HOKKAIDO UNIVERSITY

Title	Essays on Determinants and Measurements of Productivity
Author(s)	黄, 霂寒
Citation	北海道大学. 博士(経済学) 甲第12525号
Issue Date	2017-03-23
DOI	10.14943/doctoral.k12525
Doc URL	http://hdl.handle.net/2115/66267
Type	theses (doctoral)
File Information	Muhan_HUANG.pdf

Instructions for use

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

Essays on Determinants and Measurements of Productivity

Muhan Huang

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Graduate School of Economics and Business Administration

Hokkaido University

Sapporo, Japan

2017

Acknowledgment

First and foremost, I would like to express my deepest gratitude to my advisor, Professor Atsushi Ohyama, who mentored me throughout my doctoral studies. This dissertation would not have been possible without his constant guidance.

Besides my advisor, I would like to thank the members of my dissertation committee, Professor Kazuo Machino at Hokkaido University, Professor Yoichi Hizen at Kochi University of Technology, for their valuable comments and suggestions.

I also wish to express my sincere appreciation to the faculty members at Hokkaido University, especially Professor Junichi Itaya and Professor Yukiko Abe, for their constructive feedback on my research; Professor Noritaka Kudoh and Professor Yuka Ohno, for their continuing support. I would also like to thank the staffs at Hokkaido University, especially Kumiko Tsukada, for her help during my graduate school.

My heartfelt thanks also go to my colleagues in graduate school, especially Kengo Kurosaka, Nobushige Kanamori, Shingo Yamazaki, Lei Wang, Xuelong Wang, and Wenjun Chen, for their friendship and assistance.

Finally, I would like to extend my deepest thanks to my parents, Ming Huang and Yulin Sha, for their understanding, endless patience and encouragement.

Abstract

In this dissertation, I theoretically and empirically examine effects of new entry, resource misallocation and economic policies on firm-level and region-level productivity, and I try to measure real technical improvement more precisely when observed productivity changes arise from both supply and demand factors. More specifically, I propose a theoretical approach to examine the entry effect on productivity and to distinguish the physical productivity improvement and revenue productivity change. I also use Chinese firm-level data to examine to what extent resource misallocation accounts for observed regional productivity differences between Beijing and Shanghai. The main contributions of this dissertation are threefold. The first contribution is to improve our understanding of sources and mechanisms that determine productivity. The second one is to propose a methodology of measuring technical efficiency improvement more precisely. The third one is to provide policy implications for fostering economic efficiency at the regional level.

In Chapter 2, I examine how new entry influences incumbent firm's measured productivity when the highest quality product is introduced to the market by incorporating both demand and supply factors into a single analytical framework. First I build a theoretical model by extending the model of Johnson and Myatt (2003) where both consumers and firms take product quality into account when they decide their optimal behaviors. The extended model allows me to link physical productivity and revenue productivity under several types of new entry. The key insight form this analysis is that incumbent firm's revenue productivity can be affected by both business

stealing effects and technical improvement effects. Based on the results from this theoretical model, I discuss potential problems of using revenue productivity measures, and the importance of distinguishing physical productivity improvement and revenue productivity improvement in an empirical analysis. The separation of revenue productivity from physical productivity is critical for understanding the sources of an observed productivity change accurately.

In Chapter 3, I use firm-level Chinese manufacturing data to investigate how input market distortions affect the aggregate productivity differences between two major cities in China, Beijing and Shanghai. In this empirical analysis, I use an extended version of Hsieh and Klenow (2009) approach and an alternative approach developed from Midrigan and Xu (2014) to estimate productivity losses from resource distortions. This empirical analysis reveals that the aggregate productivity level is lower in Beijing than that in Shanghai, and the input market distortions, especially the capital misallocation is more severe in Beijing than that in Shanghai.

In Chapter 4, I attempt to offer a possible mechanism through which regional productivity are affected by labor market misallocation between Beijing and Shanghai. In doing so, I construct a theoretical model and investigate a possible role of Hukou allocation system (a unique household registration policy in China) in influencing regional productivity through firm's strategic behavior with respect to the retention of workers. The theoretical analysis shows that a firm has an incentive to retain inefficient match between the firm and workers under some Hukou system and this theoretical insight indicates that the level of regional productivity is low due to inefficient labor market allocation arising from Hukou system. This analysis offers a new explanation for an observed regional productivity difference between Beijing and Shanghai.

Contents

Chapter 1.

Introduction

What determines productivity? How can we estimate productivity precisely? These questions have been central for a long time to researchers in many fields who are interested in investigating the source and nature of economic growth and development. This is because, as endogenous growth theory indicates, productivity growth is a critical determinant of economic growth, especially under the condition that resources such as labor and capital are unlikely to expand. Thus, these productivity issues have been attracting a lot of attention from economists and policy makers, and in this dissertation I try to contribute to this important line of the literature on determinants of productivity and productivity measurements.

1.1 Motivation

Productivity is commonly defined as a ratio of volume measure of output to a volume measure of input. It is essentially a measure of efficiency that describes the ability to transform raw inputs into intermediate or final goods. High level of productivity allows firms or economies to produce a larger amount of outputs from a given level of inputs, and it thus leads to a higher GDP and economic growth. In other words, economic growth can be achieved with an improved productivity level even if inputs such as labor and capital remain unchanged. Productivity is therefore recognized as the most fundamental and crucial determinant of economic growth, and it plays an important role for explaining why some regions or economies are richer than others and why some regions or economies were poor in the past are now getting richer.

Past research regarding productivity has revealed that many factors can influence productivity. For example, Syverson (2011) lists several mechanisms through which productivity is determined and presents related evidence in his comprehensive literature survey. Among several determinants of productivity, he mentions the degree of competition in market, resources allocation among firms, product innovation, management, investment in capital, and government policy. In his survey, it is stressed that productivity can improve not only by firms' behaviors such as R&D investment and process innovations, but also by other firms' behaviors through competition and spillovers and institutional settings.

Among factors for productivity improvement, competition is one of the important determinants. First, competition affects a firm's incentive to innovate because the firm may conduct process or product innovation in order to escape from competitive pressure when facing fierce competition. Conversely, a firm may reduce its incentive to innovate because competition may result in eroding benefits from the innovation (Aghion et al., 2005). In either way, competition influences firms' incentives to innovate, and this in turn affect firms' productivity. Competition also brings about selection effect and reallocation effect. That is, fierce competition is likely to force some firms to exit from a market and reallocate scarce resources from one firm to another firm. Although a lot of studies have examined a relationship between competition and firms' productivity or a relationship between competition and firm's product choices, only a few researchers examined competition, productivity and product choices in a single framework, and as a result we have known very little about how competition affects demand and supply factors for the determination of productivity at the firm and industry level. In order to fill this gap in the literature, in this dissertation, I build a theoretical model by incorporating all the three elements (i.e., competition, productivity and product choices) and examine entry effects on productivity by taking both demand and supply factors into account.

Recently, effects of resource misallocation on productivity have increasingly attracted attentions from economic researchers, because it plays an important role in explaining productivity differences between developed countries like the U.S. and developing countries like China and India. Past studies have mainly focused on differences in productivity across countries, but there are not many studies that empirically investigate regional differences in productivity within a country arising from resource misallocation. In this dissertation, I use the Chinese manufacturing data to examine whether regional productivity differences are due to resource misallocation. Moreover, I dig deeper this resource misallocation problem by investigating how a particular policy in China leads to such resource distortion in the labor markets of Beijing and Shanghai, and whether the policy accounts for a possible source of regional productivity differences between Beijing and Shanghai.

While factors mentioned above can affect the productivity differences among firms or regions directly, Syverson (2011) also mentions that the productivity differences may be attributable to simply different approaches of productivity measurement. More specifically, two main approaches, which differ in output measures, have been used in past research, namely, physical productivity (TFPQ) and revenue productivity (TFPR). The physical productivity approach measures the output in physical units (e.g., number of product, number of customers served), whereas the revenue productivity approach is based on an amount of output in monetary units (e.g., value added, sales).

TFPR approach is widely used to measure productivity in past studies mainly because of data availability. However, this approach can be sometimes problematic when we are interested in examining the technical efficiency. Since the TFPR approach measures output in monetary value units, the measurement of TFPR not only contains physical efficiency differences, but also reflects price effects arising from the demand side (Foster et al., 2008), input market distortions (Hsieh and Klenow, 2009, 2014) and other factors. Therefore, a lower TFPR may be due to a lower demand or market distortions rather than lower technical efficiency. This TFPR approach may fail to estimate real technical efficiency change and may end up with capturing a mixture of changes arising from both supply and demand effects. Researchers in this area recognize that the conventional TFPR is insufficient to understand some fundamental economic issues like entry effect or regional development.

It is important to distinguish physical productivity (TFPQ) and revenue productivity (TFPR) when we are interesting in examining and comparing entry effects on productivity of incumbent firms (Foster et al., 2008). Since the TFPR can be an imprecise measure of physical productivity, we must understand to which degree physical productivity measure is incorrectly estimated when using the TFPR, and we must provide a method of correcting biases from the TFPR. In Chapter 2 of this dissertation, I offer a theoretical model that raises potential problems of the TFPR, and then propose an empirical approach that links TFPR to TFPQ.

The past studies show that observed TFPR dispersion partly reflects market distortions (Hsieh and Klenow, 2009, 2014). They show that even if TFPQ dispersed, TFPR will not disperse in an economy without market distortions. This suggests the importance of distinguishing TFPR and TFPQ when we compare productivity or economic growth across regions or countries. In Chapter 3 of this dissertation, I use Chinese manufacturing data and shows differences in dispersion of TFPR and TFPQ clearly.

1.2 Productivity Changes and Entry

New product offering has profound impacts on existing product through both supply and demand sides. On the demand side, such entry allows consumers to choose from an extended variety of products, and leads to demand switch from existing products. On the supply side, such entry forces incumbent firms to face a tougher competition, and in order to respond to it, they may improve their productivity through selection and incentive.

I build a theoretical model that includes demand and supply factors, and competition in order to examine how new entry by offering a higher quality product affects existing firm's measured productivity. Chapter 2 of this dissertation is devoted to achieving this goal. I also decompose such entry effects on supply and revenue into technical improvement effect and other demand factors like business stealing effect, and separate physical productivity improvement from revenue productivity change cleanly.

More specifically, in so doing, I extend the model of Johnson and Myatt (2003), where both consumers and firms take a quality of product into consideration when they decide their optimal behaviors. In the framework of this study, consumers choose a particular quality of the product that gives them their highest utility level, whereas firms decide the quality and quantity of a particular quality of the product to supply. The key idea is that a derived first order condition from the model contains fundamental parameters of both firms and consumers, and reflects effects from both demand and supply factor. I then use the optimal conditions to separate supply factors from demand factors, which finally allows me to separate TFPQ from TFPR clearly, and make a connection between entry effects on firm revenue productivity and those on firm physical productivity.

The main insight from the theoretical analysis can be summarized as follows. When a new entry occurs by introducing the highest quality of the product, existing firms' supply and revenues are affected by both business stealing effects and physical productivity improvement effects. While the business stealing effects can be critical for the revenue of the highest quality product before such entry, the revenue of other lower quality products is unaffected. On the other hand, physical productivity improvement effects can impact the revenue of all the existing products. Since firm revenues reflect these two effects, the TFPR measure is different from TFPQ unless business-stealing effects can be controlled for properly. The model shows that the technical efficiency improvement can be underestimated if we use revenue productivity measurement in general.

1.3 Regional Productivity Difference and Resource Allocation

Empirical research on input market distortions has started using firm-level micro-data to advance our understanding of sources and mechanism of productivity improvement and economic growth (e.g., Hsieh and Klenow 2009). One important finding from this line of research is that developing countries like China and India indicate a lower level of the aggregate productivity than countries like the U.S. because of greater resource misallocation among firms. The resource misallocation literature has mainly focused on differences in productivity across countries, but there are not many studies that empirically investigate regional differences in productivity within a country arising from such efficiency of resource allocation.

In Chapter 3 of this dissertation, I estimate productivity losses from labor and capital resource misallocations as well as potential gains from efficient resource allocation, and closely examine sources of productivity differences in two Chinese major cities, Beijing and Shanghai. The estimation method I used to estimate productivity losses is developed by Hsieh and Klenow (2009) and extended by Gong and Hu (2016) who relax the assumption of constant returns to scale. In order to take a comprehensive approach, I also use an alternative approach for measuring resource misallocation based on the model in Midrigan and Xu (2014). I compare the empirical results from the Hsieh and Klenow approach and this alternative approach to examine whether these results are consistent between the two different approaches. Furthermore, I also decompose productivity loss into the one arising from capital misallocation and the one from labor misallocation.

Main findings from this study are summarized as follows. First, there is a difference in productivity loss between Beijing and Shanghai. The average aggregate productivity loss of Beijing from resource misallocation is larger than that of Shanghai, implying that resource allocation is more efficient in Shanghai than in Beijing. There is indeed a regional productivity difference due to resource misallocation. Second, the aggregate productivity level of Beijing would improve more significantly than that of Shanghai if resources were allocated to the efficient level. This result again indicates that there is some variation in the degree of resource misallocation across regions. Finally, the empirical analysis of this dissertation reveals that capital misallocation is more severe in Beijing than labor misallocation, and this indicates that some frictions likely exist in Beijing's capital markets.

The empirical results from Chapter 3 raise a question of what is a possible mechanism that generates such productivity differences between Beijing and Shanghai. In Chapter 4 of this dissertation, I try to offer a possible mechanism that may cause some labor misallocation between Beijing and Shanghai. I theoretically investigate a possible role of Hukou allocation system (a unique household registration policy in China) in influencing regional productivity through labor market distortion. I demonstrate that the level of regional productivity can be low due to inefficient labor allocation arising from firms' strategic behavior in their response to the Hukou policy.

1.4 Contributions of the Dissertation

In this dissertation, I examine the effects of new entry, resources misallocation and a particular policy on productivity theoretically and empirically. These determinants will help us to understand the sources and mechanisms of productivity improvement more accurately, and then contribute to the line of research on what determines productivity. Second, I propose an approach to examine the entry effects on productivity by controlling for both supply and demand factors, and this framework allows me to separate the entry effect on TFPR from the entry effect on TFPQ. By investigating a relationship between TFPQ and TFPR, we can more clearly understand potential problems of using TFPR measures, and we can measure the productivity improvement more precisely. Finally, I empirically examine how the resource distortions affect the aggregate productivity and whether a different pattern of resource distortions leads to a gap in aggregate productivity between regions. The result helps identify the sources and the extent of resource market distortions as well as to suggest policy implications for improving economic efficiency. I also offer a theoretical explanation for the findings from my empirical analysis, and the main insight in my theoretical analysis is novel in this research area.

1.5 Structure of the Dissertation

This dissertation is organized as follows. In Chapter 2, I theoretically examine a relationship between productivity and competition through new entry and then relate TFPR measure to TFPQ measure. In Chapter 3, I use the firm-level Chinese manufacturing data and estimate how much productivity loss of Beijing and Shanghai is attributable to labor and capital resource misallocation. In Chapter 4, I theoretically investigated a possible role of Hukou allocation system in order to provide one possible explanation for the empirical findings I found in Chapter 4. Chapter 5 concludes.

Chapter 2.

Measuring Physical Productivity Changes when New Entry Occurs

2.1 Introduction of Chapter 2

Entry of some firms by offering a new product has profound impacts on both supply and demand sides. On the demand side, such entry will allow consumers to choose from an extended variety of products, and some consumers may actually switch to another product from the product they are currently consuming. On the supply side, such entry would force incumbent firms to face a tougher competition, and they may take some actions to respond to it. For example, firms may upgrade the quality of their products or expand the line of their products to make an optimal adjustment to the new competitive environment. Also, such entry may cause resource reallocation between incumbent firms or between new entrants and exiting firms, and it may eventually leads to the transformation of market structures. A tougher competition and resource reallocation may, in turn, affect both firm-level and industry-level productivity through selection and incentive.

There have been a lot of theoretical and empirical studies that examine a relationship between competition and firms' productivity or a relationship between competition and firm's choices about a variety and quality of products. However, many studies have not examined competition, productivity and product choices in a single framework, and as a result we have known very little about how entry through a new product offering affects demand and supply factors for the determination of the level and growth of productivity at the firm- and industry-level. In this research, I build a theoretical model by incorporating all the three elements (i.e., competition, productivity and product choices) and, based on insights from the theoretical analysis, I propose an empirical framework that allows us to estimate physical productivity (TFPQ) cleanly and to infer effects of entry on productivity. To achieve this goal, I extend the model of Johnson and Myatt (2003), where both consumers and firms take a quality of product into consideration when they decide their optimal behaviors, by allowing for several types of new entry as well as by explicitly linking physical productivity (TFPQ) to revenue productivity (TFPR). In the framework of this study, consumers choose a particular quality of the product that gives them their highest utility level, whereas firms decide which quality of the product and what quantity of a particular quality of the product to supply. The key idea is that a derived first order condition from the model contains fundamental parameters of both firms and consumers. I use the optimal condition to separate supply factors from demand factors, which allows me to separate physical productivity (TFPQ) from revenue productivity (TFPR) cleanly.

It is important to distinguish physical productivity (TFPQ) and revenue productivity (TFPR) when examining entry effects on productivity of incumbent firms and industry-level productivity (Foster et al., 2008). Revenue productivity (TFPR) is often used to measure firm productivity mainly because of data availability. This conventional approach, however, has several drawbacks. First, the revenue productivity (TFPR) approach cannot distinguish whether an observed productivity change is due to technical improvement or a change in a market structure because product prices are part of the revenue productivity measurement. Therefore, this productivity measurement may fail to capture technical changes and may end up with capturing a mixture of changes resulting from both market structures and technical changes. This can be very problematic when we are interested in examining and comparing technical efficiency of new entrants and incumbent firms or spillover effects of new entrants on incumbent firms. Second, the revenue productivity (TFPR) may not be a precise measure of productivity when products are differentiated by quality. A difference in quality of particular product may be partially reflected in prices. Therefore, including prices may work as adjusting quality differences among seemingly homogeneous products when measuring firm productivity. This favors the use of the revenue productivity (TFPR). But prices also reflect other aspects such as market structures and consumers' willingness to pay. Therefore, the revenue productivity (TFPR) can be contaminated with demand conditions and it fails to capture physical productivity accurately when products are differentiated by quality. In short, the conventional approach is insufficient to understand fundamental economic issues about entry of some firms by offering a new product. The approach proposed by this study tries to overcome such shortcomings by controlling for both supply and demand factors, and provides a way of recovering physical productivity (TFPQ) from revenue data.

The main insight from the theoretical analysis of this chapter is summarized as follows. Firms' revenues are affected by business stealing effects and physical productivity improvement effects when a new entry occurs by offering the highest quality of the product. The business stealing effects can be critical for the product of the ex-highest quality product (i.e., the second highest quality product after such entry) but the revenue of other lower quality products is unaffected. On the other hand, physical productivity improvement can impact the revenue of all the existing products. Since these two effects are reflected in revenues, revenue productivity measure can be different from physical productivity unless business-stealing effects can be controlled for properly. In general, the model demonstrates that the technical efficiency improvement can be underestimated if we use revenue productivity measurement.

Many past studies assumed homogeneous products and demonstrated that if prices are controlled for, an entry effect on firm technical efficiency improvement is uniform across producers and the improvement can be estimated by using revenue productivity measurement. However, this study shows that if products are differentiated by quality, the physical productivity improvement have different impacts on the producers. This suggests that some alternative approach is needed when products are not homogeneous in order to estimate the entry effect on firm physical productivity improvement precisely. Therefore, the first contribution of this chapter is to propose a simple and novel method of estimating physical productivity changes by using quantity data only. This methodology overcomes shortcomings of conventional revenue productivity measures. Second, the framework of this analysis allows us to decompose the entry effect on firm revenue into technical improvement effect and business stealing effect. This decomposition helps us understand whether a revenue change stems from its own behavior or stems from others' behaviors. Third, the framework of this analysis permits us to relate the entry effect on firm revenue productivity to its on physical productivity. This result helps us to judge whether revenue productivity tends to over- or under-estimate physical productivity.

The chapter is organized as follows. In section 2.2, I review the literature about entry effects on productivity by focusing on theoretical insights and empirical findings relevant to this research. In section 2.3, I lay out the model proposed by Johnson and Myatt (2003) in order to prepare for presenting my expanded model. In section 2.4, I extend their model and construct the framework for several types of new entry and specify an equilibrium for each case. I also decompose entry effects on firm supply and revenue into physical productivity improvement effects and other demand side effects. Section 2.5 concludes.

2.2 Related Literature

This research is related to three streams of exiting research (i) effects of competition on productivity, (ii) quality upgrading, and (iii) the relationship between revenue productivity and physical productivity. In this research, I try to make academic contributions in these lines of research by theoretically examining how a new entry impacts firms' choice about the quality of their product, and how the quality of their product affects measurements of firms' productivity. Below I will discuss extant studies in detail in order to help to understand where my thesis study stands in the literatures as well as facilitate interpretations about insights from my theoretical analysis presented in Section2. 3.

2.2.1 Effect of Competition on Productivity

The question of whether or not competition improves firm productivity has long been one of the central topics to economics, and many studies have been trying to answer this question as well as understanding relationships between market competition and firm productivity. Past empirical studies have documented a positive correlation between product market competition and the level and growth of productivity both at the firm level and the industry level. Mainly, three explanations for this positive relationship have been proposed in the literature. The first explanation is based on selection effects. Firms with low productivity are more likely to exit from the market when market competition is intensified. Since more productive firms stay in the market and less productive firms are driven out of the market, this results in improving the average or the industry level productivity. The second explanation hinges on resource allocation. In response to some exogenous changes in the economy such as trade liberalization, resources are likely reallocated toward high productivity firms within a given industry. This reallocation leads to the improvement of the industry level productivity. The third explanation focuses on productivity improvement of a particular firm. When competition is intensified, an incentive to escape from competition becomes large. This in turn increases firm's innovative efforts and raises firm's productivity.

2.2.1.1 Selection Effect

Both theoretical and empirical literatures on industry evolution and firm survival found that high productivity firms are more likely to grow whereas low productivity firms are more likely to exit from the market. In a competitive environment, entry of more efficient firms into the market intensifies market competition and also forces less efficient firms to exit the market. This market selection process is an important channel for the improvement of the overall industrial productivity.

