proof of Lemma B.5.1. Denote $\mathbf{\Omega} = \mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R^N} + \mathbf{R^N}$ and $\mathbf{\Sigma} = \mathbf{B} + \mathbf{R}^* + \mathbf{R^N}$. By Lemma 4.3.2, we know that $\mathbf{\Omega q^N} \geq \mathbf{\Sigma q}^*$. Note that both $\mathbf{\Omega q^N}$ and $\mathbf{\Sigma q}^*$ are two nonnegative vectors. By Lemma 4.3.1,

$$W(\mathbf{q}^*) = (\mathbf{q}^*)^T\mathbf{\Sigma}\mathbf{\Sigma}^{-1}(\frac{1}{2}\mathbf{B} + \mathbf{R}^*)\mathbf{\Sigma}^{-1}\mathbf{\Sigma}\mathbf{q}^* \leq (\mathbf{q^N})^T\mathbf{\Omega}\mathbf{\Sigma}^{-1}(\frac{1}{2}\mathbf{B} + \mathbf{R}^*)\mathbf{\Sigma}^{-1}\mathbf{\Omega}\mathbf{q^N},$$

where we have replaced $\mathbf{\Sigma q}^*$ by $\mathbf{\Omega q^N}$. Expand this expression further,

$$\begin{aligned}
&W(\mathbf{q}^*)\\
\leq&\frac{1}{2}(\mathbf{q^N})^T\mathbf{\Omega}\mathbf{\Sigma}^{-1}(\mathbf{B} + 2\mathbf{R}^*)\mathbf{\Sigma}^{-1}\mathbf{\Omega}\mathbf{q^N}\\
=&\frac{1}{2}(\mathbf{q^N})^T\mathbf{\Omega}(\mathbf{B} + 2\mathbf{R^N})^{-0.5}\underbrace{(\mathbf{B} + 2\mathbf{R^N})^{0.5}\mathbf{\Sigma}^{-1}(\mathbf{B} + 2\mathbf{R}^*)\mathbf{\Sigma}^{-1}(\mathbf{B} + 2\mathbf{R^N})^{0.5}}_{\mathbf{\Delta}}\\
&(\mathbf{B} + 2\mathbf{R^N})^{-0.5}\mathbf{\Omega}\mathbf{q^N}.
\end{aligned}$$

By the definition of the maximum eigenvalue, this expression is upper bounded by,

$$W(\mathbf{q}^*) \leq \frac{1}{2}\lambda_{\max}\{\mathbf{\Delta}\}(\mathbf{q^N})^T\mathbf{\Omega}(\mathbf{B} + 2\mathbf{R^N})^{-1}\mathbf{\Omega}\mathbf{q^N} \tag{B.5}$$

Now let us focus on this composite matrix $\mathbf{\Delta}$. By the property of similar matrices, $\lambda_{max}\{\Delta\} = \lambda_{\max}\{\mathbf{\Sigma}^{-1}(\mathbf{B} + 2\mathbf{R}^*)\mathbf{\Sigma}^{-1}(\mathbf{B} + 2\mathbf{R^N})\}$. Under the Assumptions 4.2.1 and 4.2.2, it is clear that matrix $\lambda_{\max}\{\mathbf{\Delta}\} \geq 0$ since $\mathbf{\Delta}$ is positive semi-definite. Expand this matrix,

$$
\begin{aligned}
&\lambda_{\max}\{\Delta\}\\
=&\lambda_{\max}\{(\mathbf{B}+\mathbf{R}^*+\mathbf{R}^{\mathbf{N}})^{-1}(\mathbf{B}+2\mathbf{R}^*)(\mathbf{B}+\mathbf{R}^*+\mathbf{R}^{\mathbf{N}})^{-1}(\mathbf{B}+2\mathbf{R}^{\mathbf{N}})\}\\
=&\lambda_{\max}\{\left(\mathbf{I}+(\mathbf{B}+\mathbf{R}^*+\mathbf{R}^{\mathbf{N}})^{-1}(\mathbf{R}^*-\mathbf{R}^{\mathbf{N}})\right)\left(\mathbf{I}+(\mathbf{B}+\mathbf{R}^*+\mathbf{R}^{\mathbf{N}})^{-1}(\mathbf{R}^{\mathbf{N}}-\mathbf{R}^*)\right)\}\\
=&\lambda_{\max}\{\mathbf{I}-\left((\mathbf{B}+\mathbf{R}^*+\mathbf{R}^{\mathbf{N}})^{-1}(\mathbf{R}^*-\mathbf{R}^{\mathbf{N}})\right)^2\}\\
=&1-\lambda_{\min}\{\left((\mathbf{B}+\mathbf{R}^*+\mathbf{R}^{\mathbf{N}})^{-1}(\mathbf{R}^*-\mathbf{R}^{\mathbf{N}})\right)^2\}.
\end{aligned}
$$

It is clear that $\lambda_{\min}\{\left((\mathbf{B}+\mathbf{R}^*+\mathbf{R}^{\mathbf{N}})^{-1}(\mathbf{R}^*-\mathbf{R}^{\mathbf{N}})\right)^2\}\geq 0$ because it is also a positive semidefinite matrix. As a result, $\lambda_{\max}\{\Delta\}\leq 1$. From Equation (B.5), $W(\mathbf{q}^*)\leq \frac{1}{2}(\mathbf{q}^{\mathbf{N}})^T\mathbf{\Omega}(\mathbf{B}+2\mathbf{R}^{\mathbf{N}})^{-1}\mathbf{\Omega}\mathbf{q}^{\mathbf{N}}$. □