The main models of industry dynamics (Jovanovic, 1982; Hopenhayn, 1992; Ericson and Pakes 1995) all indicate that more productive firms stay and grow in the market and less productive firms are eventually driven out of the market in a competitive industry. In the model proposed by Jovanovic (1982), an industry is a collection of firms with heterogeneous productivity. At the advent of an industry, firms are given a time invariant efficiency parameter which is gradually revealed through the process of Bayesian learning after entry into a market. One of the findings from this model is that those firms who discover that they are efficient will survive and grow to their optimal size, while those who discover that their firms are inefficient will tend to exit. This model is known as a passive learning model, because firms do not take any actions to improve their productivity over time. In contrast to this passive learning model, Ericson and Pakes (1995) provided an active learning model in which firms can invest to enhance their productivity under competitive pressure from both within and outside the industry. Firm's profitability changes over time as a response to the stochastic outcomes of the firm's own investment, and the outcomes of other firms in the same market. They found a similar result that firm grows on average if they are successful in improving productivity as well as profitability, and exits if they are unsuccessful. Hopenhayn (1992) extends Jovanovic's model by providing a steady-state analysis of the dynamics of heterogeneous firms within a perfectly competitive industry. In a stationary equilibrium, a fraction of firms that enter the industry is equal to a fraction of firms that exit from the industry. Similar to Jovanovic (1982), this model predicted that firms whose productivity is below a given threshold level must exit the market, whereas other more productive firms will survive. This selection process will then lead to the productivity growth at industry level.

Most empirical studies have found that exiting firms are typically located in the lowest part of the productivity distribution. For example, Bellone et al. (2006) showed that exiting firms is about 5% less productive than their surviving competitors in France over the 1990-2002 periods. One of the most important findings in this line of empirical research is that the competition and market selection contributes positively to aggregate productivity growth through the selection of efficient firms. Such evidence has been found in a variety of countries. For example, Baily et al. (1992), Haltiwanger (1997), and Foster et al. (2001) have reported such evidence for the United States, Griliches and Regev (1995) for Israel, Aw et al. (2001) for South Korea and Taiwan, Disney, Haskel and Heden (2003) for UK, and Nishimura et al. (2005) for Japan.

Although we observe, by and large, that exiting firms are less productive than surviving firms, the literature has also reported that there is variation in market selection processes across countries, industries and over time. Scarpetta et al. (2002) found that market selection mechanism does not function well in mature and/or restructuring industries because high productivity firms tend to find profitable opportunities in other industries and therefore exit from those sunset industries. Aw, Chen and Roberts (2001) compared data for Taiwan and South Korea from 1983 to 1993, and showed that the market selection process against less productive firms is more effective in Taiwan than in South Korea and firm turnovers contributed more significantly to the productivity growth of manufacturing industries in Taiwan than in Korea. Bellone et al. (2006) presented the evidence that market selection mechanisms may be less efficient in France than in the US over the 1990-2002 periods.

Several empirical studies attempted to dig deeper by quantifying the contribution of this market selection effects to productivity growth. Bailey et al. (1992) decomposed industry productivity growth into the contributions of the incumbents, the entrants, and the exits in order to examine the contribution of those firms to industry productivity growth. They found that incumbent plants contributed very little to the aggregate US productivity growth. They argued that the most important channel for the overall industry productivity growth is the replacement of low productivity plants by high productivity plants. Foster, Haltiwanger, and Krizan (2006) found that the aggregate productivity growth in the U.S. retail sector is almost exclusively through the exit of less efficient single-store firms. Harris and Li (2008) found that 79% of UK productivity growth arises from the replacement of low productivity firms.

2.2.1.2 Resource Allocation

The resource reallocation among firms within a given industry is also an important channel of productivity improvement. Both theory and empirical evidence suggest that higher level of competition within an industry contributes to enhancing allocative efficiency by shifting market share and factor resources from less efficient firms toward more efficient firms, and then more productive firms should be larger or becoming larger while less productive firms should be smaller or becoming smaller. This is a mechanism through which improvement in industry level productivity is achieved.

Melitz (2003) proposed a model with firm heterogeneity in order to investigate how industry competition influences firms' export behavior and their productivity levels. One of the key results from his theoretical analysis is that greater competition shifts resources from less to more productive firms and this results in raising industry level productivity. In his model, profit-maximizing firms with heterogeneous marginal costs produce a differentiated good with increasing returns to scale in a monopolistically competitive market. Firms pay an entry cost to enter the sector and draw their productivity from a common distribution. Under this setting, Melitz found that more productive firms will survive and expand because they can charge lower prices and then have a higher market share. He also found that less productive firms are negatively affected by increasing foreign competition. In other words, free entry will induce more productive firms to enter the export market and force the least productive firms to exit. One important implication of his work is that increasing competition raises average productivity in a sector by allocating resources (e.g., market share, labor and capital) away from less productive firms towards more productive ones. While Melitz model emphasizes a resource reallocation mechanism within industries, Bernard, Redding and Schott (2007) and Okubo (2009) extended this model with multiple industries, and found greater competition can leads to resources reallocations not only within industries but also across industries.

The main implications from the Melitz model are tested against data and obtain strong empirical supports. For example, Trefler (2004) studied the impact of trade liberalization on labor productivity in Canada, and showed that in industries where fierce competition took place, labor productivity rose by 15 percent partly because low-productivity plants contracted. Pavcnik (2002) found large market share reallocations within industries, and the reallocation of resources and output from the less to more efficient firms contributes to achieving 66% of aggregate productivity improvements when firms are exposed to competition from international trade in Chile. Eslava et al. (2013) found similar evidence in Colombia that more productive firms may increase their market share not only form the less productive continuing firms but also from the less productive firms which exited the market. They also showed that this productivity gain would be lowered if physical productivity measure is used instead of revenue productivity measure.

Beyond the international trade literature, many researchers found similar results that resources are reallocated across firms and this lead to an aggregate productivity growth under competitive process. Ericson and Pakes (1995), and Asplund and Nocke (2006) constructed models that make a connection between marker share shift and productivity to investigate firm evolution. They showed that low productivity plants are less likely to survive than their more efficient rivals because of a decreased market share. The reallocation of market shares to more efficient producers can emerge from among incumbents as well as entry and exit, and this productivity-survival link plays a crucial role on industry productivity growth. Arnold et al. (2011) pointed out that productivity growth is largely driven by market reallocation from less to more productive firms, rather than through within-in firm improvements in productivity by using data from OECD countries. Baldwin and Gu (2006) took advantage of Canadian manufacturing data to examine the contribution of competition to productivity growth. They showed that the output reallocation across firms driven by competition accounts for about 70% of the overall labor productivity growth in Canadian manufacturing industries. They suggested that most previous studies likely underestimate the contribution of output reallocation results from competition to aggregate labor productivity growth.

While most of the extant studies investigated changes in resource allocation and competition, Syverson (2004) examined the difference in competition in a geographically isolated concrete industries and its impact on productivity. In his model, products are physically homogeneous and have very high transport costs. Then differences in competitiveness across markets are related to the density of concrete firms in the market. It is harder for inefficient firms to be profitable in dense markets because customers can easily shift to their more efficient competitors if they charge high prices to cover their costs. This implies that when competition is intense, inefficient firms are unable to keep market share as well as to survive, and this leads to the truncation of a productivity distribution from below. They also found in a homogeneous industry in the U.S., markets with denser construction activity have higher lower-bound productivity levels, higher average productivity, and less productivity dispersion. This empirical evidence is consistent with prediction from his model.

Moreover, there is a fast-growing literature that seeks to identify factors generating resource misallocation in a competitive environment. Many researches on market regulations argued that appropriate regulations can play in facilitating the reallocation of resources. Bartelsman et al. (2004) showed that inappropriate regulations such as high start-up costs are likely to prevent the resources reallocation from low to high productivity incumbents, leading to low firm turnover and low productivity growth.

Other works, for example Foster et al. (2008), Heish and Klenow (2009) and Midrigan and Xu (2014) suggested that efficient firms fail to attract productive labor and capital, and this causes resource distortions among firms as well as a low aggregate productivity, especially in developing countries. Hsieh & Klenow (2009) showed that job misallocating is an important part of the reason which TFP is lower in China compared to the United States, and they argued that if labor and capital could be allocated to highly efficient firms in China and India similar to the United State, then the aggregate productivity could be increased by about 50%.

2.2.1.3 With-in Effect

When competition is fierce, inefficient firms are unable to stay in the market in the long run. In order to gain cost advantages and raise survival probability, firms have a strong incentive to improve their productivity through innovation, product shift, management and other ways. In contrast to selection effects and resource reallocation, this with-in effect comes from productivity improvement at firm-level and generates an aggregate level productivity change through the productivity improvement of individual firms.

When competition is strong, firms may suffer a pressure on managers. Some studies supposed that firm make a great effort to reduce cost through improve management such as ensure production process and resources are effectively production process and resources are used effectively, and this leads to a productivity increase. For example, Bloom and Van Reenen (2010) find that strong product market competition appears to boost average management practices through a combination of eliminating the tail of ineffective production process and pushing firms to improve their practices. Based on a cross-country survey of management practices covering more than 6,000 firms, they showed that higher competition results in higher management practices, and that this also appears to improve performance. Griffith (2001) isolates the impact of competition on managerial effort by dividing her sample of UK firms into single establishments and group establishments. Single establishments are typically managed and owned by the same person, and are therefore less susceptible to principal and agency problems. On the other hand, management and ownership are typically separated in group establishments, and principal and agency problems are critical for management efficiency. She finds that an exogenous rise in competition increased the productivity of the firms likely to have principal agent problems, but not that of firms without these problems. Martin (1993) suggested that the effects of fierce competition on managers' efforts are often subtle and ambiguous. They found that competition could lead to less effort if managers are highly responsive to monetary incentives, as the opportunity for performance-related pay is reduced by intense competition.

Another channel of competition affecting firm productivity is to change the distribution of firm products. Melitz, Mayer, and Ottaviano (2014) built a theoretical model in which tougher competition increase productivity by shifting down the entire distribution of markups across products and inducing firms to skew their sales toward their better performing products. They also found empirical support for this competitive effect from the data on French exporters. Feenstra and Ma (2008) and Eckel and Neary (2010) examine product line decisions of multi-product firms, and showed an effect of competition can improve firms' productivity by dropping their worst performing products.

The presence of competition may encourage innovation and innovation can be one of the most important ways to achieve productivity improvement. For example, Cameron (2003) examines data from UK manufacturing firms and finds that a 1%

increase in R&D (closely related to innovation) raised TFP by 0.2 to 0.3%. Geroski (1990) looked at 4378 major innovations in the UK between 1945 and 1983, and found evidence against the hypothesis that increases in competitive rivalry decrease innovativeness. Blundell, Griffith, Van Reenen (1995) studied relationship between competition and innovation based on 375 firms listed on the London International Stock Exchange between 1972 and 1982, and found that dominant firms tend to innovate more and that industry concentration weakens innovative activity. Griffiths, Harrison and Simpson (2010) looked at the effect of the introduction of the Single Market Programme in Europe in the early 1990s on innovation and productivity. They find that measures to reduce internal non-tariff barriers to trade and open up competition did have the effect of increasing product market competition. This, in turn, spurred innovation. They also examined industry-level effects, and find that the Single Market Programme increased R&D intensity by 1.2% in the UK metal products industry, which was associated with a 0.7 percentage point increase in TFP growth.

The relationship between competition and innovation seems non-monotone. Although nearly all studies indicated a positive relationship between competition and innovation, Aghion et al. (2005) found an evidence of an inverted-U shape relationship between competition and innovation based on UK data. Where competition in a market is not intense, an increase in the level of competition will tend to stimulate innovative activities But, beyond a certain point, further increases in competition may have adverse effects on innovation. Correa and Ornaghi (2014) take approach similar to Aghion et al. (2005) and examine the relationship between competition and innovation by using US manufacturing data. They found a positive relationship between competition and innovation, rather than an inverted-U shape. Patent counts, TFP and labor productivity are all positively correlated with the degree of competition. They conclude that when there are well-defined intellectual property rights in a market, increases in competition will generally lead to greater levels of innovation which in turn leads to higher levels of productivity.

2.2.2 Competition and Quality Upgrade

2.2.2.1 Theoretical Analysis

In response to entry or intensified competition, firms often adjust the quality of their products in order to survive. Theoretical studies focus on investigating how firms alter their products in response to intensive market competition and such studies have found an ambiguous relationship between competition and quality adjust. This is because of competition likely affects firms' incentives of quality upgrade in opposing ways. On the one hand, competition has a positive direct effect. In a more competitive environment, firms have an incentive to increase quality in order to attract consumers. On the other hand, it has a negative indirect effect on quality, because intensive competition decreases firms' profitable margins and may undermine firms' incentives to invest in quality. The overall impact of increased competition on quality upgrade then depends on the relative strength of these two opposing forces, which is highly sensitive to model specifications.

In a seminal paper on firm's choice of vertically differentiated product quality, Mussa and Rosen (1978) considered a monopolist facing privately informed consumers with heterogeneity in their willingness to pay for quality. In order to use this consumer heterogeneity to maximize its profits, the monopolist designs a product line which offers a range of products at different qualities with different prices, and then lets consumers choose themselves. The main insight is that quality provision can be distorted. More specifically, a firm may offer inefficiently low qualities for consumers with lower valuation for quality in order to reduce substitution possibilities of consumers with higher valuation for quality and to extract more surpluses.

A large theoretical literature on optimal product quality decision extended the model of Mussa and Rosen (1978) by considering multiple product, multi-dimensional consumer types or competition. For example, Gal-Or (1983) extended Mussa and Rosen's model to an oligopoly case. In the model, she explored how the change in the number of firms, as a measure of the intensity of competition, affects firms' equilibrium choice of price and quality. She demonstrated that competition tends to decrease the average level of quality in equilibrium. Ma and Burgess (1993) considered a case in which firms may use both quality and price to compete for customers in imperfectly competitive markets. They show that firms' incentives to invest in quality can be undermined by the prospect of triggering fierce price competition rivals. However, when prices are fixed and then firms can only choose quality to attract consumers, the potential margin reducing effect of competition disappears. That is, the absence of price competition can help adjust the distorting effect of quality competition.

Itoh (1983) considered a discrete number of products in the Mussa-Rosen framework and analyzes the effects on product prices and qualities of introducing new products. One important insight of this work is that the introduction of a new product has no effect on the optimal price of lower quality products and has an effect of indeterminate sign on the optimal price of higher quality products offered by the same firm. Moreover, he studied a special case where consumer's willingness to pay for marginal increases in product quality is uniformly distributed and showed that introducing a new product has no effect on the price of all other products from the same firm in this case. Itoh's model implicitly used an upgrades approach, which views higher product qualities as a base quality plus a series of quality upgrades in the range and then each upgrade can be associated to a price premium.

This upgrades approach was pioneered by Johnson and Myatt (2003). They emphasized that upgrades approach is useful in analyzing product quality and pricing choices of multiply product firms in monopoly and Cournot duopoly cases. Utilizing this approach, they studied the effects of competition on multiproduct firm's product quality decisions by considering the product quality choices as a response of an incumbent to entry by another multiproduct firm. In their model, a single firm enters a market originally dominated by a monopolist. Each of the duopolists then offers a range of quality-differentiated products and competes in quantities. As a response to new entrant, the incumbent either offers "fighting brands" by expanding their product line into a lower quality product with a lower price, or engages in "line pruning" by eliminating a range of product qualities. Their major result is that whether the incumbent will choose to expand or contract its product qualities depends on the shape of the marginal revenue curves in the market. When marginal revenue is decreasing, competitor's entry induces a restriction in the output of the incumbent's low-quality products and as a result leads to incumbent's exit from the lower qualities product markets. On the other hand, when marginal revenue is increasing, incumbent's optimal response to competitor's entry will be introducing lower quality products and expand into lower quality product markets.

2.2.2.2 Empirical Analysis

Theoretical predictions about the effect of competition on quality adjustment are not unambiguous, and they are an interesting empirical question. In fact, empirical research found that changes in competition levels can have positive or negative effects on quality depending on the particularity of markets under consideration. Intensified competition can cause quality levels to rise if firms decide to compete in quality attributes, or to fall if price competition leads to a reduction in quality as production costs are cut. Therefore, we usually need rely on empirical work to determine how quality will change in response to varying degrees of competition in most particular markets.

There are some empirical studies that documented a positive effect of competition on product quality. In particular, such studies found that increasing product variety quality is a common strategy in response to entry in some particular markets. For example, Geroski (1995) reports some case studies where the incumbents introduce new products which they had been holding back. Smiley (1988) shows that 26% of the established firms choose to respond by increasing their product variety in the US manufacturing. Matsa (2010) studied the effect of competition on firm's incentive to provide product quality in US supermarket. He reported that how incumbents adjusted their inventories in response to entry of Wal-Mart stores and found that chain stores tended to respond to Wal-mart's entry by improving the quality of their products. Bennett et al. (2013) examined empirically whether more competition among New York's vehicle emissions testing centers led to a lower quality. They found that competition among these emissions testing centers can induce firms to increase quality for their customers in ways that are both illegal and socially costly.

Other empirical studies reported negative effects of competition on product quality. McMaster (1995) investigated how the quality of health care services in the U.K changes when competition is intensified among health care providers. They found that the quality worsens when competitive bidding was introduced. Becker and Milbourn (2011) looked at corporate bond and issuer ratings between the mid-1990s and mid-2000s and concluded that increased competition among the ratings agencies reduced rating quality. They indicated that competition most likely weakens reputational incentives for providing quality in the ratings industry and, thereby, undermines their quality. Fan and Yang (2016) studied how a change in competition affects the number of product offerings in the U.S. smart phone market. They showed that a reduction in competition as a result of a merger decreases both the number and variety of products.

In some markets, results in the relationship between quality and competition are mixed. For example, in the airline industry, empirical studies report both positive and negative effects. Rose (1990) investigated the competition on firms' incentives to provide high quality services in the US airlines industry by using airline incident data. One result from this work is that that increased competition will erode firm profits, and this is indeed correlated with declines in safety and quality. However, Dunn (2008) reported an opposed effect of competition on quality in the same airline industry. He provides evidence that a firm is more likely to start a high-quality service in response to entry if the firm has an existing low-quality service in the market.

Several empirical studies try to identify factors for the conflicting results about competition's effect on quality. For example, in hospital services markets, empirical studies presented how competition affects quality depends on price regulation, which is consistent with theoretical conclusions. Under a regulated price regime, empirical evidence shows that competition improves quality. Gaynor and Town (2011) detected significant improvements in mortality and reductions in the average length of stay without changes after introducing competition in the UK health system in 2006. Bijlsma et al. (2010) found that in the Dutch hospital sector after implementing pro-competitive reforms, competition between hospitals put pressure on profits margins and force hospitals to pay more attention to quality as well as to improve quality indicators. On the other hand, the evidence is mixed without price regulation. Kessler and McClellan (2000) examined the quality of US hospital services and show that quality is significantly lower in markets that are more concentrated. Vogt and Town reviewed ten studies about the effect of competition on quality of US hospital services and concluded that quality declines when hospital market concentration increases. However, Sohn and Rathouz (2008) reported the mortality was lower for hospitals that faced more competition.

Seim and Viard (2011) study how entry into US cellular phone markets affects the number of calling plans offered by each incumbent firm. They reported that when the initial number of firms is small in a local market, entry reduces the number of calling plans offered by incumbents. However, when the initial number of firms is large, incumbent firms increase the number of calling plans.

Furthermore, based on the theoretical prediction that competition increases quality in market where price competition is absent, some empirical study tries to investigate how competition change product quality by controlling for the effect of competition on price. Manez-Castillejo (1999) investigates the intensity of price competition across the quality gradient in the UK supermarkets. He found that price competition is weaker for higher quality branded products and stronger for lower quality branded products. Deltas et al. (2010) found results by using personal computer price data in the BuyDirect market. They indicated that price competition is stronger for lower quality computers than for high quality computers.

2.2.3 Physical Productivity and Revenue Productivity

Productivity is commonly defined as a ratio of output to input. The output could be the products produced or services rendered, and the input includes the resources used to produce the output. There are many different approaches to measuring productivity. To list a few, total factor productivity (TFP), multi factor productivity (MFP) and labor productivity (LP) are such measurements, which differ in terms of the categories of included input. Most economists prefer the TFP measures since it is the most comprehensive of possible productivity statistics.

The basic definition of total factor productivity is the rate of transformation of total input into total output. Two main approaches to measure this productivity have been used in empirical research: physical productivity (TFPQ) and revenue productivity (TFPR). Revenue productivity approach measures the output in financial value (e.g., Value added, sales) whereas physical productivity approach measures the output in physical units (e.g., number of customers served, number of product). TFPQ measurement is ideal because it reflects the physical effectiveness and efficiency of a production, and it is not affected by price fluctuations. However, data on physical units of outputs and inputs are hardly available for empirical research. Because of data availability, TFPR is commonly used as productivity indicator in past studies, but this way of measuring is sometimes problematic. For example, Foster et al. (2008) argue that a higher revenue productivity may not imply higher technologically efficient if prices reflect demand shifts or market power variation rather than quality or production efficiency differences.

2.2.3.1 Physics Productivity Measurement

The physics productivity (TFPQ) is based on quantities of physical outputs. According to Foster et al. (2008), physical productivity of goods i can be simply expressed as

$$\mathrm{TFPQ_i} = \frac{q_i}{x_i} = \omega_i \ ,$$

where q_i denote the quantity of good i, x_i and ω_i is the input and the technical efficiency level respectively. From the equation, we can see clearly that the TFPQ should not be affected by the price of the output or input. It simply shows the rate of transformation of input to a quantity of output, and equals the producer's true technical efficiency of goods i.

Since the physical productivity reflects the true efficiency of transformation, this is an ideal productivity measurement when we are interested in examining technical efficiency or cost changes realized, or comparing efficiency differences between two firms. However, it has several difficulties when we apply it to empirical works. First, Diewert and Nakamura (2007) indicated that the information on output in quantities is typically unavailable to the researcher, especially in the industry or nation level. Second, it is difficult to make comparison with others. To be meaningfully interpreted, productivity measures usually need to be placed in a comparative context. However, Beveren (2012) discussed that firm's physics productivity is difficult to aggregate into industry level if firms produce multiple products within the same industry, because the unit which used to measure output is quite distinct and is nonsense to compare to others. Therefore, physics productivity can be problematic when we try to compare with different firms and industries.

2.2.3.2 Revenue Productivity Measurement

The revenue productivity (TFPR) is based on an amount of monetary output, such as value-added and sales which are commonly used. Revenue productivity of goods i can be formally expressed as

$$TFPR_i = \frac{p_i q_i}{x_i} = p_i \omega_i$$

where p_i is the price of goods i. The equation shows clearly that the $TFPR_i$ positively correlates with true productivity ω_i and includes the effect of price. If we have the information about p_i, then we can make a connection between revenue productivity and true productivity ω_i, as well as convert one into the other. However, the information on firm-level output price is usually unavailable in empirical work.

Empirical work typically uses revenue productivity measurement simply because of its convenience. First, it is easy to collect data, because information needed for estimating this TFP measure can be found in firm's financial report directly. Second, it is convenient to compare with different output. Beveren (2012) discussed that since revenue based output is measured in monetary units, it allows the aggregation of different output, and then can be easily compared through firms and industries which produce different output.

Although revenue productivity measurement is a widely used approach of measuring productivity, there are several problems in using it, especially when we use micro data. First, as the equation above shows, its measurement is potentially distorted by price effects. Foster et al. (2008) pointed out that revenue productivity is not able to cleanly measure the true efficiency. This is simply because revenue productivity is basically measured by a monetary unit of outputs, and then contains price components. Second, TFPR approach cannot tell us whether a productivity change is due to technical improvement, change in market structures, taste changes of consumers and so forth. While the TFPR is positively correlated with true productivity, Foster et al. (2008) discussed that it captures a mixture of demand and supply effects and fails to capture the real productivity change. Changes in prices can be caused by demand-side factors like consumers' willingness to pay rather than supply-side factor such as cost savings, and then the high productivity firm may not be particularly efficient. For example, firms can have a high TFPR levels because they are efficient, but this can also be driven by high price resulting from a producer-specific demand. Moreover, Beveren (2012) explained that a bias in estimated TFPR may arise when we use firm-level output data without the information on firm-level output price in empirical work. In the absence of information on firm-level prices, industry-level price indices are usually applied in traditional production function estimates. However, if the firm-level prices is different from the industry-level price, it will lead to a biased estimation of that particular firm's output for a given inputs, and hence result in a bias of the real productivity.

These difficulties do not mean that researchers should choose using measures of physical-based compared to revenue-based. TFPR measures have the advantage that they reflect profitability factors beyond TFPQ. That is, in general TFPR measures reflect both technology and demand factors. Since both factors are likely important for firm growth, it is useful to capture both when we interesting in studying the relationship between profitability and evolution of firms.

2.2.3.3 Distinction between TFPR and TFPQ

The physical productivity approach and revenue productivity approach are different in output measures, and have pros and cons when they are used in empirical work. Therefore understanding distinctions of these measurements and how empirical results correspond to these distinctions becomes important.

Foster et al. (2008) investigated differences between revenue-based and physical output measure of productivity by using both standard micro data and rare quantity data. They found that the correlation between the revenue-based productivity and the physical-based productivity is 0.75. But, revenue-based measure may not be able to estimate the real technical efficiency consistently because price effects cannot be ignored. They compare the impacts of revenue-based productivity measure and physical-based productivity measure on selection and productivity growth in industries. An important result from their research is that within industry productivity dispersion observed in physical-based productivity measure is more dispersed than that in revenue-based productivity measure, and they explained that this differences arise from physical productivity is negatively correlated with firm-level prices while revenue productivity is positively correlated with price.

The differences between TFPQ and TFPR also have been highlighted by Hsieh and Klenow (2009, 2014). Under some specific assumptions about the functional forms of the demand and the production functions which are different than Foster et al. (2008), they showed that there will be no dispersion in TFPR in an economy without distortions, even if there is dispersion in TFPQ. Their argument is that the observed TFPR dispersion must partly reflect distortions. This result shows the importance of distinguishing between TFPQ and TFPR.

Foster et al. (2008) argued that by distinguishing between TFPQ measures and TFPR measures, one can distinguish and understand the impact of demand side factors and supply side factors for the evolution of firms. They decomposed the prices into technology and demand fundamentals, and show that the fact that firms with lower revenue productivity are more likely exit is due to lower demand rather than lower technical efficiency. They suggest that the demand variations across firms are the dominant factor in firm survival in fact. Eslava et al. (2004) investigated and the role of demand side factors on productivity growth using Colombian data, and suggests that both the demand side differences and the TFPQ differences are important to the differences in firm size and firm growth, but the former are even more important. This is consistent with the result from Foster et al. (2008).

A number of studies have utilized firm-level price data in order to understand the relevance of these two measurements as well as to explore insights by distinguishing these two measurements. For example, using Colombian data, Eslava et al. (2013) decomposed revenue into the price and quantity components by using firm-level prices, and estimated the physical-based productivity. They reported that the correlation between the revenue-based productivity and the physical-based productivity is 0.69, which is close to the number reported in Foster et al. (2008). They also found that high levels of dispersion in revenue-based productivity and physical-based productivity within industries, and that dispersion in TFPQ exceeds that of TFPR.

Many studies have already noted potential limitations on revenue-based productivity measures as well before the study by Foster et al. (2008), and provided alternative methods that attempt to overcome these difficulties arising from unavailable firm-level price data. Abbott (1992) pointed out the extent of price dispersion within industries, and outlines possible limitation for revenue productivity measurement aggregation. Klette and Griliches (1996) and Mairesse and Jaumandreu (2005) considered how within industry price fluctuations affect production function and as a result to affect revenue productivity estimates. Melitz (2000) and Loecker (2005) have extended these analyses to multi-product producers and factor price variation. Katayama, Lu, and Tybout (2003) argued that technical efficiency or product quality can be very misleading by revenue-based output measures, and developed an alternative approach of inferring the quantities, qualities and prices of each goods from the observed revenues and expenditures.

2.3 Model

The main purpose of this section is to lay out the model proposed by Johnson and Myatt (2003) in order to understand a mechanism through which the distribution of consumers' taste parameters and firms' productivity jointly determine optimal supply of products with different quality as an equilibrium outcome. In section 2.4, this basic model is extended to allow for new entry and the extended model provides novel insights regarding effects of new entry on measured productivity.

2.3.1 Consumers

I consider a model where each consumer purchases one unit of a product from a set of n different qualities of the product. These n different qualities of the product are differentiated vertically in the sense that all consumers agree to order these n types of the product from the highest quality to the lowest quality. For convenience, let q_j denotes the j-th quality level of the product, and assume that $q_1 < q_2 < \cdots < q_n$.

Consumers have a different taste over quality of the product, and their utility from the product depends on a quality of the product and its price. Formally, consumer i's utility from the purchase of the product with quality q_j is specified as

$$v_{ij} = \theta_i q_j - p_j \ , \tag{2-1}$$

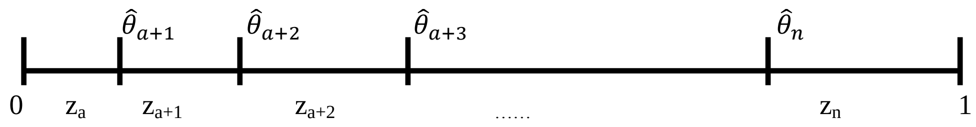
where θ_i is consumer i's taste parameter, and p_j is a price of the product with quality q_j. The marginal valuation of quality increases with θ so that this utility function satisfies the single-crossing property. The taste parameter θ is distributed according to a distribution function *F*.

Given that a consumer's utility is normalized to zero when she does not purchase the product at all, the consumer with $\hat{\theta}_1$ is indifferent between buying the q_1 quality product and not buying any types of product, where $\hat{\theta}_1$ satisfies $\hat{\theta}_1 q_1 = p_1$. Similarly, the consumer with $\hat{\theta}_j$ is indifferent between buying q_j quality product and buying q_{j-1} quality product, where $\hat{\theta}_j$ satisfies $\hat{\theta}_i(q_i - q_{i-1}) = p_i - p_{i-1}$. Denoting a demand for q_j quality product by z_j, I can write

$$\hat{\theta}_j = 1 - \sum_{k=j}^{n} z_k \quad . \tag{2-2}$$

Figure 2.1 shows a relationship between θ and z graphically.

Figure 2.1



A fraction of "potential" buyers of the product with quality q_j is given by

$$\sum_{k=j}^{n} z_k = 1 - F\left[\left(p_j - p_{j-1}\right)/\left(q_j - q_{j-1}\right)\right] \quad . \tag{2-3}$$

Using equation (2-3) and defining $q_0 = p_0 = 0$, an inverse demand function of q_j quality product is written as

$$p_j - p_{j-1} = \left(q_j - q_{j-1}\right)H\left(Z_j\right) \ , \tag{2-4}$$

where

$$Z_j = \sum_{k=j}^{n} z_k \ , \tag{2-5}$$

and

$$H\left(Z_j\right) = F^{-1}\left(1 - Z_j\right) \ \text{ for } \ Z_j \in (0,1) \quad . \tag{2-6}$$

The cumulative variable Z_j is interpreted as the total demand at quality q_j and above. Equation (2-4) suggests that a price of quality upgrade from q_{j-1} to q_j is $p_j - p_{j-1}$ and the price of upgrade depends only on the cumulative variable Z_j. A benefit of this approach is that I do not need to pay attention to how Z_j consists of from each z_i.

The product with quality q_j is in positive supply in an equilibrium if $Z_j - Z_{j+1} > 0$, and it is in zero supply if $Z_j - Z_{j+1} = 0$. If q_j is the minimum quality in positive supply, I have the $Z_k = Z_j$ for $k < j$. Using this fact and $p_j = \sum_{i=1}^{j}(p_i - p_{i-1})$, I must have

$$p_j = H\left(Z_j\right)q_j \quad . \tag{2-7}$$

Equations (2-4) and (2-7) characterize the demand system for different qualities of the product.

2.3.2 Firms

This section examines a firm's optimal choice of product supply. I consider a case where there are M firms in this industry and these firms compete through the quantity of the product supplied to the market. Let us assume that there is a constant marginal cost, c_j, of providing quality q_j product.

Firm m chooses $(z_{m,1}, \dots, z_{m,n})$ so as to maximize its profit $\Pi^m = \sum_{j=1}^{n} z_{m,j}(p_j - c_{m,j})$. Using equation (2-7), the maximization problem of firm *m* can be formulated as

$$\text{Max } \sum_{j=1}^{n} Z_{m,j}\left[(q_j - q_{j-1})H\left(\sum_{k=1}^{M} Z_{k,j}\right) - (c_{m,j} - c_{m,j-1})\right]$$

$$\text{subject to } Z_{m,j} \le Z_{m,j-1} \text{ for each } j \ .$$

This formulation greatly simplifies the optimization problem since I do not need to consider how each of $(z_{m,1}, \dots, z_{m,n})$ is combined to maximize the profit.

Differentiating the objective function with respect to each $Z_{m,j}$, I obtain first order conditions for this optimization problem as

$$H\left(\sum_{k=1}^{M} Z_{k,j}^*\right) + Z_{m,j}^* H'\left(\sum_{k=1}^{M} Z_{k,j}^*\right) \ge \frac{c_{m,j} - c_{m,j-1}}{q_j - q_{j-1}} \ , \tag{2-8}$$

with equality if $Z_{m,j}^* < Z_{m,j-1}^*$.

Define a "marginal" productivity index as

$$\omega_{m,j} \equiv \frac{c_{m,j} - c_{m,j-1}}{q_j - q_{j-1}} \ . \tag{2-9}$$

When $\omega_{m,j}$ increases with j, the inequality that $Z_{m,j}^* < Z_{m,j-1}^*$ is guaranteed under regularity conditions. Therefore, equation (2-8) holds with equality:

$$H\left(\sum_{k=1}^{M} Z_{k,j}^*\right) + Z_{m,j}^* H'\left(\sum_{k=1}^{M} Z_{k,j}^*\right) = \omega_{m,j} \ . \tag{2-10}$$

In this case, a firm provides all levels of qualities of products, $\left(z_{m,1}, \dots, z_{m,n}\right)$. On the other hand, when $\omega_{m,j}$ decreases with j, I must have $Z^*_{m,j} = Z^*_{m,j-1}$ for any j. The firm therefore provides only the highest quality of product, $z_{m,n}$.

To proceed further, I assume that $\omega_{m,j}$ increases with j for all m so that equation (2-10) holds for any j and m. As in the Cournot model, a Nash equilibrium consists of $(Z^*_{1,1}, .., Z^*_{1,n}, \dots, Z^*_{M,1}, .., Z^*_{M,n})$ that satisfy a system of equation (2-10). Once I obtain $(Z^*_{1,1}, .., Z^*_{1,n}, \dots, Z^*_{M,1}, .., Z^*_{M,n})$, I can derive an equilibrium supply of each firm, $(z^*_{1,1}, .., z^*_{1,n}, \dots, z^*_{M,1}, .., z^*_{M,n})$, from equation (2-5).

2.3.3 Relationship between Cost and Conventional Productivity Measures

To see how $\omega_{m,j}$ is related to the conventional productivity measure, I write the revenue of firm m from the j-th quality product as

$$R_{m,j} = z^*_{m,j} H\left(\textstyle\sum_{k=1}^{M} Z^*_{k,j}\right) q_{m,j} \ ,$$

by using equation (2-7). Combining it with the first-order condition (2-10), I have

$$R_{m,j} = z^*_{m,j} q_{m,j} \left(\omega_{m,j} - Z^*_{m,j} H'\left(\textstyle\sum_{k=1}^{M} Z^*_{k,j}\right)\right) \ . \qquad (2\text{-}11)$$

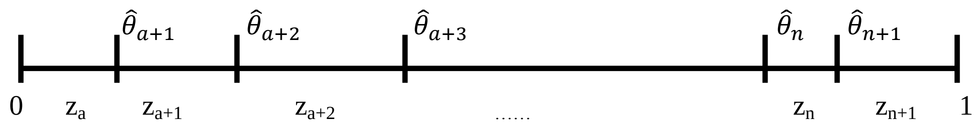
Equation (2-11) shows that the value-added approach is problematic when I am interested in measuring cost changes (i.e., physical productivity changes) realized by productivity improvement.[1] A change in R_j can be caused by demand, even if ω_j is held constant. For example, an introduction of a new product may cause a change in market share, and this result in changing R_j, even though there is no change in physical productivity ω_j. Similarly, equation (2-11) also implies that productivity improvements can affect the measurement R_j through a change in $Z^*_{k,j}$. Overall, the value-added approach captures a mixture of demand change and cost change, and cannot isolate ω_j from other factors.

[1] In this discussion, we assume that material costs are negligible in the sector.

2.4 Entry of Higher Quality Product

I extend the basic model by allowing firms to produce a new product whose quality is higher than the highest quality of exiting products. In this case, a new cutoff $\hat{\theta}_{n+1}$ emerges and consumers with taste parameter $\hat{\theta}_{n+1}$ and above are willing to buy the higher quality product q_{n+1} instead of the product with quality q_n. Figure 2.2 shows the new relationship between θ and z graphically.

Figure 2.2



The introduction of such a new product can influence revenues of all the existing products through two channels, and as a result affects revenue productivity. One channel is that the supply of existing products is optimally readjusted. Since there are some consumers who switch to the new product, this leads to a decrease in the demand for all the existing products, and lowers the revenue of the existing products. The other channel is that such a new entry may also bring technology advancement to the industry, and may improve productivity of all the existing products through spillover effects. Such physical productivity improvement increases revenues of the existing products, and affects their revenue productivity. There is no change in physical productivity, but revenue productivity changes in the former case whereas in both physical productivity and revenue productivity change in the latter case. In this section, I distinguish changes in physical productivity and revenue productivity and I analyze sources of changes in revenue productivity when new entry occurs.

2.4.1 Pre-Entry Equilibrium

In this section, I derive pre-entry equilibrium outcomes by assuming that consumer's taste parameter is distributed uniformly in order to obtain sharper theoretical insights. As mentioned in 2.3.1, θ_j is the taste parameter of consumer j and can be written as

$$\hat{\theta}_j = 1 - \sum_{k=j}^{n} z_k \equiv 1 - Z_j \quad \text{for } j = 1,2,\dots,n \quad .$$

To proceed, assume that consumer's taste parameter θ is uniform distribution. Then the density function is given by

$$f(\theta) = \begin{cases} 1 & \text{for } 0 \le \theta \le 1 \\ 0 & \text{otherwise} \end{cases} \quad ,$$

and the cumulative density function is

$$F(\theta) = \theta \qquad \text{for } 0 \le \theta \le 1 \quad .$$

Using $F(\theta)$, I obtain the function of $H(\cdot)$ as

$$H\left(\sum_{k=j}^{n} z_k\right) = F^{-1}\left(1 - \sum_{k=j}^{n} z_k\right) = 1 - \sum_{k=j}^{n} z_k \qquad \text{for } \sum_{k=j}^{n} z_k \in (0,1) \quad .$$

The expression for $H'(\cdot)$ can be obtained by using the fact that

$$H'\left(\sum_{k=j}^{n} z_k\right) = -\frac{1}{f\left(1-\sum_{k=j}^{n} z_k\right)} \quad ,$$

and then I have $H'\left(\sum_{k=j}^{n} z_k\right) = -1$.

Substituting the expression of $H(\cdot)$ and $H'(\cdot)$ into the first order condition (2-10) yields

$$\left(1 - \sum_{k=j}^{n} z_k^*\right) - \left(\sum_{k=j}^{n} z_k^*\right) = \omega_j \quad . \qquad (2\text{-}12)$$

The total demand at quality q_j and above is given by

$$Z_j^* = \sum_{k=j}^{n} z_k^* = \frac{1-\omega_j}{2} \quad \text{for } j = 1,\dots,n \quad , \qquad (2\text{-}13)$$

and the equilibrium supply of product j can be obtained as

$$z_j^* = \begin{cases} Z_j^* - Z_{j+1}^* = \frac{\omega_{j+1}-\omega_j}{2} & \text{for } j = 1,\dots,n-1 \\ \sum_{k=j}^{n} z_k^* = \frac{1-\omega_j}{2} & \text{for } j = n \end{cases} \quad . \qquad (2\text{-}14)$$

From the equation (2-14), we can see that the equilibrium supply at quality q_j is only influenced by the difference in productivity between q_j and q_{j+1}, and it does not depend on the product at other quality level.

Combining the equilibrium supply (2-13) with the first order condition (2-12), the firm's revenue from the j-th quality of the product is given by

$$R_j^* = z_j^* H\left(\sum_{k=j}^{n} z_k^*\right) q_j = \begin{cases} \frac{(\omega_{j+1}-\omega_j)(1+\omega_j)}{2} q_j & \text{for } j = 1, \dots, n-1 \\ \frac{(1-\omega_j)(1+\omega_j)}{2} q_j & \text{for } j = n \end{cases} . \qquad (2\text{-}15)$$

Since the demand and supply of the product at quality j only depend on the productivity of product at quality j and j+1, the firm revenue from product j then is only affected by the productivity of itself and its upgrade product.

2.4.2 Post-Entry Equilibrium

This section allows a firm to enter the market by offering a new product, and derives pre-entry equilibrium outcomes regarding the supply of existing and new products and their revenues. Here, the quality of the new product q_{n+1} is higher than that of all existing products, q_j for j=1,2,…,n. Let $Z_j{}'$ and $\omega_j{}'$ denote new demand and productivity of product at quality level j, respectively, after such an entry takes place.

Then, a new equilibrium can be obtained by taking similar steps as outlined in section 2.4.1. The new demand of product at quality q_j and above is given by

$$Z_j^{*\prime} = \sum_{k=j}^{n+1} z_k{}^{*\prime} = \frac{1-\omega_j{}'}{2} \quad \text{for } j = 1, \dots, n+1 \ , \qquad (2\text{-}16)$$

and the new equilibrium supply is given by

$$z_j^{*\prime} = \begin{cases} Z_j^{*\prime} - Z_{j+1}^{*}{}' = \frac{\omega_{j+1}{}' - \omega_j{}'}{2} & \text{for } j = 1, \dots, n \\ \sum_{k=j}^{n+1} z_k{}^{*\prime} = \frac{1-\omega_j{}'}{2} & \text{for } j = n+1 \end{cases} . \qquad (2\text{-}17)$$

Similarly, the new firm revenue from the j-th quality of product becomes

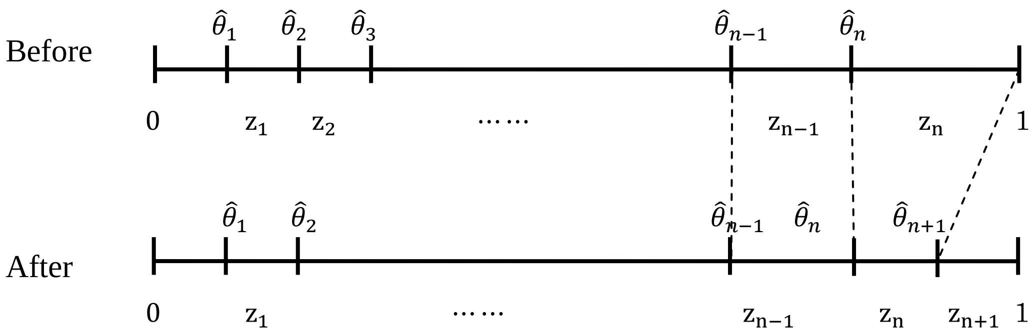
$$R_j^{*\prime} = \begin{cases} \frac{(\omega_{j+1}{}' - \omega_j{}')(1+\omega_j{}')}{2} q_j & \text{for } j = 1, \dots, n \\ \frac{(1-\omega_j{}')(1+\omega_j{}')}{2} q_j & \text{for } j = n+1 \end{cases} . \quad (2\text{-}18)$$

2.4.3 Entry Effects with Constant Productivity

I first consider a simple case where the entry of the product with quality q_{n+1} does not influence the physical productivity of the existing products, so only a market share change effect exists after the entry. Therefore, I assume constant productivity $\omega_j = \omega_j{}'$ so that the entry effect arising from demand changes is isolated from the effect of physical productivity changes.

Figure 2.3 shows the relationship between θ and z before and after an introducing a new higher quality product with constant productivity improvement graphically.

Figure 2.3



Combing this constant productivity condition and taking the difference between equilibrium supply before and after the introduction of the product with quality q_{n+1} , a change in the equilibrium supply is given by

$$\Delta z_j' = z_j^{*\prime} - z_j^* = \begin{cases} 0 & \text{for } j = 1, \dots, n-1 \\ \frac{\omega_{n+1}{}' - 1}{2} & \text{for } j = n \\ \frac{1-\omega_{n+1}{}'}{2} & \text{for } j = n+1 \end{cases} . \quad (2\text{-}19)$$

As explained in section 2.4.1, the equilibrium supply of product j at quality level below the highest quality before entry depends only on its own productivity ω_j and its upgrade productivity ω_{j+1}.

For the product with quality q_n that was the highest quality level before this entry, the equation (2-19) shows that $z_n^{*\prime} - z_n^* < 0$ when $\frac{1-\omega_{n+1}{}'}{2} > 0$. Note that $\frac{1-\omega_{n+1}{}'}{2}$ is the equilibrium supply of the new highest quality product. Then the equation (2-19) implies that as long as this new product is supplied to the market, the supply of the product with quality q_n will decrease. Since the equilibrium supply of product at quality level below q_n will not change under the constant productivity assumption, the introduction of the new highest quality product q_{n+1} only takes a market share away from the ex-highest quality product q_n. This insight is very similar to the insight we can obtain from the idea of limiting price (Holmes and Schmitz, 1990).