B.6 Lemma B.6.1 and its proof

Lemma B.6.1 $\frac{W(\mathbf{q}^{\mathbf{N}})}{W(\mathbf{q}^*)}\leq\kappa(1-\lambda_{\min}\{((\mathbf{I}-\mathbf{\Xi})(\mathbf{G}+\mathbf{I}+\mathbf{\Xi})^{-1}))^2\})$, *where* $\kappa\geq 1$ *is the Jacobian similarity factor,* $\mathbf{G}=\mathbf{\Gamma}_{\mathbf{B}}^{-0.5}(\mathbf{B}+2\mathbf{\Gamma}_{\mathbf{R}}^*)\mathbf{\Gamma}_{\mathbf{B}}^{-0.5}$ *and* $\mathbf{\Xi}=\mathbf{\Gamma}_{\mathbf{B}}^{-0.5}\mathbf{R}_{\mathbf{off}}^*\mathbf{\Gamma}_{\mathbf{B}}^{-0.5}$.

Proof of Lemma B.6.1. From Lemma 4.3.1, we obtain that

$$
W(\mathbf{q}^N) \;=\; (\mathbf{q}^{\mathbf{N}})^T(\frac{1}{2}\mathbf{B}+\mathbf{\Gamma}_{\mathbf{B}}+\mathbf{\Gamma}_{\mathbf{R}}^{\mathbf{N}})\mathbf{q}^{\mathbf{N}}=(\mathbf{q}^{\mathbf{N}})^T\mathbf{\Psi}\mathbf{\Psi}^{-1}(\frac{1}{2}\mathbf{B}+\mathbf{\Gamma}_{\mathbf{B}}+\mathbf{\Gamma}_{\mathbf{R}}^{\mathbf{N}})\mathbf{\Psi}^{-1}\mathbf{\Psi}\mathbf{q}^{\mathbf{N}},
$$

where $\mathbf{\Psi}=\mathbf{B}+\mathbf{\Gamma}_{\mathbf{B}}+\mathbf{\Gamma}_{\mathbf{R}}^{\mathbf{N}}+\mathbf{R}^*$. Making use of Lemma 4.3.5 which shows that $\mathbf{\Psi}\mathbf{q}^{\mathbf{N}}\leq(\mathbf{B}+2\mathbf{R}^*)\mathbf{q}^*$, it follows that

$$
\begin{aligned}
&W(\mathbf{q}^N)\\
\leq&(\mathbf{q}^*)^T(\mathbf{B}+2\mathbf{R}^*)(\mathbf{B}+\mathbf{\Gamma}_{\mathbf{B}}+\mathbf{\Gamma}_{\mathbf{R}}^{\mathbf{N}}+\mathbf{R}^{\mathbf{N}})^{-1}(\frac{1}{2}\mathbf{B}+\mathbf{\Gamma}_{\mathbf{B}}+\mathbf{\Gamma}_{\mathbf{R}}^{\mathbf{N}})(\mathbf{B}+\mathbf{\Gamma}_{\mathbf{B}}+\mathbf{\Gamma}_{\mathbf{R}}^{\mathbf{N}}+\mathbf{R}^{\mathbf{N}})^{-1}\\
&(\mathbf{B}+2\mathbf{R}^*)\mathbf{q}^*\\
\leq&\kappa(\mathbf{q}^*)^T(\mathbf{B}+2\mathbf{R}^*)(\mathbf{B}+\mathbf{\Gamma}_{\mathbf{B}}+\mathbf{\Gamma}_{\mathbf{R}}^*+\mathbf{R}^*)^{-1}(\frac{1}{2}\mathbf{B}+\mathbf{\Gamma}_{\mathbf{B}}+\mathbf{\Gamma}_{\mathbf{R}}^*)(\mathbf{B}+\mathbf{\Gamma}_{\mathbf{B}}+\mathbf{\Gamma}_{\mathbf{R}}^*+\mathbf{R}^*)^{-1}\\
&(\mathbf{B}+2\mathbf{R}^*)\mathbf{q}^*,
\end{aligned}
$$

where we obtain the last inequality by using the Jacobian similarity property that we discussed for matrix $\mathbf{F}(\mathbf{q}) = \mathbf{\Gamma}_{\mathbf{R}}^{\mathbf{N}}$ and $\kappa \geq 1$. In particular, an upper bound on κ is given as follows, based on the definition of the maximum eigenvalue of a positive semidefinite matrix,

$$\kappa \leq \lambda_{\max}\{(\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma}_{\mathbf{R}}^{\mathbf{N}} + \mathbf{R}^{\mathbf{N}})^{-2}(\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma}_{\mathbf{R}}^{*} + \mathbf{R}^{*})^{2}(\frac{1}{2}\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma}_{\mathbf{R}}^{\mathbf{N}}) (\frac{1}{2}\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma}_{\mathbf{R}}^{*})^{-1}\}. \quad \text{(B.6)}$$

Combine this result with $W(\mathbf{q}^*)$ shown in Lemma 4.3.1, we obtain an upper bound

$$\frac{W(\mathbf{q}^{\mathbf{N}})}{W(\mathbf{q}^*)} \leq \kappa \frac{(\mathbf{q}^*)^T(\mathbf{B} + 2\mathbf{R}^*)(\frac{1}{2}\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma}_{\mathbf{R}}^{*})(\mathbf{B} + 2\mathbf{R}^*)\mathbf{q}^*}{(\mathbf{q}^*)^T(\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma}_{\mathbf{R}}^{*} + \mathbf{R}^*)(\frac{1}{2}\mathbf{B} + \mathbf{R}^*)(\mathbf{B} + \mathbf{\Gamma_B} + \mathbf{\Gamma}_{\mathbf{R}}^{*} + \mathbf{R}^*)\mathbf{q}^*}.$$