Proposition 2-1. *If physical productivities of existing products do not improve after entry, the supply of highest quality product before entry decrease, and the supply of other existing products keep constant.*

A change in firm revenue from the j-th quality product can be expressed as

$$\Delta R_j' = R_j^{*\prime} - R_j^* = \begin{cases} 0 & \text{for } j = 1, \dots, n-1 \\ \frac{(\omega_{j+1}-1)(1+\omega_j)}{2} q_j & \text{for } j = n \\ \frac{(1-\omega_{j+1})(1+\omega_{j+1})}{2} q_{j+1} & \text{for } j = n+1 \end{cases} . \qquad (2\text{-}20)$$

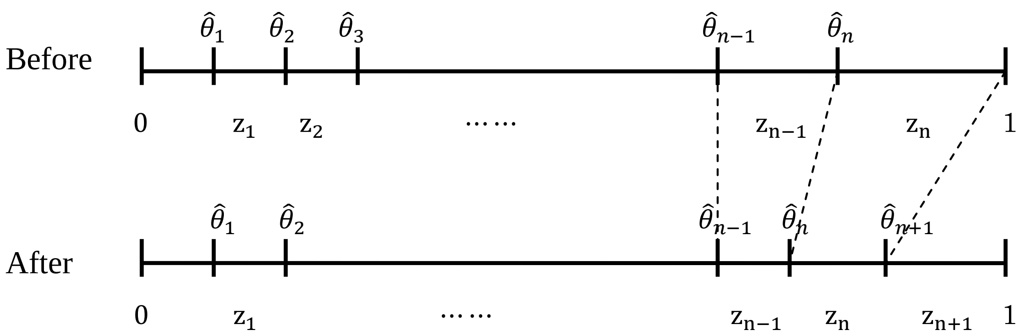
Equation (2-20) shows clearly that the introduction of the new highest quality product will not influence the revenue of the product at quality below the quality n because there is no demand and supply change in these products as long as physical productivity does not change at all. For the product with quality q_n, a part of consumers switch to the new highest quality product, and this leads to a loss of a market share as well as a loss of firm revenue.

Proposition 2-2. *If physical productivities of existing products do not improve after new entry, the revenue from the ex-highest quality product,* q_n*, decreases, and the revenues from other existing products remain constant.*

2.4.4 Entry Effects with Limited Productivity Improvement

In this section, I relax the assumption that physical productivity is unaffected by the introduction of a new product. I begin my analysis by assuming that the introduction of the new highest quality product q_{n+1} affects only the physical productivity of the ex-highest quality product q_n. This case is called a limited productivity improvement case. Figure 2.4 illustrates the relationship between θ and z before and after introducing the product with quality q_{n+1} in the limited productivity improvement case.

Figure 2.4



2.4.4.1 Equilibrium with Limited Productivity Improvement

Consider the limited productivity improvement case in which $\omega_n{}' - \omega_n < 0$

holds. I will demonstrate how the equilibrium supply and firm revenue from the product with q_n has been influenced in the limited productivity case.

By subtracting equation (2-14) from equation (2-17), I get a change in equilibrium supply of all the existing products as following:

$$\begin{cases} \Delta\hat{z}_{n-1} = {z^*_{n-1}}' - z^*_{n-1} = \frac{\omega_n' - \omega_n}{2} \\ \Delta\hat{z}_n = {z^*_n}' - z^*_n = \frac{(\omega_{n+1}' - \omega_n') - (1-\omega_n)}{2} \\ \Delta\hat{z}_{-n,-n-1} = 0 \end{cases}, \quad (2\text{-}21)$$

where $\Delta\hat{z}_{-n,-n-1}$ is the supply change of the existing products except the products n and n-1. Equation (2-21) indicates that the supply of the quality n product and its the quality n-1 product will change whereas the supply of other existing products remain constant. It follows from equilibrium supply equation (2-14) that the supply of the j quality product only depends on the physical productivity of itself and the physical productivity of the j+1 product. Therefore, physical productivity improvement of the quality n product only influences the supply of itself and the quality n-1 product.

Since I assume that $\omega_n' - \omega_n < 0$, it holds that $\hat{z}_{n-1} = \frac{\omega_n' - \omega_n}{2} < 0$. That is, the supply of the n-1 quality product decreases. This is a business steal effect. More specifically, the productivity improvement of the n quality product is translated into a lower price than before, and consequently consumers are attracted to this product more than the n-1 quality product. This decreases the demand for the n-1 quality product, and its market share drops. However, a change in the supply of the n quality product is ambiguous. The productivity improvement of the n quality product works for increasing its demand by taking the market share away from the n-1 quality product, but its market share can also be taken away by the n+1 quality product. Therefore, the supply of the n quality product may decrease or increase depending on the magnitudes of these two opposing effects.

I now turn to investigating entry effects on revenues of the products. Using the revenue equations (2-15) and (2-18), I have a change in the firm revenue from all the existing products:

$$\begin{cases} \Delta\widehat{R}_{n-1} = {R^*_{n-1}}' - R^*_{n-1} = \frac{({\omega_n}' - \omega_n)(1+\omega_{n-1})}{2} q_{n-1} \\ \Delta\widehat{R}_n = {R^*_n}' - R^*_n = \frac{({\omega_{n+1}}' - {\omega_n}')(1+{\omega_n}') - (1-\omega_n)(1+\omega_n)}{2} q_n \\ \Delta\widehat{R}_{-n,-n-1} = 0 \end{cases} . \quad (2\text{-}22)$$

We can see that $\Delta\widehat{R}_{n-1} = \frac{({\omega_n}' - \omega_n)(1+\omega_{n-1})}{2} < 0$, because ${\omega_n}' - \omega_n < 0$. That is, the revenue of the n-1 quality product decreases because of the business stealing effects I mentioned above.

As in the supply case, however, a change in the revenue of the n quality product is not clear. The productivity improvement increases its supply and thus will raise the revenue, but its market share can also decrease due to the introduction of the n+1 quality product and this will reduce the revenue. If the revenue increase arising from productivity improvement exceeds the revenue decrease resulting from the market share loss, the revenue from the n quality product will increase. The revenue will decrease when the situation above is reversed. The revenue from other existing products remains constant because there is no change in demand for and supply of those products.

In sum, when the introduction of the new highest quality product brings physical product improvement to the n quality product, the entry effect on product revenue is negative for the n-1 quality product, ambiguous for the n quality product, and zero for the product whose quality is below n-1.

2.4.4.2 Relationship between Physical Productivity and Revenue Productivity

In this section, based on the theoretical analysis above, I examine how physical productivity is related to the revenue productivity that is frequently used as conventional measures of productivity. In so doing, specify a change in ω_j as physical productivity improvement, and a change in revenue R_j as revenue productivity improvement.

I first decompose the entry effect on product supply into direct entry effects through market share change and indirect entry effect through productivity improvement. Notice that the equation (2-21) can be rewritten as

$$\begin{cases} \Delta\hat{z}_{n-1} = 0 + \frac{\omega_n' - \omega_n}{2} < 0 \\ \Delta\hat{z}_n = \frac{(\omega_{n+1}' - 1)}{2} + \frac{(\omega_n - \omega_n')}{2} \\ \Delta\hat{z}_{-n,-n-1} = 0 + 0 \end{cases} . \tag{2-23}$$

The first term of the right hands captures the supply changes when physical productivity is unaffected by the introduction of the n+1 quality product (See equation (2-19)). Since I assume the limited productivity improvement, the second term captures the entry effect on product supply arising only from the productivity improvement of the n quality product. That is, the first term explains the direct entry effect on supply, and the second is the indirect entry effect through physical productivity improvement on supply.

As showed above, there is a negative entry effect on the supply of n-1 quality product. The direct entry effect on the supply is zero because the market share of n-1 quality product is not affected. But the indirect entry effect on its supply is negative because its market share is affected by the productivity improvement of the n quality product.

The entry effect on the n quality product is more complicated. On the one hand, the direct entry effect of market share change is negative. On the other hand, the indirect entry effect arising from productivity improvement on product supply is positive.

Proposition 2-3. *When new product only impacts the physical productivity of the ex-highest quality product, direct effects (i.e., business stealing effects) are zero for the products whose quality is n-1 or below, and is negative for the n-quality product. Indirect effects (i.e., limited productivity improvement) are zero for the products whose quality is n-2 or below, negative for the n-1 quality product, and positive for the n-quality product.*

Similarly, the entry effect on product revenues can also be decomposed into direct entry effects through market share change and indirect entry effect through limited productivity improvement. The equation (2-22) can be rewritten as

$$\begin{cases} \Delta\widehat{R}_{n-1} = 0 + \frac{(\omega_n{}' - \omega_n)(1+\omega_{n-1})}{2} q_{n-1} < 0 \\ \Delta\widehat{R}_n = \frac{(\omega_{n+1}{}' - 1)(1+\omega_n)}{2} q_n + \frac{(\omega_n - \omega_n{}')[(1-\omega_{n+1}{}') + (\omega_n + \omega_n{}')]}{2} q_n \\ \Delta\widehat{R}_{-n,-n-1} = 0 + 0 \end{cases} , \qquad (2\text{-}24)$$

where the first term of the right hand side is the revenue changes when productivity is held constant. The second term expresses the indirect entry effect on product revenue only arising from limited productivity improvement.

The first term of equation (2-24) is either 0 or negative. The revenue will remain constant or decrease when physical productivity is held constant. This suggests that a change in the revenue productivity $\widehat{R}_n$ can be negative even if there is no change in physical productivity ω_n.

Now, consider the second term of the indirect entry effect. For the n-1 quality product, the second term of equation (2-24) is negative since I assumed $\omega_n{}' - \omega_n < 0$. This means the productivity improvement of the n quality product leads to a decrease in the revenue of the n-1 quality product. For the n quality product, a change in its revenue is not clear. The productivity improvement increases its supply and thus will raise the revenue, but its market share can also decrease.

Proposition 2-4. *When new product only impacts the physical productivity of the ex-highest quality product, direct effects on a change in revenue (i.e., business stealing effects) are zero for the products whose quality is n-1 or below, and negative for the n-quality product. Indirect effects on a change in revenue (i.e., limited productivity improvement) are zero for the products whose quality is n-2 or below, negative for the n-1 quality product, and positive for the n-quality product.*

The equation (2-24) also indicates that the revenue productivity measures the physical productivity precisely for the products whose quality is n-2 or below because a change in their revenues coincides with a change in their physical productivity. For the n-1 quality product, however, revenue productivity underestimates a change in its physical productivity. For the n quality product, it is possible that revenue productivity underestimates or overestimates a change in its physical productivity.

2.4.4.3 Identification of Change in Physical Productivity

In this section, I propose a simple and novel approach of estimating physical productivity changes by using quantity data only about the existing products based on the model outlined above. Equation (2-21) suggests that a change in the supply of the quality n product can be rewritten as

$$\Delta\hat{z}_n = \frac{(\omega_{n+1}{}' - 1)}{2} - \frac{(\omega_n{}' - \omega_n)}{2} \ .$$

Notice that $\frac{1-\omega_{n+1}{}'}{2}$ is the supply of the quality n+1 product $\Delta\hat{z}_{n+1}$. A physical productivity change of the n quality product becomes

$$\Delta\omega_n = \omega_n{}' - \omega_n = -2\Delta\hat{z}_{n+1} - 2\Delta\hat{z}_n \ . \tag{2-25}$$

It follows from equation (2-21) that a change in the supply of the quality n-1 is given by

$$\Delta\hat{z}_{n-1} = \frac{\omega_n{}' - \omega_n}{2} \ .$$

The physical productivity change of the n quality product can be also expressed as

$$\Delta\omega_n = -2\Delta\hat{z}_{n+1} - 2\Delta\hat{z}_n = 2\Delta\hat{z}_{n-1} \ .$$

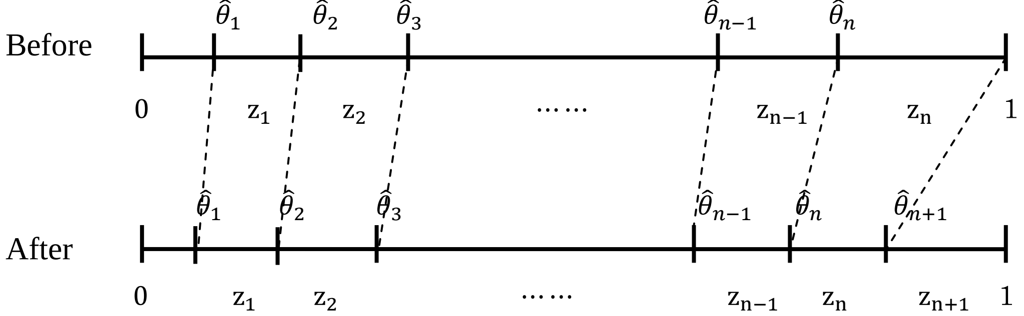
Therefore, I can measure a physical productivity change if the data about quantity of the quality n-1 product are available before and after the entry takes place.

Proposition 2-5. *When new product only impacts the physical productivity of the ex-highest quality product, this physical productivity improvement is proportional to the supply change of the n-1 quality product.*

2.4.5 Equilibrium with Productivity Improvement

In this section, I relax the limited productivity improvement assumption by assuming that the market introduction of the new highest quality improves all the existing products' physical productivity through spillover. Figure 2.5 depicts the relationship between θ and z before and after introducing the n+1 quality product when this product introduction improves all the existing products' physical productivity.

Figure 2.5



2.4.5.1 Equilibrium with Productivity Improvement

Assume that the market introduction of the n+1 quality product can improve all the existing product's physical productivity, and that the inequalities $\Delta\omega_j = \omega_j{}' - \omega_j < 0$ and $|\Delta\omega_{j+1}| > |\Delta\omega_j|$ hold. These inequalities imply that the productivity improvement effect becomes weaker as the quality level goes down from n, n-1, to 1. I first show how such an entry influences the equilibrium supply and revenue of existing products.

It follows from equations (2-14) and (2-17) that changes in the equilibrium supply of all the existing products are given by

$$\Delta\hat{z}_j = z_j^{*\prime} - z_j^{*} = \begin{cases} \frac{(\omega_{j+1}{}' - \omega_{j+1}) - (\omega_j{}' - \omega_j)}{2} & \text{if } j = 1, \dots, n-1 \\ \frac{(\omega_{j+1}{}' - 1) - (\omega_j{}' - \omega_j)}{2} & \text{if } j = n \end{cases} . \qquad \text{(2-26)}$$

The equations show that a change in the supply of the quality j product depends on the difference in the extent of productivity improvement between itself and the j+1 quality product.

I obtain $\Delta\hat{z}_j = \frac{(\omega_{j+1}{}' - \omega_{j+1}) - (\omega_j{}' - \omega_j)}{2} < 0$ for j=1,…, n-1 under the assumption that $|\Delta\omega_{j+1}| > |\Delta\omega_j|$. In other words, the supply of the product whose quality is n-1 or below will decrease, and the magnitude of this supply change becomes smaller as the quality level lowers. The improvement on physical productivity causes its price to fall, and this price fall is greater as the quality is higher. Consumers who are willing to buy an upgrade product are more than the consumers who shift from a downgrade product and as a result, the supply of these products decreases.

For the good with quality q_0, the equilibrium supply change $\Delta\hat{z}_0 = \frac{(\omega_1{}' - \omega_1) - (\omega_0{}' - \omega_0)}{2} < 0$ can be calculated from the total demand before and after the new higher quality entry in a similar way. Notice that the firms do not produce this q_0

quality goods, and the consumers with taste parameter $\hat{\theta}_0$ do not buy any types of product. Therefore, the equilibrium supply change $\Delta\hat{z}_0$ implies the change on the amount of consumers who do not buy at all. The physical productivity improvement leads to a decrease in the price of q_1 quality product, and then some consumers who did not purchase any product are now willing to buy the q_1 quality product. Therefore, the amount of consumers who choose to buy nothing becomes smaller, and this also means more consumers choose to purchase product when the new entry can improve all existing products' physical productivity.

However, the supply of the n quality product may increase or decrease, depending on the magnitude of two opposing forces. The productivity improvement lowers its price, and this leads to an expansion of its supply by stealing a market share of the lower quality products but some consumers now switch to the n+1 quality product.

Next, I examine the entry effect on product revenues. Using the revenue equations (2-15) and (2-18), changes in revenue from each product can be written as

$$\Delta\widehat{R_j} = R_j^{*\prime} - R_j^* = \begin{cases} \frac{(\omega_{j+1}{}' - \omega_j{}')(1+\omega_j{}') - (\omega_{j-1} - \omega_j)(1+\omega_j)}{2} q_j & \text{if } j = 1, \dots, n-1 \\ \frac{(\omega_{j+1}{}' - \omega_j{}')(1+\omega_j{}') - (1-\omega_j)(1+\omega_j)}{2} q_j & \text{if } j = n \end{cases} . \quad (2\text{-}27)$$

Equation (2-27) shows that the effects on revenues are the same as those on the supply of products.

2.4.5.2 Relationship between Physical Productivity and Revenue Productivity

In this section, I will do the same decomposition exercise as in Section 2.4.4.2, but it is assumed this time that the market introduction of a new product has positive impacts on all the existing products from quality 1 to quality n.

Note that the equation (2-26) can be rewritten as

$$\Delta\hat{z}_j = \begin{cases} 0 + \frac{(\omega_{j+1}{}' - \omega_{j+1}) - (\omega_j{}' - \omega_j)}{2} < 0 & \text{if } j = 1, \dots, n-1 \\ \frac{(\omega_{j+1}{}' - 1)}{2} + \frac{(\omega_j - \omega_j{}')}{2} & \text{if } j = n \end{cases}. \quad (2\text{-}28)$$

Again, the first term on the right hand side of equation (2-28) is the supply changes when productivity change is held constant, and the direct entry effect on a change in supply (i.e., business stealing effects). The second term captures the indirect entry effect on supply arising from physical productivity improvement.

For the product whose quality is n-1 or below, the direct entry effect on supply is zero. On the other hand, the indirect effect on supply is negative and becomes weaker as the quality level goes down along with quality ladder because of the assumption that $|\Delta\omega_{j+1}| > |\Delta\omega_j|$. In other words, the n+1 quality product cannot attract consumers who previously bought such products unless there is physical product improvement, but it increases physical productivity of the existing products and its increment becomes larger for higher quality products. As a result, the product whose quality is below n loses its market share.

The direct effect on the n quality product is $\frac{(\omega_{j+1}{}' - 1)}{2}$ and negative whereas the indirect effect is $\frac{(\omega_j - \omega_j{}')}{2}$ and positive. The overall effect depends on the magnitude of these two terms and thus is uncertain. Intuitively, the new product with quality n+1 reduces the market share for the n quality product because a part of consumers are willing to switch to the new product. On the other hand, physical productivity improvement reduces production cost and leads to expand its market share by stealing a market share of the lower quality products.

Proposition 2-6. *When entry impacts all products' physical productivity, direct effects on a change in supply (i.e., business stealing effects) are zero for the products whose* *quality is n-1 or below, and negative for the n-quality product. Indirect effects on a change in supply (i.e., productivity improvement) are negative for the products whose quality is n-1 or below and the magnitude becomes smaller as the quality ladder goes down from n-1 to 1. They are positive for the n-quality product.*

Similarly, the entry effect on firm revenue from existing products can also be decomposed into direct substitution effect and indirect productivity improvement effect. Rewrite the equation (2-27) as

$$\Delta\widehat{R}_j = \begin{cases} 0 + \frac{(\omega_{j+1}' - \omega_{j+1}) - (\omega_j' - \omega_j) + (\omega_{j+1}' - \omega_j')\omega_j' - (\omega_{j+1} - \omega_j)\omega_j}{2} q_j < 0 & \text{if } j = 1, \dots, n-1 \\ \frac{(\omega_{j+1}' - 1)(1 + \omega_j)}{2} q_j + \frac{(\omega_j - \omega_j')[(1 - \omega_{j+1}') + (\omega_j + \omega_j')]}{2} q_j & \text{if } j = n \end{cases} . \quad (2\text{-}29)$$

The logic behind this result is the same as the one I discussed above about a change in the supply of different quality products.

Proposition 2-7. *When entry impacts all products' physical productivity, direct effects on a change in revenue (i.e., business stealing effects) are zero for the products whose quality is n-1 or below, and negative for the n-quality product. Indirect effects on a change in revenue (i.e., productivity improvement) are negative for the products whose quality is n-1 or below and the magnitude becomes smaller as the quality ladder goes down from n-1 to 1. They are positive for the n-quality product.*

2.4.5.3 Identification of Change in Physical Productivity

In this section, I explain how to estimate physical productivity changes by using quantity data only about the existing products when entry impacts productivity of all the existing products. The method I explain below is fundamentally the same as the one I discussed in Section 2.4.4.3.

It follows from equation (2-26) that the supply change of the quality n product is given by

$$\Delta\hat{z}_n = \frac{(\omega_{n+1}{}' - 1)}{2} - \frac{(\omega_n{}' - \omega_n)}{2} \ .$$

The physical productivity change of the quality n product is

$$\Delta\omega_n = \omega_n{}' - \omega_n = -2\Delta\hat{z}_{n+1} - 2\Delta\hat{z}_n \ . \qquad (2\text{-}30)$$

Using the equation (2-26), I have

$$\Delta\hat{z}_{n-1} = \frac{(\omega_n{}' - \omega_n)}{2} - \frac{(\omega_{n-1}{}' - \omega_{n-1})}{2} \ .$$

Combining this equation with equation (2-30), I obtain the physical productivity change of the n-1 quality product as

$$\Delta\omega_{n-1} = \omega_{n-1}{}' - \omega_{n-1} = -2\Delta\hat{z}_{n+1} - 2\Delta\hat{z}_n - 2\Delta\hat{z}_{n-1} \ .$$

Similarly, I can apply the same procedure to obtain the physical productivity change of the product with quality level n-2, n-3, and so on. Finally, I obtain the physical productivity change of the j quality product as

$$\Delta\omega_j = \omega_j{}' - \omega_j = -2\sum_{k=j}^{n+1} \Delta\hat{z}_k \ . \qquad (2\text{-}31)$$

Thus, the physical productivity change of all existing products can be estimated using this equation when quantity data are available,

Proposition 2-8. *When entry can impact all products' physical productivity, the physical productivity improvement is proportional to the supply change of itself and above.*

2.5 Result Summary

The theoretical model connects the new entry, and fundamental parameters of both firms and consumers, and then allows me to separate the technical improvement effect of entry from business stealing effect. I showed the changes in product supply and revenue after an entry and decompose it into terms arising from direct product substitution and arising from indirect productivity improvement in section 2.4. Based on the insights from the theoretical analysis explained above, this section summarizes the results of how a higher quality product entry affects supply of existing products and firm revenue, as well as the relationship between the revenue productivity change and physical productivity improvement.