As all quantities are in the social optimum setting, we will skip the superscript on matrices. Denote $\mathbf{G} = \mathbf{\Gamma}_{\mathbf{B}}^{-0.5}(\mathbf{B} + 2\mathbf{\Gamma_R})\mathbf{\Gamma}_{\mathbf{B}}^{-0.5}$ and $\mathbf{\Xi} = \mathbf{\Gamma}_{\mathbf{B}}^{-0.5}\mathbf{R}_{\mathbf{off}}\mathbf{\Gamma}_{\mathbf{B}}^{-0.5}$, the expression becomes

$$\frac{W(\mathbf{q}^{\mathbf{N}})}{W(\mathbf{q}^*)} \leq \kappa \frac{(\mathbf{q}^*)^T\mathbf{\Gamma}_{\mathbf{B}}^{0.5}(\mathbf{G} + 2\mathbf{I})\mathbf{\Gamma}_{\mathbf{B}}^{0.5}\mathbf{q}^*}{(\mathbf{q}^*)^T\mathbf{\Gamma}_{\mathbf{B}}^{0.5}(\mathbf{G} + \mathbf{I} + \mathbf{\Xi})(\mathbf{G} + 2\mathbf{\Xi})^{-1}(\mathbf{G} + \mathbf{I} + \mathbf{\Xi})\mathbf{\Gamma}_{\mathbf{B}}^{0.5}\mathbf{q}^*}.$$

Using the Rayleigh-Ritz Theorem, the upper bound can be simplified as the follows. We skip some derivations as they follow exactly the same steps in the proof for Theorem 4.3.4.

$$\begin{aligned}\frac{W(\mathbf{q}^{\mathbf{N}})}{W(\mathbf{q}^*)} &\leq \kappa\lambda_{\max}\{(\mathbf{G} + 2\mathbf{I})(\mathbf{G} + \mathbf{I} + \mathbf{\Xi})^{-1}(\mathbf{G} + 2\mathbf{\Xi})(\mathbf{G} + \mathbf{I} + \mathbf{\Xi})^{-1}\} \\ &\leq \kappa(1 - \lambda_{\min}\{((\mathbf{I} - \mathbf{\Xi})(\mathbf{G} + \mathbf{I} + \mathbf{\Xi})^{-1}))^2\}).\end{aligned}$$

□

B.7 Proof for Proposition 4.5.1

B.7.1 Preliminary result: two indices for the post-merged (multiple-product) setting

In the pre-merger setting where each service provider only provides a single service, the competition index adjusted with internalized congestion effect can be written as follows, $\mathbf{r}^{pre} = (r_1^{pre}, \ldots, r_n^{pre}) = |\mathbf{\Gamma_B}^{-1}(\mathbf{B_{off}} + 2\mathbf{\Gamma_R})|_1$, where $|.|_1$ is the L_1 norm (the row sum of absolute values of the elements in a matrix). The external-cost-to-benefit ratio can also be written in a compact form, i.e., $\rho^{pre} = (\rho_1^{pre}, \ldots, \rho_n^{pre}) = |\mathbf{\Gamma_B}^{-1}\mathbf{R_{off}}|_1$.

When the firms merge, the merged firm needs to make decisions on the service level for all m types of services. Matrix $\mathbf{B_M} = \begin{bmatrix} \beta_{11} & \ldots & \beta_{1m} \\ \vdots & \ddots & \vdots \\ \beta_{m1} & \ldots & \beta_{mm} \end{bmatrix} \in \mathbb{R}^{m \times m}$ represents the merged firm's price change of all m products with respect to a unit change in the quantity of these products. Let $\mathbf{\Gamma_B}^{post}$ assembles each firm's own quantity sensitivity coefficients into a block diagonal matrix and $\mathbf{B_{off}}^{post} = \mathbf{B} - \mathbf{\Gamma_B}^{post}$ represents each firms's price change with respect to his competitors' output change. Similarly, as the merged firm determines his service level, he also takes into the congestion effect experienced by all m types of services. We define $\mathbf{\Gamma_R}^{post}$ and $\mathbf{R_{off}}^{post}$ in a similar way.

$$\mathbf{\Gamma_B}^{post} = \begin{bmatrix} \mathbf{B_M} & & & \\ & \beta_{m+1,m+1} & & \\ & & \ddots & \\ & & & \beta_{nn} \end{bmatrix}, \mathbf{\Gamma_R}^{post} = \begin{bmatrix} \mathbf{R_M} & & & \\ & l_{m+1,m+1} & & \\ & & \ddots & \\ & & & l_{nn} \end{bmatrix}, \text{ where}$$

$$\mathbf{R_M} = \begin{bmatrix} l_{11} & \ldots & l_{1m} \\ \vdots & \ddots & \vdots \\ l_{m1} & \ldots & l_{mm} \end{bmatrix}.$$

Thus, the competition index for each service in the post-merger setting can be written as $\mathbf{r}^{post} = (r_1^{post}, \ldots, r_n^{post}) = \left|(\mathbf{\Gamma_B}^{post})^{-1}(\mathbf{B_{off}}^{post} + 2\mathbf{\Gamma_R}^{post})\right|_1$. Similarly, the post-merger external-cost-to-benefit ratio is defined as $\Theta^{post} = (\theta_1^{post}, \ldots, \rho_n^{post}) = |(\mathbf{\Gamma_B}^{post})^{-1}\mathbf{R_{off}}^{post}|_1$.

B.7.2 Proof for Proposition 4.5.1

We use one property of block diagonal matrices to prove this result: An inverse of a block diagonal matrix is still a block diagonal matrix. In particular, $(\mathbf{\Gamma_B}^{post})^{-1} = \begin{bmatrix} \mathbf{B_M}^{-1} & & & \\ & 1/\beta_{m+1,m+1} & & \\ & & \ddots & \\ & & & 1/\beta_{nn} \end{bmatrix}$, where $\mathbf{B_M}^{-1} = \frac{1}{\beta-1}(\mathbf{I} - \frac{1}{\beta-1+\mathbf{m}}H) \in \mathbb{R}^{m\times m}$, where $\mathbf{I}$ is an identity matrix and $\mathbf{H}$ is a matrix of all 1s. Given the definitions for $\mathbf{r}^{post}$ and ρ^{post}, it is straightforward to show the desired results. $\square$

B.8 Proof for Proposition 4.5.2

Firstly, we observe that for a market with n types of services, the optimal total welfare, W^*, remains the same before or after the merger as W^* is the outcome when a central planner jointly determines the service level for all types of services, irrespective of who owns them. With symmetric service providers and self-contained congestion effect, the lower bound on efficiency loss in Theorem 4.3.7 is tight, i.e., $1 - W^{pre}/W^* = 1/(2 + r^{pre})^2$, where $r^{pre} = (n - 1 + 2l)/\beta$. After the merger, the service providers are no longer "symmetric" due to the differences in their sizes. Thus, Theorem 4.3.7 serves as a lower bound to the quantity $1 - W^{post}/W^*$. By Proposition 4.5.1, $r_i^{post} \leq r_i^{pre}$ for all i. Thus, $1 - W^{post}/W^* > 1/(2 + r^{pre})^2$, which leads to the conclusion that $W^{post} < W^{pre}$. $\square$

B.9 Proof for Proposition 4.5.3

The bound on $1 - W^N/W^* \leq (\frac{\rho-1}{\rho+1})^2$ is tight for symmetric service providers as shown in Theorem ?? when $\rho > 1$. After the merger, ρ^{post} decreases, indicating a welfare improvement. When $\rho \geq 1$, the bound depends on both ρ and r. For a given ρ^{post}, one could construct examples by varying r^{post} to have W^{post} which could be higher or lower than the pre-merger value. $\square$

B.10 Proof for Proposition 4.6.1 and Proposition 4.6.2

With the utilization rate ϕ, the societal welfare $W(\mathbf{q}, \phi)$ is given by

$$\begin{aligned} W(\mathbf{q}, \phi) &= CS(\mathbf{q}, \phi) + PS(\mathbf{q}, \phi) + \phi TR(\mathbf{q}, \phi) \\ &= \mathbf{q}^T(\bar{\mathbf{p}} - \bar{\mathbf{l}} - \mathbf{B}/2\mathbf{q} - \mathbf{R}\mathbf{q}) + (\phi - 1)\mathbf{q}^\mathbf{T}\mathbf{t} \\ &= \mathbf{q}^T(\bar{\mathbf{p}} - \bar{\mathbf{l}} - \mathbf{B}/2\mathbf{q} - \mathbf{R}\mathbf{q}) + (\phi - 1)\mathbf{q}^\mathbf{T}(\bar{\mathbf{p}} - \bar{\mathbf{l}} - (\mathbf{B} + \mathbf{R} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R})\mathbf{q}) \end{aligned}$$

We can then show that the optimal service level, $\mathbf{q}^*(\phi)$ is given by

$$\mathbf{q}^*(\phi) = ((2 - 1/\phi)\mathbf{B} + \mathbf{2R} + 2(1 - 1/\phi)(\mathbf{\Gamma_B} + \mathbf{\Gamma_R}))^{-1}(\bar{\mathbf{p}} - \bar{\mathbf{l}}),$$

which decreases in ϕ. Since the optimal congestion pricing is given by $\mathbf{t}^* = \bar{\mathbf{p}} - \bar{\mathbf{l}} - (\mathbf{B} + \mathbf{R} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R})\mathbf{q}^*$, it follows that $\mathbf{t}^*$ increases with ϕ. Moreover, the optimal societal welfare is given by

$$W^*(\phi) = W(\mathbf{q}^*, \phi) = 1/2(\bar{\mathbf{p}} - \bar{\mathbf{l}})^T((2 - 1/\phi)\mathbf{B} + \mathbf{2R} + 2(1 - 1/\phi)(\mathbf{\Gamma_B} + \mathbf{\Gamma_R}))^{-1}(\bar{\mathbf{p}} - \bar{\mathbf{l}}),$$

which also increases in ϕ. $\square$

B.11 Proof for Proposition 4.6.3

When we redistribute the revenue collected from congestion pricing as suggested, producer surplus for service provider i becomes $\overline{PS_i} = PS_i^* + \alpha_i TR^* = PS_i^* + \frac{\psi PS_i^N - PS_i^*}{TR^*} TR^* = \psi PS_i^N$. Similarly, we can show that $\overline{CS_i} = \psi CS_i^N$. Since $\psi =$

$W^*/W^N \geq 1$, both the service providers and the users are better off than their counterparts in the unregulated setting after the revenue is redistributed. $\square$