2.5.1 Direct Entry Effect

The direct entry effect is an effect arising from substitution of higher quality product (i.e., business stealing effect) and it reflects an impact from demand side. When a new higher quality product is introduced to a market, a part of consumers are willing to buy the new product instead of existing products, and this leads to a loss of market share for existing products. Firms then have to adjust its supply in order to respond to this demand reduction, and finally change its revenue. I conclude the direct entry effect on product supply and firm revenue in the following proposition.

Proposition 2-9. *The entry effect on supply and revenue arising from business stealing effects is negative for the ex-highest quality product, and is zero for other existing products*.

The direct entry effect exists even in the absence of productivity improvement of existing product. This is exactly the case of entry effect under constant productivity assumption as explained in section 2.4.3. In the absence of physical productivity improvement, introducing a new higher quality product can only reduce the supply and revenue of ex-highest quality product and do not affect those of other existing products.

When a new product with a higher quality is marketed, a part of consumers who choose the ex-highest quality product are willing to switch to the new product. This results in a loss of market share, and finally reduces the supply and revenue of this ex-highest quality product. However, the new higher quality product cannot attract consumers who choose quality level under ex-highest quality, and cannot steal a market share from those products as well as affect revenue of those products. Therefore, when facing new higher quality competitors, the ex-highest quality product will lose its market share and revenue unless it improves its productivity.

2.5.2 Indirect Entry Effect

The indirect entry effect explains an effect arising only from physical productivity improvement. When a new higher quality product is intoroduced to a market, it may also bring technical improvement to existing products. This productivity improvement may reduce the production cost of existing products, and may lead to an expansion of market share as well as an increase in firm revenue. I summarize this indirect entry effect on product supply and firm revenue as the following proposition.

Proposition 2-10. *The entry effect on supply and revenue arising from productivity improvement is positive for the ex-highest quality product, and is 0 or negative for other existing products.*

If the productivity of existing products has improved after the new higher quality product enters, such productivity improvement expands the supply and increases the revenue of the ex-highest quality product. The physical productivity improvement can reduces its production costs and leads to an increase in demand by stealing a market share from its downgrade quality products. On the other hand, the other existing products will lose their market share and revenue as long as their upgrade product's productivity improved. Productivity improvement of an existing product can lower its production cost and expand its demand by taking the market share away from its downgrade product, but its market share can also be taken away by its upgrade product. As a result, the supply and revenue of an existing product under the ex-highest quality may decrease or increase depending on the magnitudes of these two opposite effects. Intuitively, the effect of entry on technical improvement becomes weaker as the quality level far from the new product's quality level. That is, the productivity improves more as the quality goes up. Therefore, for each existing product except the ex-highest quality product, consumers who switch to its upgrade product are more than the consumers who shift from its downgrade product, and results in reductions of supply as well as revenue.

Therefore, when a new higher quality product has spillover effects on the physical productivity of all products, only the ex-highest quality product can expand demand as well as raise the revenue. Moreover, although the ex-highest quality product suffers from a negative business stealing effect of entry on revenue, it is the only product that has a change to obtain a revenue gain in total.

2.5.3 Relationship between Revenue Productivity and Physical Productivity

The theoretical framework of this study allows me to decompose the entry effect on revenue change into direct effect and the indirect effect arising from productivity improvement. Based on this decomposition, I explicitly show a link between the revenue productivity change and physical productivity and find the relationship between them as following:

Proposition 2-11. *Revenue productivity captures a change arising from both demand factor and physical productivity improvement, and it varies even if physical productivity is kept constant.*

Under the physical productivity constant assumption in section 2.4.3, the firm revenue of the ex-highest quality product changes because the effects of a loss in a marker share outweigh the effects of its physical productivity improvement. This implies that when a new product with higher quality is introduced to a market, revenue productivity of the ex-highest quality product varies even though there is no physical productivity change, and this implies that the revenue productivity measurement may fail to capture such a change in the real physical productivity.

When a new product can only impact a part of existing product’s physical productivity, as I showed in section 2.4.4, such productivity improvement affects not only the revenue of productivity improved product, but also the revenue of its downgrade products. That is, the revenue of an existing product may be affected a change in a physical productivity improvement of its upgrade products, rather than a physical productivity improvement of itself. As a result, even if its physical productivity remains unchanged, a product’s revenue productivity may change because of other product’s productivity improvement, and the revenue productivity measurement may fail to capture a physical productivity change arising only from itself.

When new product can impact all existing products’ physical productivity, the analysis in section 2.4.5 showed that the revenue change of existing product relies not only on the demand side factor such as business stealing effect, but also on the physical productivity improvement of itself and other products. The physical productivity improvement of itself helps to reduce production costs and leads to a revenue gain, but other product's physical productivity improvement reduces its market share and leads to a revenue loss. Since the revenue productivity depends on the magnitude of these two forces, it can be lower or even have an opposite direction comparing with physical productivity. The revenue productivity measurement captures a mixture of business stealing effect and physical productivity effect, and thus cannot isolate its real physical productivity change from other factors.

Therefore, when a new product is introduced to a market, the revenue productivity improvement captures a mixture of changes steming from both business stealing effect and technical productivity effect. Then the real physical productivity improvement can be underestimated if we use the revenue productivity approach. I also showed that the real physical productivity improvement is proportional to the supply change of itself and goods with higher quality, and can be identified precisely by using the supply change data.

2.6 Conclusion of Chapter 2

In this analysis, I built a theoretical model that includes competition, consumption and production in order to examine the entry effect on firm productivity. Although many extant studies have examined how new entry affects product quality choices, this chapter examined how new entry by offering the highest quality product affects firm's measured productivity by taking both demand and supply factors into account. Based on the insights from the theoretical model, I decompose the effect of offering a new product on supply and revenue into technical improvement effect and other demand factors, and this allows me to separate physical productivity from revenue productivity cleanly.

It is important to distinguish physical productivity and revenue productivity when examining entry effect on productivity of incumbent firms and industry level. The revenue productivity measurement tends to capture a change arising from both technical improvement and demand side factors, and then cannot distinguish whether an observed productivity change is due to technical improvement or a change in market structure. This can be very problematic when we are interested in comparing technical efficiency of new entrants and incumbent firms or examining the spillover effects of new entrants on incumbent firms. The separation of revenue productivity from physical productivity allows us to identify the determination of an observed productivity change accurately, and then helps to understand fundamental economic issues about entry effect such as spillover effect clearly.

Chapter 3.

Resources Misallocation and Regional Productivity Difference

3.1 Introduction of Chapter 3

Why are some regions richer than others? Why some regions that were poor in the past are now getting richer? Empirical research on economic growth started using firm-level micro-data to advance our understanding of sources of economic growth. Among many important factors for economic growth, resource misallocation has recently attracted a lot of attention from economic researchers (e.g., Hsieh and Klenow 2009, Midrigan and Xu 2014). One of the stylized facts from this line of research is that developed countries like the United States of America attain a higher level of the aggregate productivity because of more efficient resource allocation than developing countries like China and India. The resource misallocation literature has mainly focused on differences in productivity across countries, but there are not many studies that empirically investigate regional differences in productivity within a country arising from resource misallocation.

In this chapter, I use the Chinese manufacturing data to examine regional productivity differences due to resource misallocation. In particular, I estimate productivity losses from resource misallocation as well as potential gains from efficient resource allocation, and closely examine sources of productivity differences in two Chinese major cities, Beijing and Shanghai. The method of estimating productivity loss has been proposed by Hsieh and Klenow (2009) and is easily implementable, but it assumes constant returns to scale. However, it has been widely known that production technology in China does not exhibit constant returns to scale. It may sound innocuous to maintain the constant returns to scale assumption, but Gong and Hu (2016) raise the concern that resource misallocation tends to be over-estimated when production technology is decreasing or increasing returns to scale. This is because variation in physical productivity across firms is added to the misallocation measure when production technology does not exhibit constant returns to scale. In the estimation of this study, I estimate productivity losses by dropping the constant returns to scale assumption in addition to using the estimation method of Hsieh and Klenow (2009). I also decompose productivity losses into the one arising from capital misallocation and the one from labor misallocation. This decomposition helps us to understand which resource misallocation needs to be corrected in order to improve an aggregate productivity. Moreover, in order to examine the reliability of productivity loss from HK approach, I construct an alternative approach for measuring such aggregate TFP losses arising from resource misallocation based on the model in Midrigan and Xu (2014). Their original model focused on how capital misallocation impact aggregate TFP, but I add labor misallocation in this model and also provide simple estimation methodology when firm-level data is available.

The main findings from this study are summarized as follows. First, there is a difference in productivity losses between Beijing and Shanghai. The average aggregate productivity loss of Beijing from resource misallocation is larger than that of Shanghai, implying that resource allocation is more efficient in Shanghai than in Beijing. There is indeed a regional productivity difference due to resource misallocation. Second, the estimation result of this study indicates that the aggregate productivity level of Beijing will significantly improve when resources are allocated to the efficient level. While

Hsieh and Klenow (2009) estimate the productivity gain from correcting resource misallocation to be around 100 for the whole Chinese economy, the average estimated productivity gain of this study is 113 percent for Beijing and 67 percent for Shanghai. The productivity gain of Beijing is similar to the national average in magnitude whereas the productivity gain of Shanghai is much lower than the national average. Again, this result indicates that there is variation in the degree of resource misallocation across regions. Finally, the estimation result of this study shows that capital misallocation is more severe in Beijing than labor misallocation. This indicates that some frictions exist in Beijing's capital markets.

The rest of chapter is organized as follows. In section 3.2, I briefly review the literature about resource misallocation effect on productivity by focusing on empirical findings relevant to this research. In section 3.3, I outline the model proposed by Hsieh and Klenow (2009) and the extended model that relaxes the constant returns to scale assumption. In section 3.4, I present an alternative approach of measuring productivity losses based on Midrigan and Xu (2014). This approach takes dynamic aspects of firm decisions into account to capture the labor misallocation effect on aggregate TFP. Section 3.5 describes the data and construction of research variables. In section 3.6, I present the estimation results of TFP and resources misallocation, and analyze their differences between Beijing and Shanghai. Section 3.7 concludes.

3.2 Brief Literature Review

As the growth of input resources slows down, more efficient resource allocations become critical to develop the economic growth in China. Product and input resource market distortions liberalized in the process of economic transformation have promoted the aggregate productivity growth to some extent, for example, the migration of labor from lower productive region to higher productive region. However, there remain some barriers that hinder the further population movement across regions or even within a region, such as the Hukou system. Past studies have attributed the remaining resources distortions to such factors as interregional trade barriers (Holz, 2009), capital market (Boyreau-Debray and Wei, 2004) local protectionism (Young, 2000) and size restriction (Guner et al., 2008).

All these market impediments in China spread the dispersions in labor and capital productivity across firms, and therefore results in lower levels of firm output and finally show up in the lower aggregate productivity in a region. In this chapter, I examine how the input resource distortions affect the aggregate productivity and whether a different pattern of input resource distortions leads to a gap in aggregate productivity. This study helps identify the sources and the extent of resource market distortions and suggest policy directions to improve economic efficiency.

Many literatures have stressed the importance of input market allocation across firms as a determinant of aggregate productivity. For example, Restuccia and Rogerson (2008) considered a version of the neoclassical growth model where firms face idiosyncratic policy distortions. The policy they considered can be levying taxes or providing subsidies to output or the capital and labor. Policy distortions are found to have considerable effects on the aggregate output and the measured TFP for the U.S. Hsieh and Klenow (2009) constructed a standard model of monopolistic competition to demonstrate that idiosyncratic distortions in capital and labor markets lower aggregate TFP. They estimated resource distortions from the residuals in first-order conditions following Chari, Kehoe and McGrattan (2007), and provide quantitative evidence of the impact of resource misallocation on aggregate manufacturing TFP in China, India and the U.S. Amaral and Quintin (2010) have examined the relationship of financial constraints and aggregate TFP, and they suggest that firms cannot optimize the amount of capital when they face financial constraints and this reduces aggregate TFP. Eslava et al. (2013) have studied the impact of trade liberalization on input market allocation as well as on labor productivity by using Colombian data, and they suggest that trade liberalization is a source of input market misallocation and labor productivity change in Columbia.

Hsieh and Klenow (2009) made a rigid assumption of constant returns to scale production function in their monopolistic competition model. However, this assumption is debatable, because direct estimates or related studies of firm-level value-added production functions such as Atkeson, Khan, and Ohanian (1996), Pavcnik (2002) and Guner, Ventura, and Xu (2008), all point to a decreasing returns production function. The production function estimation from this study also results in a decrease returns production function, which is consistent with those literatures. In fact, the assumption of constant returns to scale production function can lead to a substantial misestimate in market distortions and TFP when the firms' production function do not exhibit constant returns in data.

3.3 HK Approach Extension to Measure the Misallocation

In this section I present the model in which a link between aggregate productivity and resource misallocation arises from the existence of distortions regarding the optimal allocation of input factors at the firm-level. The basic framework developed in Hsieh and Klenow (2009) (HK hereafter) is adopted, but I also outline the extend model proposed by Gong and Hu (2016) who relax the assumption of constant returns to scale. Many studies have also showed that the output elasticities of inputs of Chinese firms are typically different from those of U.S. firms and the production function is not likely to exhibit constant returns to scale in most Chinese firms. For example, Feenstra et al. (2014) estimated Chinese manufacturing factor elasticities and showed that the average output elasticity of capital is 0.278 and the average output elasticity of labor is 0.399. This extension is thus crucial for empirical analysis of this chapter because I am able to estimate resource misallocation in China more precisely. The purpose of this section is to set up a framework that guides my empirical analysis, and I do not intend to make any original contributions here.

3.3.1 Theoretical Framework

Assume that an economy has a single final good Y produced by a representative firm in a perfectly competitive final output market. This firm combines the output Ys of S industries in the economy using a Cobb-Douglas production technology:

$$Y = \prod_{S=1}^{S} Y_S^{\theta_S} \text{ , where } \sum_{s=1}^{s} \theta_S = 1 \text{ .} \tag{3-1}$$

Cost minimization implies that industry shares θ_S are given by

$$\theta_S = \frac{P_S Y_S}{PY} \quad , \quad \text{where } P \equiv \prod_{S=1}^{S} \left(\frac{P_S}{\theta_S}\right)^{\theta_S} \quad , \tag{3-2}$$

where P_S is the price of industry gross output Y_S, and P is the price of the final good. The final good Y is assumed to be the numeraire so that I can set P = 1.

In each industry, gross output Y_S is a CES aggregate of M_S numbers of differentiated products and is expressed as

$$Y_S = \left(\sum_{i=1}^{M_S} Y_{Si}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} \quad , \tag{3-3}$$

where Y_{Si} represents the gross output of firm i and σ measures the elasticity of substitution between varieties of differentiated goods. The assumptions of free entry and monopolistic competition at the industry level imply the inverse demand equations for each product produced by firm i are given by

$$\widehat{P}_{Si} = P_S \left(\frac{Y_S}{Y_{Si}}\right)^{\frac{1}{\sigma}} .$$

Firms differ in their efficiency levels, and I also assume that firms face different output and input prices. At the firm level, the production function for each differentiated product is given by a Cobb-Douglas function:

$$Y_{Si} = A_{Si} K_{Si}^{\alpha_S} L_{Si}^{\beta_S} , \qquad (3\text{-}4)$$

where A_{Si}, K_{Si} and L_{Si} represent firm i's total factor productivity, capital stock, labor inputs, respectively. The parameters α_s and β_s are the output elasticities of capital and labor. I assume that the sum of α_s and β_s is not necessary to be 1. That is, the firm i can have increasing, decreasing or constant returns to scale of production technology. Moreover, note that factor shares can vary across industries but not across firms within the same industry.

Since there are two factors of production, it is possible that firms face two types of distortions: a capital distortion and a labor distortion. I denote the distortions that increase the marginal product of capital relative to an output distortion that changes the marginal product of capital and labor by the same proportion as the capital distortion $\tau_{K,Si}$. For example, $\tau_{K,Si}$ will be low if firms are able to access to cheap credit, but will be high if firms cannot use such credit and then have to pay a high interest. Similarly, let $\tau_{L,Si}$ denote labor distortions that increase the marginal product of labor relative to output distortion. For example, $\tau_{L,Si}$ is expected to be low for firms that benefit from government subsidies on employment, however to be high for that firms that do not have right to use these subsidies and face higher labor costs. Combining these definitions of distortions and production function (3-4), firm's profit maximum problems are given by

$$\max\{\pi_{Si} = P_{Si}Y_{Si} - (1+\tau_{L,Si})\omega L_{Si} - (1+\tau_{K,Si})RK_{Si}\}\ , \qquad (3\text{-}5)$$

where ω and R stand for the price of labor input and capital input respectively. If the distortions do not exist, all firms will face the same wage and capital input price. Note that equation (3-5) expresses the distortions in terms of capital and labor relative to the output distortion. Then, in the model, an output distortion will show up as a lower capital and labor distortions.

The first order condition from the firm's profit maximization implies

$$P_{Si} = \left(P_S Y_S^{\frac{1}{\sigma}}\right)^{\frac{\sigma\varphi_S}{\sigma-1}\cdot[1-(\alpha_S+\beta_S)]} \cdot \left[\left(\frac{\sigma}{\sigma-1}\right)^{(\alpha_S+\beta_S)}\left(\frac{\alpha_S}{R}\right)^{\alpha_S}\left(\frac{\beta_S}{\omega}\right)^{\beta_S} A_{Si}(1+\tau_{K,Si})^{-\alpha_S}(1+\tau_{L,Si})^{-\beta_S}\right]^{\frac{-\varphi_S}{\sigma-1}}, \qquad (3\text{-}6)$$

where $\varphi_s = \left[\frac{\sigma}{\sigma-1} - (\alpha_S+\beta_S)\right]^{-1}$.

I can also get the capital-labor ratio, labor allocation, capital allocation and output respectively:

$$\frac{K_{Si}}{L_{Si}} = \frac{\alpha_S}{\beta_S}\cdot\frac{\omega}{R}\cdot\frac{1+\tau_{L,Si}}{1+\tau_{K,Si}}\ ; \qquad (3\text{-}7)$$

$$K_{Si} \propto \frac{1}{1+\tau_{K,Si}}\left[A_{Si}\cdot(1+\tau_{K,Si})^{-\alpha_S}\cdot(1+\tau_{L,Si})^{-\beta_S}\right]^{\varphi_S}\ ; \qquad (3\text{-}8)$$

$$L_{Si} \propto \frac{1}{1+\tau_{L,Si}}\left[A_{Si}\cdot(1+\tau_{K,Si})^{-\alpha_S}\cdot(1+\tau_{L,Si})^{-\beta_S}\right]^{\varphi_S}\ ; \qquad (3\text{-}9)$$

$$Y_{Si} \propto \left[A_{Si}\cdot(1+\tau_{K,Si})^{-\alpha_S}\cdot(1+\tau_{L,Si})^{-\beta_S}\right]^{\frac{\sigma\varphi_S}{\sigma-1}}\ . \qquad (3\text{-}10)$$

It follows from equations (3-8) and (3-9) that allocation of resources across firms not only depends on firm's total factor productivity A_{Si}, but also on the capital and labor distortions each firm faces. However, the equation (3-7) shows that the capital-labor ratio is only affected by the extent of distortion of capital and labor, and is irrelevant with firm's production efficiency. Inefficient resources allocation will leads to distortions that can result in differences in the marginal revenue product of capital and labor across firms. The marginal revenue product of labor and capital for each firm i can be obtained by differentiating the revenue function $P_{Si}Y_{Si}$:

$$\mathrm{MRPL}_{Si} = \frac{\partial P_{Si} Y_{Si}}{\partial L_{Si}} = \omega \cdot \left(1 + \tau_{L,Si}\right) = \frac{\sigma-1}{\sigma} \cdot \frac{\beta_S P_{Si} Y_{Si}}{L_{Si}} \quad ; \tag{3-11}$$

$$\mathrm{MRPK}_{Si} = \frac{\partial P_{Si} Y_{Si}}{\partial K_{Si}} = R \cdot \left(1 + \tau_{K,Si}\right) = \frac{\sigma-1}{\sigma} \cdot \frac{\alpha_S P_{Si} Y_{Si}}{K_{Si}} \quad . \tag{3-12}$$

Equation (3-11) shows that the marginal revenue product of labor is a function of labor distortion and wage rate ω that is the labor input price without resource distortions. Equation (3-12) similarly shows the marginal revenue product of capital is a function of capital distortion and capital input price R under efficient resources allocation in input markets. Moreover, under the assumptions, the after-tax marginal revenue products of capital and labor are equalized across firms, which means that the before tax marginal revenue product of capital and labor must be high for firms with some friction of access to input markets and be low for firms with frictionless access to input markets.

In order to derive the relationship between the aggregate total factor productivity and resource misallocation, the equilibrium allocation of resources across industries is solved and it is given by

$$L_S \equiv \sum_{i=1}^{M_s} L_{Si} = L \cdot \frac{\beta_S \theta_S / \overline{\mathrm{MRPL}_S}}{\sum_{S=1}^{S} \beta_S \theta_S / \overline{\mathrm{MRPL}_S}} \quad , \tag{3-13}$$

$$K_S \equiv \sum_{i=1}^{M_s} K_{Si} = K \cdot \frac{\alpha_S \theta_S / \overline{\mathrm{MRPK}_S}}{\sum_{S=1}^{S} \alpha_S \theta_S / \overline{\mathrm{MRPK}_S}} \quad , \tag{3-14}$$

where the $\overline{\mathrm{MRPL}_S}$ and $\overline{\mathrm{MRPK}_S}$ are the weighted average of the value of the marginal product of labor and capital in industry S respectively, and can be expressed as

$$\overline{\mathrm{MRPL}_S} = \omega \cdot \left(\sum_{i=1}^{M_S} \frac{1}{1+\tau_{L,Si}} \cdot \frac{P_{Si} Y_{Si}}{P_S Y_S}\right)^{-1} \quad , \tag{3-15}$$

$$\overline{\mathrm{MRPK}_S} = R \cdot \left(\sum_{i=1}^{M_S} \frac{1}{1+\tau_{K,Si}} \cdot \frac{P_{Si} Y_{Si}}{P_S Y_S}\right)^{-1} \quad . \tag{3-16}$$

L_S and K_S represent the equilibrium labor input and the capital input in industry S. $L \equiv \sum_{s=1}^{s} L_S$ and $K \equiv \sum_{s=1}^{s} K_S$ are the aggregate supply of labor and capital in an economy. Because the aggregate supply of labor and capital in an economy is constant, we can see from equations (3-13) and (3-14) that if the average values of the marginal product of labor and capital in industry S do not change, the equilibrium labor input and the capital input will not change. I can then aggregate output as a function of L_S and K_S and total factor productivity A_S:

$$Y = \prod_{S=1}^{S} \left(A_S \cdot K_S^{\alpha_S} \cdot L_S^{\beta_S} \right)^{\theta_S} \quad .$$

3.3.2 Relationship between Resources Misallocation and Aggregate TFP Loss

So far the derivation process is similar to Hsieh & Klenow (2009), except relaxing the assumption of constant returns to scale. In their paper, in order to measure the output distortion, they introduce the definition of total factor quantity productivity (TFPQ) and total factor revenue productivity (TFPR). TFPQ means how many units of output can be produced by a firm from using one unit of mix of input factors, and can be calculation by using a firm-specific price deflator. TFPR explains how much revenue can be obtained from the same amount of input factors, and can be obtained by using an industry deflator. Hsieh & Klenow (2009) use the variance of TFPR as a measure of resources misallocation across firms, because the TFPR does not vary across firms within the same industry unless firms face distortions in their model. That is, when production function has constant returns to scale, the TFPR will not vary across firms if the prices of labor input and capital input are equal within the industry, and the TFPR will be different if the input prices vary result from the resources distortions. Therefore, they stressed that the variance of TFPR can be used to measure the extent of resource distortions, and showed that the industry aggregate TFP becomes lower as the dispersion in TFPR across firms gets larger.