B.12 Proof for Proposition 4.6.4

Consider a new marginal utility function $\tilde{u}_i(\mathbf{q}) = \theta_i - \sum_{j=1}^{n} \beta_{ij} q_j$, where θ_i represents a representative user's maximum willingness to pay for service i. Under this demand function, producer surplus of service provider i in the unregulated setting is given by $PS_i = q_i(\theta_i - \bar{l}_i - \sum_j(\beta_{ij} + l_{ij})q_j)$. The goal is to have $PS_i = \overline{PS_i}$, i.e., the value which service provider i receives in an *unregulated* setting equals to target surplus level. In the unregulated setting, the equilibrium service level with the new demand function satisfies $\tilde{\mathbf{q}}^{\mathbf{N}} = (\mathbf{B} + \mathbf{R} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R})^{-1}(\mathbf{\Theta} - \bar{\mathbf{l}})$. There exists a θ_i which satisfies $\overline{PS_i} = \tilde{q}_i^N(\theta_i - \bar{l}_i - \sum_j(\beta_{ij} + l_{ij}))\tilde{q}_j^N)$, or equivalently, $\overline{PS_i} = (\theta_i - \bar{l}_i)\sum_j M_{ij}(\theta_j - \bar{l}_j)$, where $\mathbf{M} = (\mathbf{B} + \mathbf{R} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R})^{-1}(\mathbf{\Gamma_B} + \mathbf{\Gamma_R})(\mathbf{B} + \mathbf{R} + \mathbf{\Gamma_B} + \mathbf{\Gamma_R})^{-1}$. $\square$

Appendix C

Appendix for Chapter 5

C.1 Nash Bargaining Game

A n-person Nash Bargaining game consists of a pair $(\mathcal{N}, \omega)$, where $\mathcal{N} \subseteq \mathbb{R}^n_+$ is a compact and convex set and $\omega \in \mathcal{N}$. Set $\mathcal{N}$ is the feasible set and its elements give utilities that the n players can simultaneously accrue. Point ω is the disagreement point - it gives the utilities that the n players obtain if they decide not to cooperate. Game $(\mathcal{N}, \omega)$ is said to be feasible if there is a point $v \in \mathcal{N}$ such that $v_1 > \omega_1$ and $v_2 > \omega_2$. The solution to a feasible game is the point that satisfies the following four axioms,

1. Pareto optimality: No point in $\mathcal{N}$ can weakly dominate v.
2. Invariance under affine transformation of utilities
3. Symmetry: The numbering of the players should not affect the solution.
4. Independence of irrelevant alternatives: If v is the solution for $(\mathcal{N}, \omega)$, and $\mathcal{S} \subseteq \mathbb{R}^n_+$ is a compact and convex set satisfying $\omega \in \mathcal{S}$ and $v \in \mathcal{S} \subseteq \mathcal{N}$, then v is also the solution for $(\mathcal{S}, \omega)$.

Nash Bargaining Solution (NBS) If game $(\mathcal{N}, \omega)$ is feasible then there is a unique point in $\mathcal{N}$ satisfying the axioms stated above. This is also the unique point that maximizes $\prod_{i=1}^{n}(v_i - \omega_i)$ over all $v \in \mathcal{N}$.

Bibliography

Acemoglu, D., A. Ozdaglar. 2007. Competition and efficiency in congested markets. *Mathematics of Operations Research* **32**(1) 1–31.

Adida, E., V. DeMiguel. 2010. Supply chain competition with multiple manufacturers and retailers. *Forthcoming in Operations Research* .

Allon, G., A. Federgruen. 2007. Competition in service industries. *Operations Research* **55** 37–55.

Allon, G., A. Federgruen. 2008. Service competition with general queueing facilities. *Operations Research* **56** 827–849.

Angelus, A., E. L. Porteus. 2002. Simultaneous capacity and production management of short-life-cycle, produce-to-stock goods under stochastic demand. *Management Science* **48**(3) pp. 399–413.

Arora, A., A. Gambardella. 1990. Complementarity and external linkages: The strategies of the large firms in biotechnology. *The Journal of Industrial Economics* **38**(4) 361–379.

Bassok, Y., R. Anupindi, R. Akella. 1999. Single-period multiproduct inventory models with substitution. *Operations Research* **47**(4) pp. 632–642.

Bernstein, F., A. Federgruen. 2003. Pricing and replenishment strategies in a distribution system with competing retailers. *Operations Research* **51**(3) 409–426.

Bernstein, F., A. Federgruen. 2007. Coordination mechanisms for supply chains under price and service competition. *Manufacturing and Service Operations Management* **9**(3) 242–262.

Binmore, K., A. Rubinstein, A. Wolinsky. 1986. The nash bargaining solution in economic modelling. *The RAND Journal of Economics* **17**(2) pp. 176–188.

Bordley, R. 1985. Relating elasticities to changes in demand. *Journal of Business and Economic Statistics* **3**(2) 156–158.

Bradley, J. R., P. W. Glynn. 2002. Managing capacity and inventory jointly in manufacturing systems. *Management Science* **48**(2) pp. 273–288.

Brander, J. A., A Zhang. 1990. Market conduct in the airline industry: An empirical investigation. *The RAND Journal of Economics* **21**(4) pp. 567–583.

Brander, J. A., A Zhang. 1993. Dynamic oligopoly behaviour in the airline industry. *International Journal of Industrial Organization* **11**(3) 407 – 435.