However, the variance of TFPR may not be an appropriate measure of resources distortions anymore if production function does not exhibit constant returns to scale. In our model which relaxed the assumption of constant returns to scale, using equation (3-11) and (3-12), we get the TFPR of firm i as

$$\mathrm{TFPR}_{Si} \propto \left(\mathrm{MRPK}_{Si}^{\alpha_S} \cdot \mathrm{MRPL}_{Si}^{\beta_S}\right)^{\frac{\varphi_S}{\sigma-1}} \cdot \mathrm{A}_{Si}^{\varphi_S[1-(\alpha_S+\beta_S)]}$$

$$\propto \left[\left(1+\tau_{K,Si}\right)^{\alpha_S} \cdot \left(1+\tau_{L,Si}\right)^{\beta_S}\right]^{\frac{\varphi_S}{\sigma-1}} \cdot \mathrm{A}_{Si}^{\varphi_S[1-(\alpha_S+\beta_S)]} \quad . \qquad \text{(3-17)}$$

From this equation, we can see the TFPR of firm i is proportional to its marginal revenue products of capital MRPK and labor MRPL as well as its total factor productivity A. Under the assumption of constant returns to scale, $\varphi_s[1-(\alpha_S+\beta_S)] = 0$ and then TFPR only depends on firms capital distortion $\tau_{K,Si}$ and labor distortion $\tau_{L,Si}$. This is the case Hsieh and Klenow considered. However, if the production function exhibits increasing or decreasing returns to scale, firm i's TFPR can be affected not only by the resource distortions $\tau_{K,Si}$ and $\tau_{L,Si}$, but also by its total factor productivity A_{Si}. Since total factor productivity $A_{Si} \equiv \mathrm{TFPQ}$ is different across firms, the TFPR may just reflect a difference in total factor productivity even if resource distortions do not exist at all. Therefore, the variance of TFPR may lead to an overestimate of the extent of resource distortions, because it adds variation arising from firm-level productivity differences. This is the main concern Gong and Hu (2016) raised in their paper.

Instead of the variance of TFPR, Gong and Hu (2016) propose to use the variance of MRPK and MRPL to evaluate the extent of resource distortions, and I follow my empirical analysis below. We can see from equation (3-11) and (3-12) that MRPK and MRPL do not vary across firms within the same industry when the firms do not face resource distortions. However, when resource distortions exist, MRPK and MRPL are positively related to the extent of distortions and are not affected by other factors such as the firm's TFPQ. Furthermore, compared to the variance of TFPR, using the variance of MRPK and MRPL can help to distinguish the extent of capital distortion from the extent of labor distortion clearly and to understand the effect of each distortion on industry aggregate TFP. The effect of each distortion can give us an important clue about whether an economic policy aiming to raise industry aggregate TFP should target towards labor misallocation or capital misallocation. Even in the case of constant returns to scale, if the labor distortion correlates with the capital distortion, the variance of TFPR will not be a simple geometric average of labor distortion and capital distortion anymore, and then it is very difficult to identify the effect of each distortion and which distortion is the main reason for the low TFP.

Now, I calculate the aggregate industry TFP in industry S. Since the output of industry S can be expressed as $Y_S = TFP_S \cdot K_S^{\alpha_S} \cdot L_S^{\beta_S}$, combining the equation of L_S, K_S, MRPK and MRPL above, the industry level TFP can be written as

$$TFP_S = \left\{ \sum_{i=1}^{M_S} \left[A_{Si} \cdot \left(\frac{\overline{MRPK_S}}{MRPK_{Si}} \right)^{\alpha_S} \left(\frac{\overline{MRPL_S}}{MRPL_{Si}} \right)^{\beta_S} \right]^{\varphi_S} \right\}^{\frac{1}{\varphi_S}} . \quad (3\text{-}18)$$

Equation (3-18) is the key equation I use for our empirical estimations. Also, from this equation, we can see that when resources misallocation do not exist, the marginal product of capital and labor were equalized across firms in industry S and the TFP_S then becomes to

$$TFP_S = \left(\sum_{i=1}^{M_S} A_{Si}{}^{\varphi_S} \right)^{\frac{1}{\varphi_S}} = \overline{A_S} . \quad (3\text{-}19)$$

This equation shows that if there are no resource distortions, the industry level TFP_S equals to the average TFPQ in industry S.

When $A \equiv TFPQ$, MRPL and MRPK are jointly lognormally distributed, a simple closed-form expression for the aggregate TFP:

$$\ln TFP_S = \overline{A_s} - \frac{1}{2} \cdot \frac{\sigma-1}{\sigma} \cdot \varphi_s \left[\begin{array}{c} \left(\frac{\sigma-1}{\sigma}\right) \alpha_S (1-\beta_S) \cdot var(\ln MRPK_{Si}) + \left(\frac{\sigma-1}{\sigma}\right) \beta_S (1-\alpha_S) \\ \cdot var(\ln MRPL_{Si}) + 2\left(\frac{\sigma-1}{\sigma}\right) \alpha_S \beta_S \cdot cov(\ln MRPK_{Si}, \ln MRPL_{Si}) \end{array} \right] . \quad (3\text{-}20)$$

The negative effect of distortions on aggregate TFP can be summarized by the variance of lnMRPK and lnMRPL, and the covariance between lnMRPK and lnMRPL. As mentioned above, the variance of lnMRPK and lnMRPL indicate the extent of distortion of capital and labor respectively. Then from equation (3-20), I find that the industry level TFP becomes lower, when the variance of lnMRPK and/or lnMRPL gets greater which means there is a greater distortion of capital and/or labor across firms. Equation (3-20) also tells us, if a firm faces a greater distortion of capital (labor) within industry, the marginal effect of labor (capital) distortion on industry level TFP will increase. That is, compare to firms with low capital (labor) distortion, the increase in labor (capital) distortion for firms with high capital (labor) distortion has a greater negative effect on the industry level TFP. Therefore, the level TFP gets lower as the covariance of lnMRPK and lnMRPL gets greater.

3.3.3 MRPL and MRPK Measures

To measure the effects of capital and labor misallocation, I need to calculate firm specific resource distortions and total factor productivity (TFPQ). I proceed as follows: First, I follow Levinsohn & Petrin (2003) approach to estimate the output elasticities of capital α_s and labor β_s by using real capital stock as the state variable, labor input as freely variable and intermediate input as proxy variable. I choose the LP approach because it correct the endogeneity problems in production function estimations and also solve the problem of investment proxy may not smoothly respond to the productivity shock which exist in Olley and Pakes (1996) approach.

Second, I need to set the parameter of capital price R and elasticity of substitution σ. I set the capital rent price to $R = 10\%$. Therefore the actual price of capital faced by firm i in industry S is given by $(1 + \tau_{K,Si})R$, and then differs from 10%

when the capital distortion exist. In the model, from equation of marginal product of input (3-11), (3-12), (3-15) and (3-16), we can know that the measure of resource distortions do not depend on R. If I have set R incorrectly, it affects only the average of marginal product of capital and labor, not the variance of marginal product of input MRPL and MRPK. I set the elasticity of substitution between firm to $\sigma = 3$. As equation (3-20) shows, the gains from reallocation are increasing in σ. Estimates of the substitutability of competing manufactures in the trade and industrial organization literatures such as Broda and Weinstein (2006) and Hendel and Nevo (2006) usually range this value from three to ten. Since Hsieh and Klenow (2009) set this value to 3, then it allows us to compare our estimation results to theirs.

Then, with the estimated α_s and β_s as well as the parameter R and σ, I am able to calculate the resource distortions and productivity for each firm in each country-year. Here, I follow the assumption of the substitution elasticity σ has no impact on output elasticities in Hsieh and Klenow (2009). That is, the net profits show up as payments to labor and capital pro rata in each industry. In calculating the capital distortion, labor distortion and productivity (TFPQ), I use the following equations:

$$1 + \tau_{K,Si} = \frac{\alpha_s}{\alpha_s + \beta_s} \cdot \frac{P_{Si} Y_{Si}}{R K_{Si}} \ , \tag{3-21}$$

$$1 + \tau_{L,Si} = \frac{\beta_s}{\alpha_s + \beta_s} \cdot \frac{P_{Si} Y_{Si}}{\omega L_{Si}} \ , \tag{3-22}$$

$$TFPQ_{Si} = A_{Si} = \kappa_s \cdot \frac{(P_{Si} Y_{Si})^{\frac{\sigma}{\sigma-1}}}{K_{Si}^{\alpha_s} (\omega L_{Si})^{\beta_s}} \ , \text{ where } \kappa_s = \frac{(P_S Y_S)^{-\frac{1}{\sigma-1}}}{P_S} \ . \tag{3-23}$$

Equation (3-23) describes the estimation of TFPQ for firm i. Since κ_s is a scalar that vary across industries, it do not affect the variance of output marginal product of capital and labor within an industry. Then although I do not observe this κ_s, the aggregate TFP gains from reallocation are not affected by it. Therefore, I set $\kappa_s = 1$ for each industry S for simplicity. Using the equation (3-21) to (3-23), I can calculate the estimate of $MRPK_{Si}$, $\overline{MRPK_S}$, $MRPL_{Si}$, $\overline{MRPL_S}$ and $\overline{A_s}$.

Finally, on the basis of parameters and the plant data, equation (3-18) allows us to estimate the aggregate TFP as well as to infer the aggregate TFP gains from reallocation the capital and/or labor input.

3.4 Alternative Approach to Measure the Misallocation

Based on the model in Midrigan and Xu (2014), I construct and examine an alternative approach for measuring resource misallocation arising from labor market frictions or particular policies. In contrast to the approach by Hsieh and Klenow (2009) and approach Gong and Hu (2016), this alternative approach put more emphasis on dynamic aspects of firms' decisions. I will compare the empirical results from HK approach and this alternative approach, and examine whether these results are consistent between the two different approaches.

3.4.1 Theoretical Framework

3.4.1.1 Setup

Assume that each firm has a stochastic production function using labor and capital. The production technology is given by

$$Y_t = \exp(z + s_t)^{1-\eta}(L_t^{\alpha}K_t^{1-\alpha})^{\eta} \ , \tag{3-24}$$

where Y_t is the amount of output produced at time t, L_t and K_t are the amount of labor and capital employed, respectively. The parameter z is a permanent of the producer's productivity, whereas s_t is a random variable that evolves over time according to a finite-state Markov process with a transition matrix $Q(j, i) = Pr(s_{t+1} = j|s_t = i)$.

In period t, profits are equal to

$$\pi_t = p_t Y_t - w_t L_t - r_t K_t - F - g(L_t, L_{t-1}) \ , \tag{3-25}$$

where w_t is wage rate, r_t is rental price, and F is a fixed cost. The function g captures costs arising from some friction of re-allocating labor forces.

In this setup, the Bellman equation is given by

$$V(s, L) = \max_{K,L'} \begin{Bmatrix} pe^{(z+s)(1-\eta)}(L'^{\alpha}K^{1-\alpha})^{\eta} - wL' - rK - F - g(L', L) \\ +\beta \sum_{s'} V(s', L')\, Q(s', s) \end{Bmatrix} . \tag{3-26}$$

3.4.1.2 Efficient Allocation

To compute an efficient allocation where $g(L', L) = 0$, first write the social planner's problem as

$$\max_{L_i,K_i} \int_{i \in M} \exp(z + s_i)^{1-\eta} (L_i^{\alpha} K_i^{1-\alpha})^{\eta} di \ , \tag{3-27}$$

subject to

$$\int_{i \in M} L_i di \leq \bar{L} \ , \tag{3-28}$$

$$\int_{i \in M} K_i di \leq \bar{K} \ , \tag{3-29}$$

where time subscript t is suppressed for simplifying our notation. The first order conditions are given by

$$\exp(z + s_i)^{1-\eta} \eta (L_i^{\alpha} K_i^{1-\alpha})^{\eta-1} \alpha L_i^{\alpha-1} K_i^{1-\alpha} = \lambda_1 \ \text{ for any } \ i \in M \ , \tag{3-30}$$

and

$$\exp(z + s_i)^{1-\eta} \eta (L_i^{\alpha} K_i^{1-\alpha})^{\eta-1} (1 - \alpha) L_i^{\alpha} K_i^{-\alpha} = \lambda_2 \ \text{ for any } \ i \in M. \tag{3-31}$$

After some manipulations, I can get

$$L_i = \left(\frac{\bar{L}}{\bar{Y}}\right)^{\frac{1}{1-\eta}} \left(\frac{\bar{L}}{\bar{K}}\right)^{\frac{(1-\alpha)\eta}{1-\eta}} \exp\,(z + s_i) \ \text{ for any } \ i \in M \ , \tag{3-32}$$

$$K_i = \left(\frac{\bar{K}}{\bar{Y}}\right)^{\frac{1}{1-\eta}} \left(\frac{\bar{L}}{\bar{K}}\right)^{\frac{\alpha\eta}{1-\eta}} \exp\,(z + s_i) \ \text{ for any } \ i \in M \ . \tag{3-33}$$

These equations imply that

$$\bar{L} = \left(\frac{\bar{L}}{\bar{Y}}\right)^{\frac{1}{1-\eta}} \left(\frac{\bar{L}}{\bar{K}}\right)^{\frac{(1-\alpha)\eta}{1-\eta}} \int \exp\,(z + s_i) di \ \text{ for any } \ i \in M \ ,$$

$$\bar{K} = \left(\frac{\bar{K}}{\bar{Y}}\right)^{\frac{1}{1-\eta}} \left(\frac{\bar{L}}{\bar{K}}\right)^{\frac{\alpha\eta}{1-\eta}} \int \exp\,(z + s_i) di \quad \text{for any } \; i \in M \;\; .$$

Substituting these into the production function, I obtain

$$\bar{Y} = \left[\int_{i \in M} \exp(z + s_i) di\right]^{1-\eta} (\bar{L}^{\alpha} \bar{K}^{1-\alpha})^{\eta} \quad . \tag{3-34}$$

The aggregate TFP level under the efficient allocation is given by

$$TFP^{e} = \frac{\bar{Y}}{(\bar{L}^{\alpha} \bar{K}^{1-\alpha})^{\eta}} = \left[\int_{i \in M} \exp(z + s_i) di\right]^{1-\eta} \;. \tag{3-35}$$

3.4.1.3 Distorted Allocation

When $g(L', L) \neq 0$, distortions arise in the allocation of labor across the firms. The first order conditions can be derived as

$$pe^{(z+s_i)(1-\eta)} \eta (L_i^{\alpha} K_i^{1-\alpha})^{\eta-1} \alpha L_i^{\alpha-1} K_i^{1-\alpha}$$
$$= w + \frac{\partial g(L_i', L_i)}{\partial L'} - \beta \sum_{s'} \frac{\partial V(s_i', L_i')}{\partial L'} Q(s', s) \quad , \tag{3-36}$$

and

$$pe^{(z+s_i)(1-\eta)} \eta (L_i^{\alpha} K_i^{1-\alpha})^{\eta-1} (1-\alpha) L_i^{\alpha} K_i^{-\alpha} \; = r \;\; . \tag{3-37}$$

The envelope theorem then implies

$$\frac{\partial V(s, L)}{\partial L} = -\frac{\partial g(L', L)}{\partial L} \;\; .$$

Evaluating this at (s', L'), I get

$$\frac{\partial V(s', L')}{\partial L'} = -\frac{\partial g(L'', L')}{\partial L} \;\; . \tag{3-38}$$

Substituting this equation (3-38) into the first order condition (3-36), I get

$$pe^{(z+s_i)(1-\eta)} \eta (L_i^{\alpha} K_i^{1-\alpha})^{\eta-1} \alpha L_i^{\alpha-1} K_i^{1-\alpha}$$
$$= w + \frac{\partial g(L_i', L_i)}{\partial L'} + \beta \sum_{s'} \frac{\partial g(L_i'', L_i')}{\partial L'} Q(s', s) \quad , \tag{3-39}$$

where $L_i' = L'(L_i, s_i)$ and $L_i'' = L''(L_i, s_i, s_i')$. Define

$$\mu_i \equiv \mu(L_i, s_i, s_i') = \frac{\partial g(L_i', L_i)}{\partial L'} + \beta \sum_{s'} \frac{\partial g(L_i'', L_i')}{\partial L'} Q(s', s) \;\; ,$$

the first order-conditions are simplified as

$$\alpha\eta p Y_i L_i^{-1} = w + \mu_i \ , \tag{3-40}$$

and

$$(1-\alpha)\eta p Y_i K_i^{-1} = r \ . \tag{3-41}$$

These first order conditions imply

$$L_i = \frac{\alpha r}{(1-\alpha)(w+\mu_i)} K_i \ .$$

Substituting this back into (3-18) yields

$$K_i = \frac{(1-\alpha)\eta p}{r} e^{(z+s_i)(1-\eta)} \left(\frac{\alpha r}{(1-\alpha)(w+\mu_i)}\right)^{\alpha\eta} K_i^{\eta} \ ,$$

which is

$$K_i = \left[\left(\frac{1-\alpha}{r}\right)^{1-\alpha\eta} \alpha^{\alpha\eta} \eta p\right]^{\frac{1}{1-\eta}} (w+\mu_i)^{-\frac{\alpha\eta}{1-\eta}} e^{(z+s_i)} \ . \tag{3-42}$$

Equation (3-42) implies

$$\bar{K} = \left[\left(\frac{1-\alpha}{r}\right)^{1-\alpha\eta} \alpha^{\alpha\eta} \eta p\right]^{\frac{1}{1-\eta}} \int_{i\in M} (w+\mu_i)^{-\frac{\alpha\eta}{1-\eta}} e^{(z+s_i)}\, di \ . \tag{3-43}$$

Finally, I get

$$K_i = \omega_{Ki} \bar{K} \ , \tag{3-44}$$

where

$$\omega_{Ki} = \frac{(w+\mu_i)^{-\frac{\alpha\eta}{1-\eta}} e^{(z+s_i)}}{\int_{i\in M} (w+\mu_i)^{-\frac{\alpha\eta}{1-\eta}} e^{(z+s_i)} di} \ . \tag{3-45}$$

Similarly, I have

$$L_i = \left(\frac{\alpha\eta p}{w+\mu_i}\right) e^{(z+s_i)(1-\eta)} \left(\frac{(1-\alpha)(w+\mu_i)}{\alpha r}\right)^{(1-\alpha)\eta} L_i^{\eta} \ ,$$

which is

$$L_i = \left[\left(\alpha^{(1-\alpha)\eta-1} \eta p\right) \left(\frac{1-\alpha}{r}\right)^{(1-\alpha)\eta}\right]^{\frac{1}{1-\eta}} (w+\mu_i)^{\frac{(1-\alpha)\eta-1}{1-\eta}} e^{(z+s_i)} \ . \tag{3-46}$$

This implies

$$L_i = \omega_{Li} \bar{L} \ , \tag{3-47}$$

where

$$\omega_{Li} = \frac{(w+\mu_i)^{\frac{(1-\alpha)\eta-1}{1-\eta}} e^{(+s_i)}}{\int_{i\in M}(w+\mu_i)^{\frac{(1-\alpha)\eta-1}{1-\eta}} e^{(z+s_i)} di} \quad . \tag{3-48}$$

Equations (3-44) and (3-47) together imply

$$\overline{Y} = \int_{i\in M} e^{(z+s_i)(1-\eta)} (\omega_{Li}^{\alpha}\omega_{Ki}^{1-\alpha})^{\eta} di\, (\overline{L}^{\alpha}\overline{K}^{1-\alpha})^{\eta} \quad . \tag{3-49}$$

Therefore, I can write the level of TFP under the distorted allocation is

$$TFP = \int_{i\in M} e^{(z+s_i)(1-\eta)} (\omega_{Li}^{\alpha}\omega_{Ki}^{1-\alpha})^{\eta} di \quad , \tag{3-50}$$

which can be written as

$$TFP = \frac{\left[\int_{i\in M}(w+\mu_i)^{-\frac{\alpha\eta}{1-\eta}} e^{(z+s_i)} di\right]^{1-(1-\alpha)\eta}}{\left[\int_{i\in M}(w+\mu_i)^{\frac{(1-\alpha)\eta-1}{1-\eta}} e^{(z+s_i)} di\right]^{\alpha\eta}} \quad . \tag{3-51}$$

Note that when $\mu_i = 0$, I have

$$\omega_{Ki} = \omega_{Li} = \frac{e^{(z+s_i)}}{\int_{i\in M} e^{(z+s_i)} di} \quad .$$

Hence, when $\mu_i = 0$, I recover the TFP level under the efficient allocation:

$$TFP = \int_{i\in M} e^{(z+s_i)(1-\eta)} \left[\frac{e^{(z+s_i)}}{\int_{i\in M} e^{(z+s_i)} di}\right]^{\eta} di = \left[\int_{i\in M} e^{(z+s_i)}\, di\right]^{1-\eta} . \tag{3-52}$$

3.4.1.4 Relationship between Efficient TFP and Distorted TFP

Using (3-35) and (3-51), the loss of TFP under misallocation is

$$\text{TFP losses} = \ln\left[\int_{i\in M} e^{s_i}\, di\right]^{1-\eta} - \ln \frac{\left[\int_{i\in M}(w+\mu_i)^{-\frac{\alpha\eta}{1-\eta}} e^{s_i} di\right]^{1-(1-\alpha)\eta}}{\left[\int_{i\in M}(w+\mu_i)^{\frac{(1-\alpha)\eta-1}{1-\eta}} e^{s_i} di\right]^{\alpha\eta}} \quad . \tag{3-53}$$

Substituting equation (3-40), I can rewrite it as

$$\text{TFP losses} = \ln\left[\int_{i\in M} e^{s_i}\, di\right]^{1-\eta} - \ln \frac{\left[\int_{i\in M}\left(\frac{Y_i}{L_i}\right)^{-\frac{\alpha\eta}{1-\eta}} e^{s_i} di\right]^{1-(1-\alpha)\eta}}{\left[\int_{i\in M}\left(\frac{Y_i}{L_i}\right)^{\frac{(1-\alpha)\eta-1}{1-\eta}} e^{s_i} di\right]^{\alpha\eta}} \quad . \tag{3-54}$$

When s_i and $\frac{Y_i}{L_i}$ are jointly normally distributed, equation (3-54) simplifies to

$$\text{TFP losses} = \frac{1}{2}\frac{\alpha\eta-(1-\alpha)\alpha\eta^2}{1-\eta}\text{var}\left(\log\left(\frac{Y_i}{L_i}\right)\right) \ , \qquad (3\text{-}55)$$

the TFP losses are proportional to the variance of output-labor ratio, i.e., labor productivity. Dispersion in the labor productivity generates the losses of TFP. Note that equation (3-54) or (3-55) hinges on the assumption that there is no market power, and more importantly the assumption that there is no technological difference among firms in the economy. This should be modified somehow when considering empirical analysis.