Brueckner, J. K. 2002. Airport congestion when carriers have market power. *American Economic Review* **92**(5) 1357–1375.

Brueckner, J. K. 2005. Internalization of airport congestion: A network analysis. *International Journal of Industrial Organization* **23**(7-8) 599 – 614.

Brueckner, J. K., P. T. Spiller. 1994. Economies of traffic density in the deregulated airline industry. *Journal of Law and Economics* **37**(2) pp. 379–415.

Brueckner, Jan K. 2003. The benefits of codesharing and antitrust immunity for international passengers, with an application to the star alliance. *Journal of Air Transport Management* **9**(2) 83 – 89.

Bulow, J. I., J. D. Geanakoplos, P. D. Klemperer. 1985. Multimarket oligopoly: Strategic substitutes and complements. *The Journal of Political Economy* **93**(3) 488–511.

Cachon, G. P. 2003. *Handbook in Operations Research and Management Science: Supply Chain Management*, chap. Supply chain coordination with contracts. Elsevier, Amsterdam, 229–339.

Cachon, G. P., P. T. Harker. 2002. Competition and outsourcing with scale economies. *Management Science* **48**(10) 1314 – 1333.

Cachon, G. P., A. G. Kok. 2010. Competing manufacturers in a retail supply chain: On contractual form and coordination. *Management Science* **56**(3) 571–589.

Cachon, G. P., M. A. Lariviere. 1999. Capacity choice and allocation: Strategic behavior and supply chain performance. *Management Science* **45**(8) pp. 1091–1108.

Cachon, G. P., M. A. Lariviere. 2001. Contracting to assure supply: How to share demand forecasts in a supply chain. *Management Science* **47**(5) 629–646.

Cachon, G. P., M. A. Lariviere. 2005. Supply chain coordination with revenue-sharing contracts: Strengths and limitations. *Management Science* **51**(1) 30–44.

Caldentey, R., L. M. Wein. 2003. Analysis of a decentralized production-inventory system. *Manufacturing and Service Operations Management* **5**(1) 1.

Carr, S., W. Lovejoy. 2000. The inverse newsvendor problem: Choosing an optimal demand portfolio for capacitated resources. *Management Science* **46**(7) pp. 912–927.

Carr, S. M., U. S. Karmarkar. 2005. Competition in multiechelon assembly supply chains. *Management Science* **51**(1) 45–59.

Caves, D. W., L. R. Christensen, M. W. Tretheway. 1984. Economies of density versus economies of scale: Why trunk and local service airline costs differ. *The RAND Journal of Economics* **15**(4) pp. 471–489.

Chintagunta, P. K., A. Bonfrer, I. Song. 2002. Investigating the effects of store-brand introduction on retailer demand and pricing behavior. *Management Science* **48**(10) 1242–1267.

Chod, J., N. Rudi. 2006. Strategic investments, trading, and pricing under forecast updating. *Management Science* **52**(12) pp. 1913–1929.

Choi, S. C. 1991. Price competition in a channel structure with a common retailerwinter. *Marketing Science* **10**(4) 271–296.

Corbett, C. J., G. A. DeCroix, A. Y. Ha. 2005. Optimal shared-savings contracts in supply chains: Linear contracts and double moral hazard. *European Journal of Operational Research* **163**(3) 653 – 667.

Corbett, C. J., U. S. Karmarkar. 2001. Competition and structure in serial supply chains with deterministic demand. *Management Science* **47**(7) 966–978.

Correa, J. R., A. S. Schulz, N. E. Stier-Moses. 2004. Selfish routing in capacitated networks. *Mathematics of Operations Research* **29**(4) 961–976.

Correa, J. R., A. S. Schulz, N. E. Stier-Moses. 2007. Fast, fair, and efficient flows in networks. *Operations Research* **55**(2) 215–225.

Dafermos, S. C., F. T. Sparrow. 1969. The traffic assignment problem for a general network. *Journal of research of the National Bureau of Standards* **73**(2) 91 – 118.

Daniel, Joseph I. 1995. Congestion pricing and capacity of large hub airports: A bottleneck model with stochastic queues. *Econometrica* **63**(2) 327–370.

Daniel, Joseph I. 2001. Distributional consequences of airport congestion pricing. *Journal of Urban Economics* **50**(2) 230 – 258.

Deneckere, R., C. Davidson. 1985. Incentives to form coalitions with bertrand competition. *RAND Journal of Economics* **16**(4) 473 – 486.

Ellison, S. F., I. Cockburn, Z. Griliches, J. Hausman. 1997. Characteristics of demand for pharmaceutical products: An examination of four cephalosporins. *The RAND Journal of Economics* **28**(3) 426–446.

Farahat, A., G. Perakis. 2009. Profit loss in differentiated oligopolies. *Operations Research Letters* **37**(1) 43 – 46.

Farahat, A., G. Perakis. 2010a. The anarchy of price competition. *Working paper*.

Farahat, A., G. Perakis. 2010b. A nonnegative extension of the affine demand function and equilibrium analysis for multiproduct price competition. *Operations Research Letters* **38**(4) 280–286.

Farrell, J., C. Shapiro. 1990. Horizontal mergers: An equilibrium analysis. *American Economic Review* **1**(80) 107–126.

Faulhaber, G. R., C. Hogendorn. 2000. The market structure of broadband telecommunications. *The Journal of Industrial Economics* **48**(3) 305–329.

FCC Technical Paper. 2010. Mobile broadband: The benefits of additional spectrum. `http://hraunfoss.fcc.gov/edocs-public/attachmatch/DOC-302324A1.doc`.