3.4.2 Estimation Strategy

The first step is to estimate equation (3-55) from our data. In doing so, I need to estimate parameters α and η. This can be done by using and estimating the production function (3-24). Taking the logarithm of (3-24) from both sides yields

$$\ln Y_t = (1-\eta)(z+s_t) + \alpha\eta \ln L_t + (1-\alpha)\eta \ln K_t \ .$$

Write this equation as

$$\ln Y_{it} = \beta_0 + \beta_1 \ln L_{it} + \beta_2 \ln K_{it} + \omega_{it} + \varepsilon_{it} \ , \qquad (3\text{-}56)$$

where $\beta_1 = \alpha\eta$, $\beta_2 = (1-\alpha)\eta$ and $\omega_{it} = (1-\eta)(z_i + s_t) - \beta_0$. I can recover the parameter α from

$$\frac{\beta_1}{\beta_2} = \frac{\alpha}{1-\alpha} \ ,$$

which implies

$$\alpha = \frac{\beta_1}{\beta_1+\beta_2} \ . \qquad (3\text{-}57)$$

The parameter η can be recovered from

$$\eta = \beta_1 + \beta_2 \ . \qquad (3\text{-}58)$$

In estimating equation (3-56), I can use the standard techniques of production function estimations such as OP approach or LP approach I discussed above.

3.5 Data and Methodology

3.5.1 Data Description

My data come from annual surveys of Chinese industrial firms for the period 1998-2005. The National Bureau of Statistics of China conducted the annual surveys and gathered firm's basic information such as the firm's name, address, ownership, birth year, etc., financial information such as revenue, total liabilities, export sales, tax, etc. and product information such as main product, product type, sales of products, input, etc. To be included in the surveys, firm's annual sales must exceed 5 million yuan.[2] The firm-level data cover industrial firms that increase in size from 165,118 firms in 1998 to 271.835 firms in 2005, and the annual increases rate is slightly less than10% on average. The number of the firms in our data set is about 20% of all industrial firms in China, but the total production quantity of these firms accounts for 95% of Chinese industrial production quantity. For 1998-2003 and 2005, all aggregates from our data are identical with the corresponding information from the Chinese Statistical Yearbook. In 2004 there are small discrepancies: the value added aggregate and the sales aggregates are higher than the variables reported in Yearbook.

The aggregate across firms in Beijing indicate that, the number of industrial firms in Beijing increases from 4,497 to 6,300 over the 1998-2005 periods, or is about 2.5% of the total number of firms in the data set. On the other hand, the number of industrial firms in Shanghai raises from 9,382 to 14,809 over the same period, nearly twice as large as the number in Beijing. I have the detail information on these firms including output, revenues, export, intermediate materials, employment, wage, capital

---

[2] Five million yuan amounts to $600.000 roughly over this period.

stock, birth year, ownership and industry classification, etc.

To calculate productivity growth, I need to link firms over time. A unique firm IDs in our data allows us to do so, and nearly 80% of firms in a given year can be linked to a firm in the previous year.[3] Firms are dropped in our TFP growth calculations when they do not appear in consecutive years. I also focus on manufacturing firms in order to eliminate effects from industry differences and as a result firms in other industries are excluded from our analysis.

3.5.2 Variable Construction

I follow the standard procedure to construct research variables for productivity analysis.[4] The details of each variable are explained below.

A. Value Added

Real value added is constructed by subtracting real intermediate input and the net of indirect taxes from the real output.

To calculate the real output, I deflate gross output reported in the data with the Brandt, Van Biesebroeck and Zhang (2012) benchmark output deflator. They construct these deflators using information from the 1998-2003 surveys, for which firms report their output in both nominal and real prices, and two-digit ex-factory price index in the 2007 China Statistical Yearbook for 2004-2006.

Deflating intermediate input is calculated using the value of intermediate input from the firm-level data and the Brandt, Van Biesebroeck and Zhang (2012)

[3] As Brandt et al. (2012) explain in detail, some firms received a new firm ID. To deal with such cases, we used the program provided by Brandt et al. (2012).

[4] To construct most research variables, we use the programs provided Brandt, Van Biesebroeck and Zhang (2012). See their paper and online appendix for the details.

benchmark input deflators. Brandt, Van Biesebroeck and Zhang (2012) construct the input deflator using the output deflators and information from the 2002 National Input-Output table. These deflators are mainly at the 3-digit industry level.

B. Employment and Wage Payment

Total annual employment and wage payment for each firm are reported in this firm-level data set. I calculate the labor input by using the wage payment rather than its employment, because incomes per worker may vary more across firms result from differences in hours worked and human capital per worker than result from worker rents.

I can calculate the labor share as wage payment divided by real value added. There is a slight decrease in the wage share in value added over the 1998-2005 period in both aggregate of Beijing and Shanghai. The labor's share of value added is only around 30% in Shanghai, and 33% in Beijing. However, the national income accounts indicate a share of labor in all industrial firms that is between 55% and 60%. Although Hsieh and Klenow (2009) inflate wage payment for all firms and compute a wage share to be consistent with the national accounts, I do not make this adjustment in this chapter. I conducted robustness check analysis and confirmed this difference does not affect our main findings qualitatively.

C. Real Capital Stock

I do not have the information on the fixed investment of each firm, but the value of each firm's fixed capital stock at original purchase prices and capital stock at original purchase prices less accumulated depreciation are reported in our data. Since these book values are nominal, I have to convert them into real values in order to compare them across time and firms.

To construct the real capital stock in each year, I need the information on the real annual investment for each firm. Our estimate of the real annual investment begins with calculating the investment in the year which the firm first appears in our data. To simplify the presentation, assume that it is 1998. I construct estimates of r_{nc}, the average growth rate of nominal capital stock for each province and each industry over period 1993-1998, using information from the 1993 annual enterprise survey. The firm's nominal capital stock in the year which it was established is then calculated by dividing the nominal capital stock in 1998 by $(1+r_{nc})^{n}$, where n is the age of the firm. And the nominal capital stock in each year up through 1998 can be calculated as multiplying the initial nominal capital stock by $(1+r_{nc})^{m}$, where m is the number of years since the firm was established. Then, I estimate the nominal annual investment as the change in nominal capital stock between years:

$$NI_{it} = NK_{it+1} - NK_{it} \ .$$

For years after the year which the firm first appears, the nominal annual investment is simply defined as the change in the firm's nominal capital stock at original purchase prices. The real annual investment is obtained by deflating the nominal investment with the investment deflator for China constructed by Brandt and Rawski (2008).

The real capital stock is estimated using the perpetual inventory method with geometric depreciation as

$$K_{it+1} = (1-\delta)K_{it} + I_{it} \ ,$$

where δ is the depreciation rate, and I_{it} is the real investment. I follow Brandt, Van Biesebroeck and Zhang (2012) and set the depreciation rate at $\delta = 0.09$.

3.6 Empirical Analysis

3.6.1 Summary Statistics

I first provide summary statistics on important research variables separately for Beijing and Shanghai. The location of each firm in our data is identified by mainly using information on the firm's post-code. Table 3.1 reports the number of firms, nominal value-added, sales, outputs, employment, and fixed assets. For the period 1998-2005, the total number of firms increases from 4497 to 6300 in Beijing, and from 9382 to 14809 in Shanghai. As a result, with the exceptions of total employment in Beijing, these aggregate of variables reported in this table all increase from 1998 to 2005.

For Beijing, the total nominal output produced by industrial firms rises from 181.6 billion yuan in 1998 to 694.6 billion yuan in 2005. Its average growth rate stands at over 20% annually. Beijing's total employment decreases from 1.26 to 1.17 million over the same period. On the other hand, for Shanghai, the total nominal industrial output raises from 519.1 to 1576.7 billion yuan, or an average growth rate of 17.5% per annum. Its total employment varies from 2.40 to 2.60 million, and increases only 1.5% per annum on average over the period 1998-2005.

The level of nominal value-added per worker for industrial firms in Beijing and Shanghai are higher than other areas of China. On average, the values of nominal value-added per worker are approximately 84.7 in Beijing and 104.4 in Shanghai between 1998 and 2005. The value is only around 60.8 in the whole of China over the same period. I can find the following interesting pattern: The differences in annual increase rate of nominal value-added per worker between Shanghai and Beijing is quite small, but the level of nominal value-added per worker is always higher in Shanghai by average 30% in any given year.

Table 3.1: Summary Statistics

Beijing						
Year	Number of firms	Value added	Sales	Output	Employment	Net value of fixed assets
1998	4497	52.50	191.32	181.65	1.26	133.75
1999	5225	58.45	209.26	200.00	1.24	137.70
2000	4572	72.27	269.11	256.54	1.13	151.72
2001	4356	75.12	300.69	290.88	1.08	157.18
2002	4551	84.04	318.28	317.35	1.08	159.12
2003	4017	101.25	388.51	380.99	1.01	162.89
2004	6871	151.88	561.61	573.33	1.14	237.71
2005	6300	167.74	727.90	694.61	1.17	259.04

Shanghai						
Year	Number of firms	Value added	Sales	Output	Employment	Net value of fixed assets
1998	9382	132.71	519.62	519.12	2.40	262.15
1999	9323	154.17	547.23	545.29	2.19	301.28
2000	8574	168.72	642.97	620.45	2.05	338.31
2001	9762	198.84	721.30	700.39	2.08	348.15
2002	10057	213.19	797.66	774.06	2.09	365.70
2003	11098	283.29	1098.26	1034.28	2.20	387.51
2004	15766	371.21	1378.22	1396.81	2.61	474.27
2005	14809	412.17	1635.37	1576.75	2.60	522.65

Note: values are denoted in billion RMB and employment in millions of workers.

3.6.2 Production Function Estimation

In this section, I present estimate for output elasticities of capital α and labor β, and examine whether I can maintain the assumption that the production function exhibits constant returns to scale for Beijing and Shanghai firms.

Table 3.2 compares parameter estimates of α and β from Olley-Pakes methodology and Levinsohn-Petrin (2003) methodology. I use the real investments and the real intermediate inputs as the proxy variable in LP estimate and OP estimate respectively. For firms in Beijing, the output elasticity of capital and labor are 0.373 and 0.609 under the OP estimation, and are 0.313 and 0.201 under the LP estimation respectively. On the other hand, for Shanghai firms, an estimate of output elasticity form the OP method is 0.378 for capital and is 0.394 for labor whereas an estimate of output elasticity form the LP method is 0.301 for capital inputs and is 0.192 for labor inputs. The output elasticities of capital and labor are considerably higher from the OP estimator than that from LP estimator. As pointed out by Levinsohn and Petrin (2003), the use of investment as a proxy variable may be problematic as many firms have zero investment. I found that about 50% firms' real investment are zero in any given year in this data, and this implies that there may be a large issue of estimating production function using this data under Olley-Pakes methodology. Therefore, I follow the LP estimation as the output elasticities of capital and labor in order to investigate the resources misallocation as well as TFP losses in the rest analysis.

Table 3.2: Production Function Estimates: OP and LP Estimation Results

Variable	OP		LP	
	Beijing	Shanghai	Beijing	Shanghai
Capital (α)	0.373	0.378	0.313	0.301
	(0.160)***	(0.020)***	(0.026)***	(0.153)***
Labor (β)	0.609	0.394	0.201	0.192
	(0.151)***	(0.007)***	(0.010)***	(0.006)***

Note: Standard errors are in parentheses. Standard errors in OP model are bootstrapped using 250 replications. ***Significant at 1% level.

The original HK model assumed constant returns to scale. However, according to my estimation results, the manufacturing industries exhibit decreasing returns to scale on average in both Beijing and Shanghai. The Wald test result rejects the constant returns to scale hypothesis under the 1% significance level. This result is consistent with Feenstra et al. (2014) and Gong and Hu (2016) who suggest that most industries reveal decreasing returns to scale in China. Then, the original HK approach may lead to a biased estimation of resources allocation, and the extension approach presented in section 3.3 allows me to measure the efficiency of resource allocation in Beijing and Shanghai more precisely.

3.6.3 TFPR and TFPQ Level and Dispersion

In this section, I investigate whether there is some difference in TFPQ level and TFPR level between Beijing and Shanghai. In doing so, I calculate TFPQ levels and TFPR levels for firms in Beijing and Shanghai by using the methodology described in section 3.3.

Table 3.3: TFPQ in Beijing and Shanghai

Year	Beijing				Shanghai			
	Mean	25th	50th	75th	Mean	25th	50th	75th
1998	6.697	5.514	6.891	7.991	7.486	6.614	7.530	8.478
1999	6.769	5.639	6.991	8.049	7.641	6.748	7.657	8.611
2000	6.949	5.905	7.202	8.283	7.794	6.919	7.802	8.759
2001	7.136	6.052	7.291	8.432	7.957	7.098	7.925	8.840
2002	7.378	6.342	7.519	8.565	7.953	7.083	7.933	8.891
2003	7.703	6.723	7.753	8.806	8.235	7.357	8.175	9.112
2004	7.664	6.706	7.769	8.851	8.235	7.326	8.208	9.175
2005	7.857	6.877	7.949	9.068	8.554	7.595	8.491	9.502
Total	7.319	6.302	7.470	8.588	8.046	7.136	8.024	9.006

Note: For firm i in Beijing, $TFPQ_i = \frac{Y_i}{K_i^{\alpha_B}(\omega_i L_i)^{\beta_B}}$; for firm i in Shanghai, $TFPQ_i = \frac{Y_i}{K_i^{\alpha_S}(\omega_i L_i)^{\beta_S}}$.

Table 3.3 reports several statistics of TFPQ levels for Beijing and Shanghai. Each firm's log(TFPQ) is a TFPQ level relative to the average TFPQ of firms in

Beijing and Shanghai. For Beijing firms, the mean of TFPQ levels ranges from 6.697 to 7.857, and its average over the period of 1998-2005 is 7.319. On the other hand, Shanghai firms record TFPQ levels ranging from 7.486 to 8.554, and the overall average is 8.046. For the period 1998-2005, both TFPQ levels of Beijing and Shanghai firms are improved. From table 3.3, year to year comparisons of mean TFPQ levels between Beijing and Shanghai show that Shanghai achieves a higher TFPQ level than Beijing in all years.

I may fail to capture important properties of TFPQ level differences between Beijing and Shanghai if I evaluate them by the mean statistics alone since the mean statistics can be effected significantly by outliers. In particular, I am concerned about the possibility that the results above are mainly driven by outliers from measurement errors. Table 3.3 also provides 25th, 50th and 75th percentiles of TFPQ levels for Beijing and Shanghai. The 25th, 50th and 75th percentiles of Beijing's TFPQ levels are all smaller than those of Shanghai's TFPQ levels for the all period 1998-2005. That is, the distribution of Beijing's TFPQ levels is shifted to the left of the distribution of Shanghai's TFPQ level over this period. Overall, this analysis suggests that Shanghai firms have higher TFPQ levels than Beijing firms.

Table 3.4 reports the evolution of TFPQ levels dispersion in Beijing and Shanghai by showing the standard deviation, the 75th minus the 25th percentiles and the 90th minus the10th percentiles of log(TFPQ). Comparing these standard deviations, 75th-25th and 90th-10th across years, we can see that the TFPQ differences for Beijing firms are much larger than those for Shanghai firms. The table also shows that although dispersion of TFPQ levels among both Beijing and Shanghai slightly declined over the period 1998-2005, it increased in some years.

Table 3.4: Dispersion of TFPQ

	1998	1999	2000	2001	2002	2003	2004	2005
	Beijing							
S.D.	2.150	2.147	2.357	2.191	2.054	1.910	2.061	2.041
75-25	2.477	2.410	2.378	2.381	2.223	2.083	2.145	2.191
90-10	5.008	5.188	5.163	4.977	4.886	4.330	4.602	4.524
N	3812	3720	3914	3945	4225	3744	6186	5114
	Shanghai							
S.D.	1.864	1.863	1.840	1.742	1.808	1.754	1.850	1.907
75-25	1.737	1.685	1.665	1.543	1.628	1.503	1.609	1.559
90-10	3.935	3.832	3.759	3.498	3.801	3.478	3.725	3.723
N	8302	8414	7771	9469	9231	10757	15123	13362

Note: Statistics are for deviations of $\log(\text{TFPQ})$ from city means; 75-25 is the difference between the 75th and 25th percentiles, and 90-10 is the difference between the 90th and the 10th percentiles; N is the number of firms.

Next, I turn to comparing TFPR levels of Beijing with those of Shanghai. Similar to TFPQ, each firm's $\log(\text{TFPR})$ is calculated as a TFPR level relative to the average TFPR in Beijing and Shanghai. Table 3.5 provides the several statistics of TFPR levels for Beijing and Shanghai from 1998 to 2005. The means of Beijing's TFPR levels increase from 2.924 in 1998 to 3.603 in 2005, and the means of Shanghai's TFPR levels rise from 3.384 to 4.094 through the same period. The overall average of TFPR levels is 3.28 for Beijing and 3.755 for Shanghai. This table shows clearly that the TFPR levels of Beijing are lower than those of Shanghai for the all period 1998- 2005. Table 3.5 also reveals that all the 25th, 50th and 75th percentiles of Beijing's TFPR levels are lower than those of Shanghai's TFPR levels. In other words, the distribution of Beijing's TFPR levels is shift down comparing to the distribution of Shanghai's TFPR level over the period.

Table 3.5: TFPR in Beijing and Shanghai

Year	Beijing				Shanghai			
	Mean	25th	50th	75th	Mean	25th	50th	75th
1998	2.924	2.249	2.971	3.676	3.384	2.854	3.426	3.999
1999	2.941	2.297	3.040	3.702	3.482	2.951	3.499	4.092
2000	3.054	2.465	3.171	3.844	3.551	3.028	3.581	4.149
2001	3.149	2.524	3.213	3.927	3.693	3.173	3.692	4.240
2002	3.296	2.680	3.370	4.036	3.672	3.143	3.681	4.259
2003	3.460	2.855	3.500	4.146	3.876	3.356	3.861	4.428
2004	3.542	2.938	3.624	4.319	3.920	3.363	3.939	4.526
2005	3.603	3.021	3.680	4.361	4.094	3.508	4.077	4.686
Total	3.280	2.642	3.346	4.063	3.755	3.196	3.764	4.364

Note: For firm i in Beijing, $TFPR_i = \frac{P_i Y_i}{K_i^{\alpha_B}(\omega_i L_i)^{\beta_B}}$; for firm i in Shanghai, $TFPR_i = \frac{P_i Y_i}{K_i^{\alpha_S}(\omega_i L_i)^{\beta_S}}$.

Table 3.6: Dispersion of TFPR

	1998	1999	2000	2001	2002	2003	2004	2005
	Beijing							
S.D.	1.292	1.282	1.422	1.327	1.261	1.176	1.290	1.261
75-25	1.427	1.405	1.379	1.403	1.356	1.292	1.381	1.340
90-10	2.964	3.007	3.044	2.942	2.944	2.683	2.848	2.773
N	3812	3720	3914	3945	4225	3744	6186	5114
	Shanghai							
S.D.	1.076	1.043	1.013	0.956	1.012	0.921	1.007	0.956
75-25	1.145	1.141	1.121	1.066	1.116	1.072	1.163	1.179
90-10	2.420	2.349	2.284	2.147	2.348	2.099	2.330	2.297
N	8302	8414	7771	9469	9231	10757	15123	13362

Note: Statistics are for deviations of $\log(TFPR)$ from city means; 75-25 is the difference between the 75th and 25th percentiles, and 90-10 is the difference between the 90th and the 10th percentiles; N is the number of firms.

Table 3.6 provides the evolution of TFPR dispersions. During the whole period of 1998-2005, the distributions of Beijing's TFPR levels are dispersed more than that of Shanghai's TFPR level. This implies the gap in TFPR levels among Beijing firms is larger than that among Shanghai firms. Although the gap in TFPR levels between firms narrowed down during the whole period for both Beijing and Shanghai, it spread in several years such as in 2004.

The evolution of TFPR level and TFPR dispersion follows a very similar pattern with that of TFPQ level in both Beijing and Shanghai. Firms in Beijing have lower TFPQ levels as well as TFPR levels than firms in Shanghai, and experience both larger TFPQ level and TFPR level dispersions than firms in Shanghai.

3.6.4 MRPL and MRPK Dispersion

To understand TFP differences between Beijing and Shanghai comprehensively, I also measure the dispersion of marginal revenue product of labor MRPL and capital MRPK in Beijing and Shanghai. MRPL and MRPK evaluate the extent of labor distortion and capital distortion, and then help to understand the effect of each resource misallocation on industry aggregate TFP clearly.

Table 3.7 displays several statics of MRPK dispersions for the period 1998-2005. Firm's MRPK dispersion measures as a MRPK level relative to the average MRPK in Beijing and Shanghai. Generally, in both Beijing and Shanghai, the standard deviation, the differences between percentile 75 and percentile 25, and the gaps between percentile 90 and percentile 10 declined over the period 1998 to 2005, although some of them increased in some years. In other words, the capital distortion among firms becomes smaller in both Beijing and Shanghai over this period. This may be partly because more firms in these cities are able to access to efficient credit markets in Beijing and Shanghai. When Beijing and Shanghai are compared, the values of MRPK are all higher for Beijing than those for Shanghai, indicating that capital distortion of Beijing firms is more severe than Shanghai firms during this period.