Federgruen, A., P. Zipkin. 1986. An inventory model with limited production capacity and uncertain demands ii. the discounted-cost criterion. *Mathematics of Operations Research* **11**(2) pp. 208–215.

Fudenberg, D., J. Tirole. 1984. The fat-cat effect, the puppy-dog ploy, and the lean and hungry look. *American Economic Review* **74**(2) 361–366.

Goodwin, P. B. 1990. How to make road pricing popular. *Economic Affairs* **10**(5) 6–7.

Goyal, M., S. Netessine. 2007. Strategic technology choice and capacity investment under demand uncertainty. *Management Science* **53**(2) 192 – 207.

Hamdouch, Y., M. Florian, D. W. Hearn, S. Lawphongpanich. 2007. Toll pricing in multimodal traffic assignment. *Transportation Research Part B* **41**(3) 275–291.

Hardin, G. 1968. Tragedy of the commons. *Science* **162**(3859) 1243–1248.

Hayrapetyan, A., E. Tardos, T. Wexler. 2007. A network pricing game for selfish traffic. *Distributed Computing* **19** 255–266.

Heyer, K. 2006. Welfare standards and merger analysis: Why not the best? Economic Analysis Group Discussions Paper, Department of Justice, Antitrust Division.

Holmes, T. J. 1989. The effects of third-degree price discrimination in oligopoly. *The American Economic Review* **79**(1) pp. 244–250.

Horn, R., C. Johnson. 1985. *Matrix Analysis*. Cambridge University Press.

Hu, X., R. Caldentey, G Vulcano. 2012. Revenue sharing in airline alliances. *Forthcoming in Management Science* .

Jain, H. 2010. Alliance revenue management in practice : techniques and simulation analysis. Master's thesis, Massachusetts Institute of Technology, USA.

Johari, G. Y., R. Weintraub, B. Van Roy. 2009. Investment and market structure in industries with congestion. *Forthcoming in Operations Research* .

Jones, P. 1991. Gaining public support for road pricing through a package approach. *Traffic engineering and control* **32**(4) 194–196.

Kapuscinski, R., S. Tayur. 1998. A capacitated production-inventory model with periodic demand. *Operations Research* **46**(6) pp. 899–911.

Keeler, J P., J P. Formby. 1994. Cost economies and consolidation in the us airline industry. *International Journal of Transport Economics* **21**(1) pp. 21– 45.

Koutsoupias, E., C. Papadimitriou. 1999. Worst-case equilibria. *Proceedings of 16th Sympos. Theoretical Aspects of Computer Science*. 404–413.

Kouvelis, P., M. A. Lariviere. 2000. Decentralizing cross-functional decisions: Coordination through internal markets. *Management Science* **46**(8) pp. 1049–1058.

Kreps, David M., Jose A. Scheinkman. 1983. Quantity precommitment and bertrand competition yield cournot outcomes. *The Bell Journal of Economics* **14**(2) 326–337.

Krishnan, H., R. A. Winter. 2007. Vertical control of price and inventory. *American Economic Review* **97**(5) 1840–1857.

Lin, X., R. Germain. 1998. Sustaining satisfactory joint venture relationships: The role of conflict resolution strategy. *Journal of International Business Studies* **29**(1) pp. 179–196.

Mahajan, S., G. Van Ryzin. 2001. Inventory competition under dynamic consumer choice. *Operations Research* **49**(5) 646–657.

Maille, P., N. E. Stier-Moses. 2009. Eliciting coordination with rebates. *Transportation Science* **43**(4) 473–492.

Malueg, D. A. 1994. Monopoly output and welfare: The role of curvature of the demand function. *The Journal of Economic Education* **25**(3) pp. 235–250.

Martinez-de Alberniz, V., G. Roels. 2010. Competing for shelf space. *Forthcoming in Production and Operation Management* .

Martinez-de Alberniz, V., D. Simchi-Levi. 2009. Competition in the supply option market. *Operation Research* **57**(5) 1082–1097.

Mayer, C., T. Sinai. 2003. Network effects, congestion externalities, and air traffic delays: Or why all delays are not evil. *American Economic Review* **93**(4) 1194–1215.

McGuire, T. W., R. Staelin. 1983. An industry equilibrium analysis of downstream vertical integration. *Marketing Science* **2**(2) 161–191.

Mela, C. F., S. Gupta, K. Jedidi. 1998. Assessing long-term promotional influences on market structure. *International Journal of Research in Marketing* **15**(2) 89 – 107.

Milgrom, P., J. Roberts. 1995. Complementarities and fit strategy, structure, and organizational change in manufacturing. *Journal of Accounting and Economics* **19**(2-3) 179 – 208.

Miller, K. D. 1992. A framework for integrated risk management in international business. *Journal of International Business Studies* **23**(2) pp. 311–331.

Moorthy, K. S. 1987. Managing channel profits: Comment. *Marketing Science* **6**(4) 375–379.

Myerson, R. B. 1979. Incentive compatibility and the bargaining problem. *Econometrica* **47**(1) pp. 61–73.

Nagarajan, M., G. Sosic. 2008. Game-theoretic analysis of cooperation among supply chain agents: Review and extensions. *European Journal of Operational Research* **187**(3) 719 – 745.

Nash, J. F. 1950. The bargaining problem. *Econometrica* **18**(2) pp. 155–162.

Netessine, S., F. Zhang. 2005. Positive vs. negative externalities in inventory management: Implications for supply chain design. *Manufacturing and Service Operations Management* **7**(1) 58–73.

Netessine, Serguei, Nils Rudi. 2003. Centralized and competitive inventory models with demand substitution. *Oper. Res.* **51**(2) 329–335.

Neven, D. J., L. H Roller. 2005. Consumer surplus vs. welfare standard in a political economy model of merger control. *International Journal of Industrial Organization* **23**(9-10) 829 – 848.

Oum, Tae Hoon, Anming Zhang, Yimin Zhang. 1993. Inter-firm rivalry and firm-specific price elasticities in deregulated airline markets. *Journal of Transport Economics and Policy* **27**(2) pp. 171–192.

Ozdaglar, A. 2008. Price competition with elastic traffic. *Networks* **52** 141–155.

Papadimitriou, C. 2001. Algorithms, games, and the internet. *STOC '01: Proceedings of the thirty-third annual ACM symposium on Theory of computing*. ACM, New York, NY, USA, 749–753.

Parker, Philip M., Lars-Hendrik Roller. 1997. Collusive conduct in duopolies: Multimarket contact and cross-ownership in the mobile telephone industry. *The RAND Journal of Economics* **28**(2) pp. 304–322.

Pasternack, B. A. 1985. Optimal pricing and return policies for perishable commodities. *Marketing Science* **4**(2) 166–176.

Perakis, G. 2007. The "price of anarchy" under nonlinear and asymmetric costs. *Mathematics of Operations Research* **32** 614–628.

Perakis, G., G. Roels. 2007. The price of anarchy in supply chains: Quantifying the efficiency of price-only contracts. *Management Science* **53**(8) 1249 – 1268.

Porteus, E. L., S. Whang. 1991. On manufacturing/marketing incentives. *Management Science* **37**(9) pp. 1166–1181.

Reyniers, D. J., C. S. Tapiero. 1995. The delivery and control of quality in supplier-producer contracts. *Management Science* **41**(10) pp. 1581–1589.

Robinson, J. 1933. *The Economics of Imperfect Competition*. Macmillan, London.

Roughgarden, T. 2005. *Selfish Routing and the Price of Anarchy*. MIT Press.

Roughgarden, T., E. Tardos. 2002. How bad is selfish routing? *Journal of ACM* **49** 236–259.

Rubinstein, A. 1982. Perfect equilibrium in a bargaining model. *Econometrica* **50**(1) pp. 97–109.

Schmalensee, R. 1981. Output and welfare implications of monopolistic third-degree price discrimination. *The American Economic Review* **71**(1) pp. 242–247.

Schwartz, M. 1990. Third-degree price discrimination and output: Generalizing a welfare result. *The American Economic Review* **80**(5) pp. 1259–1262.

Small, K. A. 1992. Using the revenues from congestion pricing. *Transportation* **19** 359–381.

Spengler, J. J. 1950. Vertical integration and antitrust policy. *The Journal of Political Economy* **58**(4) 347–352.

Sun, W. 2006. Price of Anarchy in a Bertrand Oligopoly. Master's thesis, Massachusetts Institute of Technology, USA.

The Joint Economic Committee Report. 2008. Your flight has been delayed again: flight delays cost passengers, airlines, and the us economy billions. `http://jec.senate.gov/public/?a=Files.Serve&File-id=47e8d8a7-661d-4e6b-ae72-0f1831dd1207`.

Tirole, J. 1988. *The Theory of Industrial Organization*. MIT Press.

Tyagi, R. K. 1999. On the effects of downstream entry. *Management Science* **45**(1) 59–73.

Van Mieghem, J. A. 2003. Capacity management,investment, and hedging:review and recent developments. *Manufacturing and Service Operations Management* **5**(4) 269 – 302.

Van Mieghem, J. A., M. Dada. 1999. Price versus production postponement: Capacity and competition. *Management Science* **45**(12) pp. 1631–1649.

Van Mieghem, J. A., N. Rudi. 2002. Newsvendor networks: Inventory management and capacity investment with discretionary activities. *Manufacturing and Service Operations Management* **4**(4) 313.

Varian, H. R. 1985. Price discrimination and social welfare. *The American Economic Review* **75**(4) pp. 870–875.

Vives, X. 1999. *Oligopoly pricing: Old ideas and new tools*. MIT Press.

Wang, Y., Y. Gerchak. 2003. Capacity games in assembly systems with uncertain demand. *Manufacturing and Service Operations Management* **5**(3) 252–267.

Wang, Y., Y. Gerchak. 2004. Revenue-sharing vs. wholesale-price contracts in assembly systems with Random Demand. *Production and Operations Management* **13**(1) 23–33.

Wardrop, J. 1952. Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers, Part II* 1(36) 352–362.

Weintraub, G. Y., R. Johari, B. Van Roy. 2010. Investment and market structure in industries with congestion. *Forthcoming in Operations Research* .

Weisman, E. 1990. *Trade in Services and Imperfect Competition: Application to International Aviation*. Kluwer Academic.

Whalen, W. T., D. W. Carlton, K. Heyer, O. M. Richard. 2008. A Solution to Airport Delays. *Regulation* **31**(1) 30–36.

Williamson, O. E. 1968. Economies as an antitrust defense: The welfare tradeoffs. *American Economic Review* **58**(1) 18.

Y., Luo, Shenkar O. 2002. An empirical inquiry of negotiation effects in cross-cultural joint ventures. *Journal of International Management* **8**(2).

Yoshida, Y. 2000. Third-degree price discrimination in input markets: Output and welfare. *The American Economic Review* **90**(1) pp. 240–246.