Table 3.7: Dispersion of MRPK

	1998	1999	2000	2001	2002	2003	2004	2005
				Beijing				
S.D.	1.717	1.671	1.751	1.694	1.668	1.553	1.764	1.679
75-25	1.916	1.865	1.899	1.998	2.005	1.844	2.085	1.931
90-10	3.835	3.786	3.858	3.844	3.880	3.587	4.048	3.806
N	3877	3824	3950	3980	4258	3748	6208	5135
				Shanghai				
S.D.	1.403	1.419	1.354	1.411	1.411	1.333	1.444	1.328
75-25	1.533	1.571	1.513	1.636	1.598	1.550	1.720	1.624
90-10	3.135	3.317	3.142	3.222	3.277	3.078	3.400	3.195
N	8352	8435	7796	9522	9286	10757	15125	13363

Note: For firm i in Beijing, $\mathrm{MRPK_i} = \frac{\sigma-1}{\sigma} \cdot \frac{\alpha_B P_i Y_i}{K_i}$; for firm i in Shanghai, $\mathrm{MRPK_i} = \frac{\sigma-1}{\sigma} \cdot \frac{\alpha_S P_i Y_i}{K_i}$.

Statistics are for deviations of $\log(\mathrm{MRPK})$ from city means; 75-25 is the difference between the 75th and 25th percentiles, and 90-10 is the difference between the 90th and the 10th percentiles; N is the number of firms.

Table 3.8: Dispersion of MRPL

	1998	1999	2000	2001	2002	2003	2004	2005
				Beijing				
S.D.	1.424	1.401	1.510	1.408	1.350	1.300	1.373	1.358
75-25	1.594	1.575	1.581	1.622	1.650	1.557	1.614	1.615
90-10	3.256	3.322	3.329	3.305	3.286	3.108	3.196	3.058
N	3924	3798	3966	4004	4284	3762	6262	5161
				Shanghai				
S.D.	1.253	1.170	1.144	1.097	1.153	1.053	1.131	1.079
75-25	1.394	1.379	1.374	1.317	1.368	1.311	1.377	1.386
90-10	2.831	2.746	2.685	2.574	2.715	2.531	2.715	2.669
N	8417	8460	7815	9590	9305	10883	15225	13426

Note: For firm i in Beijing, $\mathrm{MRPL_i} = \frac{\sigma-1}{\sigma} \cdot \frac{\beta_B P_i Y_i}{L_i}$; for firm i in Shanghai, $\mathrm{MRPL_i} = \frac{\sigma-1}{\sigma} \cdot \frac{\beta_S P_i Y_i}{L_i}$.

Statistics are for deviations of $\log(\mathrm{MRPL})$ from city means; 75-25 is the difference between the 75th and 25th percentiles, and 90-10 is the difference between the 90th and the 10th percentiles; N is the number of firms.

The MRPL dispersions in Beijing and Shanghai are reported in Table 3.8. Note that MRPL dispersion of a firm is measured as a MRPL level relative to each city's average MRPL. The measures of dispersion of MRPL have similar patterns to those of MRPK. First, except a few years, MRPL dispersions become smaller over time in both two cities. This seems to suggest that the labor distortion has improved over the period 1998-2005. Second, the measures of MRPL are lower in Shanghai than those in Beijing throughout the whole period. This result suggests that that firms in Shanghai benefits from a relative efficient labor allocation than firms in Beijing.

In this section I show that the dispersion of MRPL and MRPK more spread in Beijing than that in Shanghai. Recall that in the last section, Table 3.4 and Table 3.6 reveal firms in Beijing suffer a greater TFPQ and TFPR dispersions than firms in Shanghai. An important finding from these patterns of dispersions is that, a greater dispersion of TFPQ levels and TFPR levels are associated with greater distortion in labor and capital allocations, which is consistent with the findings in Hsieh & Klenow (2009) and other existing studies.

3.6.5 Comparison of Two TFP Losses Measures

To further investigate the findings explained above, I next use the HK and Midrigan and Xu (2014) approach and measure and compare TFP losses (See sections 3.3 and 3.4 for these methodologies).

Based on the equation (3-20) in section 3.3.2 and the equation (3-55) in section 3.4.1, I estimated the aggregate TFP losses from two methodologies. Table 3.9 provides estimation results of aggregate TFP losses in Beijing and Shanghai from two different approaches. One finding stands out in Table 3.9 when Beijing and Shanghai are compared. The average TFP losses of Beijing firms are larger than those of Shanghai throughout the whole periods 1998-2005. The total average TFP loss in Beijing is 0.394 from HK approach and is 0.275 from the alternative approach. The corresponding numbers for Shanghai firms are 0.260 and 0.169. Regardless of the approaches, the average TFP losses of Beijing firms are all larger than those of Shanghai firms. This finding appears to indicate that resource allocation is distorted more in Beijing than in Shanghai, and this distortion shows up in the difference in the TFP loss between Beijing and Shanghai.

Table 3.9: TFP Losses: HK Approach and MX Approach Results

Variable	TFP Loss			
	HK (Mean)		MX (Mean)	
	Beijing	Shanghai	Beijing	Shanghai
1998	0.406	0.277	0.288	0.208
1999	0.389	0.271	0.279	0.181
2000	0.431	0.253	0.324	0.173
2001	0.397	0.261	0.281	0.159
2002	0.380	0.267	0.258	0.176
2003	0.340	0.237	0.240	0.147
2004	0.413	0.273	0.267	0.169
2005	0.385	0.239	0.262	0.154

3.6.6 Gains from Reallocation Capital and/or Labor

In this section, I conduct a thought experiment by examining how much aggregate TFP would increase if capital and labor were allocated efficiently. In doing so, I calculate the ratio of actual TFP expressed in equation (3-18) to the efficient TFP level that equalizing the MRPK and/or MRPL among the same cities. After some manipulations, I get the expression of gains from reallocation capital and labor, from the capital only and from the labor only as

$$\left(\frac{Y}{Y_{efficient}}\right)_{KL} = \left\{\sum_{i=1}^{M_S}\left[\frac{A_{si}}{\overline{A_S}} \cdot \left(\frac{\overline{MRPK_S}}{MRPK_{Si}}\right)^{\alpha_s}\left(\frac{\overline{MRPL_S}}{MRPL_{Si}}\right)^{\beta_S}\right]^{\varphi_s}\right\}^{\frac{1}{\varphi_s}},$$

$$\left(\frac{Y}{Y_{efficient}}\right)_{K} = \left\{\sum_{i=1}^{M_S}\left[\frac{A_{si}}{\overline{A_S}'} \cdot \left(\frac{\overline{MRPK_S}}{MRPK_{Si}}\right)^{\alpha_s}\left(\frac{\overline{MRPL_S}}{MRPL_{Si}}\right)^{\beta_S}\right]^{\varphi_s}\right\}^{\frac{1}{\varphi_s}},$$

$$\left(\frac{Y}{Y_{efficient}}\right)_{L} = \left\{\sum_{i=1}^{M_S}\left[\frac{A_{si}}{\overline{A_S}''} \cdot \left(\frac{\overline{MRPK_S}}{MRPK_{Si}}\right)^{\alpha_s}\left(\frac{\overline{MRPL_S}}{MRPL_{Si}}\right)^{\beta_S}\right]^{\varphi_s}\right\}^{\frac{1}{\varphi_s}},$$

where

$$\overline{A_S}' = \left\{\sum_{i=1}^{M_S}\left[A_{si} \cdot \left(\frac{\overline{MRPL_S}}{MRPL_{Si}}\right)^{\beta_S}\right]^{\varphi_s}\right\}^{\frac{1}{\varphi_s}},$$

$$\overline{A_S}'' = \left\{\sum_{i=1}^{M_S}\left[A_{si} \cdot \left(\frac{\overline{MRPK_S}}{MRPK_{Si}}\right)^{\alpha_s}\right]^{\varphi_s}\right\}^{\frac{1}{\varphi_s}},$$

respectively. Using these expressions, I estimate the TFP gain from capital and/or labor in each year as $(Y_{efficient}/Y - 1) * 100$.

Table 3.10: TFP Gains from Equalizing MRPK and/or MRPL

	1998	1999	2000	2001	2002	2003	2004	2005
	Beijing (%)							
K	47.621	153.605	173.853	107.370	86.359	67.319	85.938	84.907
L	6.493	19.727	22.588	20.131	22.534	24.247	28.703	30.320
K&L	84.371	220.891	185.074	82.494	91.034	68.520	135.230	138.838
	Shanghai (%)							
K	7.239	7.914	4.609	9.491	19.972	18.257	36.153	40.853
L	3.936	8.837	10.973	18.488	13.798	19.805	25.475	28.617
K&L	17.898	21.540	17.998	30.692	105.036	58.988	130.151	155.676

Table 3.10 reports the estimation results of TFP gains in Beijing and Shanghai from fully equalizing the MRPK and MRPL across firms in each city over the periods

1998 to 2005. By fully equalizing the MRPK and MRPL, the aggregate TFP gains are 68% - 220% in Beijing, and are 17% - 155% in Shanghai. If only equalizing the MRPK, the aggregate TFP would improve by 47% - 173 in Beijing, and 7% - 40% in Shanghai. On the other hand, if only fully liberalizing the MRPL, the aggregate TFP will increase by 6% - 30% and by 3% - 28% in Beijing and Shanghai respectively. Hsieh and Klenow (2009) report that the percentage gain from the efficient allocation of labor and capital is 86.6 for China in 2005. According to Table 3.10, the estimated gain of Beijing in this study is similar to that number but the estimated gain of Shanghai is much lower than that number. This indicates that Shanghai's economy achieves a higher level of efficiency in terms of resource allocation than other Chinese cities.

The estimated TFP gains of Shanghai from efficient resource reallocation of either labor or capital is smaller than those of Beijing throughout the whole periods 1998-2005. This suggests that the aggregate TFP level of Beijing can improve more than that of Shanghai once resource misallocations are corrected. The flip side of this statement is that, relative to Beijing, there is a little room for improving the aggregate productivity of Shanghai by correcting its resource misallocation. It is also interesting to see that the capital misallocation of the Beijing economy has more profound impacts on its aggregate TFP level than its labor misallocation. This would render support to a policy of correcting capital misallocation if one wants to improve the aggregate productivity level of Beijing.

Table 3.10 also shows that even though the TFP gains from fully equalizing either the MRPK or MRPL is smaller for Shanghai than those for Beijing in all cases. The same results do not hold qualitatively when both the labor and capital are reallocated to the efficient level in year 2002 and 2005. The estimated TFP will increase by 105% in year 2002 and 155% in year 2005 from efficient allocation of both labor and capital in Shanghai, and by 91% and 138% respectively in year 2002 and 2005 in Beijing. Recall that the equation (3-20) in section 3.3.2 implies that the negative effect of distortions on aggregate TFP can arising from three factors, $\mathrm{var}(\ln \mathrm{MRPK}_{Si})$, $\mathrm{var}(\ln \mathrm{MRPL}_{Si})$ and $\mathrm{cov}(\ln \mathrm{MRPK}_{Si}, \ln \mathrm{MRPL}_{Si})$. The first two factors express the extents of capital and labor market distortions, respectively. The last term measures the extent of complementary of capital and labor distortions, and captures a synergy effect of capital and labor misallocation on aggregate TFP. For example, when high capital distortion accompanies with high labor distortion, the resource misallocation effect becomes larger, and results in a lower aggregate TFP. Therefore, this result indicates that the synergy effect of labor and capital distortion is greater in Shanghai than in Beijing for some years, and amplifies rapidly after the year 2002. This may suggest that the Shanghai aggregate TFP can improve more by correcting both labor and capital misallocation rather than by only correcting either labor or capital misallocation.

To illustrate the magnitude of these TFP gains, take 100 percent gain as an example. This means that the aggregate TFP level has doubled so that all firms now can produce twice as much output as they used to do from a given amount of inputs without any technological advancement. According to Syverson (2011), a productivity difference between a firm located at top 10 percentile in terms of productivity and a firm located at bottom10 percentile is about 100 percent. Thus, the 100 percent gain from efficient resource allocation is equivalent to the transformation of the bottom 10 percentile firm to the top 10 percentile firm. An economic growth rate in advanced countries is around 3 percent. It will take 25 years to double outputs if the growth rate is 3 percent and if productivity is only the source of economic growth. These examples illustrate the importance of resource allocation in the Beijing economy because the average productivity gain of Beijing from efficient allocation of capital and labor is 114.

3.6.7 Productivity of Entrants, Incumbents and Exiting Firms

In this section, I turn my attention to the correlation of productivity and market distortions with entrants, incumbents and exiting firms in order to investigate whether exiting firms are associated with lower productivity and whether new entrants achieve higher productivity. In doing so, I examine how each of entrants, incumbents and exiting firms contributes to TFPQ and TFPR, as well as the extent of labor and capital market distortions.

In this analysis, a firm in the data is identified as a new entrant in year t if the firm appears in year t for the first time, and is defined as an exiting firm in year t if the firm disappears from year t onward. Thus, a firm is an incumbent if it is neither an entrant nor an exiting firm.

Table 3.11: TFPR, TFPQ, MRPK, MRPL by Entrants, Incumbents and Exiting Firms

	TFPQ	TFPR	MRPK	MRPL
		Beijing		
Entrants	0.069	0.109	0.726	0.409
	(0.008)***	(0.010)***	(0.035)***	(0.028)***
Exiting Firms	-0.149	-0.124	-0.047	-0.352
	(0.006)***	(0.008)***	(0.026)*	(0.021)***
		Shanghai		
Entrants	-0.018	0.013	0.483	0.077
	(0.003)***	(0.004)***	(0.018)***	(0.015)***
Exiting Firms	-0.106	-0.073	0.048	-0.323
	(0.003)***	(0.004)***	(0.016)***	(0.013)***

Note: The dependent variables are dummies for entrants and exiting firms. The independent variables are the deviation of log(TFPQ), log(TFPR), log(MRPK) and log(MRPL) from the city means. Regressions include year fixed effects. Standard errors are in parentheses. * Significant at 10% level, ***Significant at 1% level.

Table 3.11 reports estimation results about TFPR, TFPQ, MRPK and MRPL from OLS regression of entrants and exiting firms. In this regression analysis, the base category is incumbents firms. The results show that TFPQ and TFPR level for entrants are positive, while those for exiting firms are negative in Beijing. This indicates that new entrants achieve higher TFPQ and TFPR than incumbents, and the productivity of exiting firms are lower than that of incumbents in Beijing. On the other hand, similar to firms in Beijing, the TFPQ and TFPR are both negative for exiting firms in Shanghai, suggests that Shanghai firms associated with lower productivity are more likely to exit. However, for new entrants in Shanghai, the contribution to TFPQ is estimated to be negative 0.018, though the contribution to TFPR is positive 0.013. This suggests that some Shanghai firms may less productive than incumbents when they enter into the market.

I next look at resources misallocation for entrants and exiting firms. Table 3.11 reports that MRPK and MRPL of Beijing entrants are 0.726 and 0.409, respectively. Thus, both the capital distortion and labor distortion are more severe for entrants than for incumbents in Beijing. Thus, this indicates that firms may face a higher capital price or labor price when they enter the Beijing market. On the other hand, the MRPK and MRPL of exiting firms are both negative in Beijing. This seems to suggest that even though there are some benefits from relative efficient capital and labor allocation, some firms still cannot survive and finally are forced to exit from the Beijing market. In Shanghai, the MRPK and MRPL are both positive for entrants. This suggests that new Shanghai entrants suffer capital and labor misallocation to a greater extent than incumbents and this pattern is similar to Beijing. However, the resources misallocation among exiting firms in Beijing and Shanghai exhibit somewhat different pattern. The MRPK of exiting firms is estimated to be positive 0.048, and the MRPL of exiting firms is negative 0.323 in Shanghai. This indicates that although Shanghai exiting firms suffer relative inefficient capital allocation in a larger magnitude than incumbents, they benefit from efficient labor allocation when they exit. This appears to suggest that the relative severe market distortions may not be a main reason for firms exiting from the Shanghai market. Furthermore, when Beijing and Shanghai are compared, the values of entrants' MRPK and MRPL are both higher for Beijing than those for Shanghai, implying that new entrants in Beijing suffer inefficient resources allocation more than new entrants in Shanghai.

These estimation results demonstrate that productivity of exiting firms are lower than that of incumbent firms, and the new entrants are more productive than incumbents, though there is some exception in Shanghai. Even though new entrants suffer inefficient resources allocations more than incumbents in both Beijing and Shanghai, I do find a different pattern of market distortion among exiting firms between Beijing and Shanghai.

3.7 Conclusion of Chapter 3

I estimated productivity losses of Beijing and Shanghai arising from resource misallocation by using the Hsieh and Klenow (2009) method and its extended method and by using the firm-level Chinese manufacturing data. The main findings are twofold. First, the aggregate TFP level of Beijing is lower than that of Shanghai. Second, the estimate of this study suggests that the aggregate TFP level of Beijing will improve significantly if resources are allocated to the efficient level from the current allocation. In particular, correcting capital misallocation is a key for improving the aggregate productivity of Beijing.

In this study, I was not able to empirically identify a source of misallocation because of data limitations. It is one limitation of this study, and some findings in this study should be interpreted with caution. Although a causal effect is not identified, the findings of this study still carry important policy implications. One important policy implication is that a policy intended to remove some frictions in the capital market would be more effective to raise the aggregate productivity of Beijing than a policy of intervening labor markets.

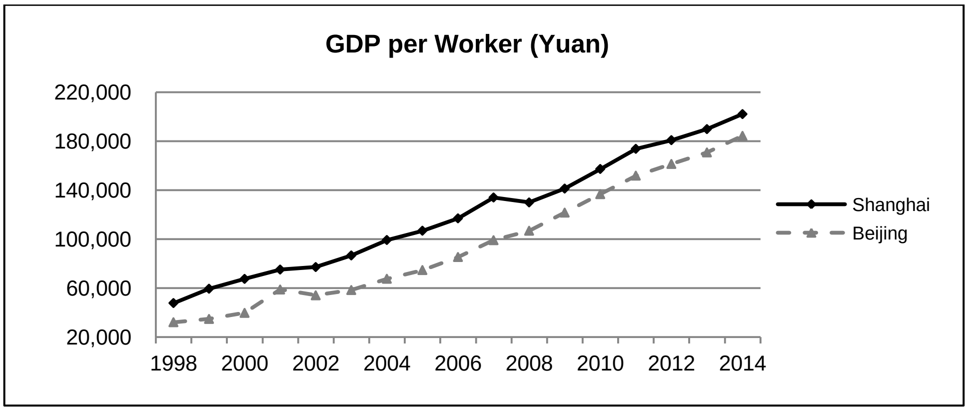
Chapter 4.

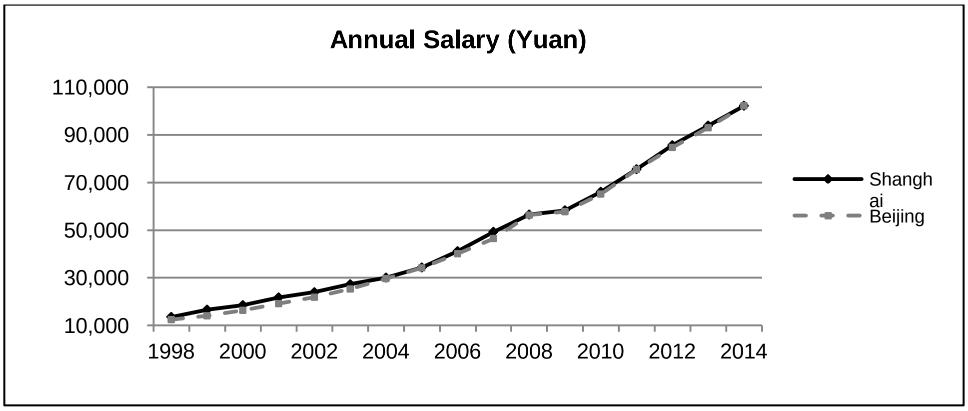
Regional Productivity, Job Turnover and Profitability under Hukou Allocation System

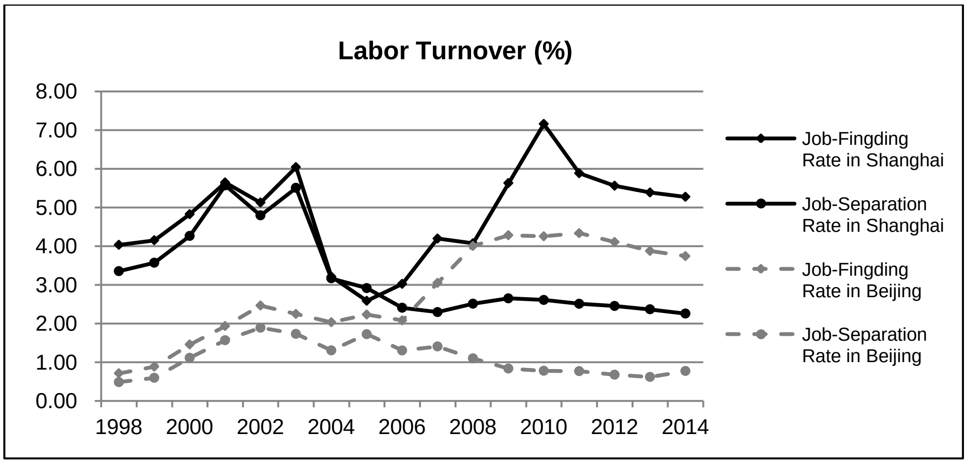
4.1 Introduction of Chapter 4

Sources of productivity disparities across regions have been studied in recent years (Ciccone and Hall, 1996; Deng and Jefferson, 2011). Understanding mechanisms that generate regional productivity differences is of primal interest to economists and policy makers, because regional development hinges crucially on productivity improvement. In this Chapter 4, I try to contribute to this line of research by investigating a source of productivity difference between the two largest cities in China, Beijing and Shanghai.

Figure 4.1: GDP per Worker, Average Salary and Labor Turnover in Shanghai and Beijing







Note: GDP per person employed was calculated by dividing the GDP by employment. The Job-Finding rate is the fraction of unemployed persons that flow out of unemployment, and Job-Separation is the fraction of workers who leave their jobs.

Source: China Statistical Yearbook 1998-2015, National Bureau of Statistics of China; China Labor Statistical Yearbook 1998-2015, Department of Population and Employment Statistics of National Bureau of Statistics of China and Ministry of Labor and Social Security of China.

Figure 4.1 shows that GDP per worker and labor turnover of Beijing is consistently lower than that of Shanghai during the period of 1998 to 2014. This figure also reports that although the annual salary of Beijing is lower than that of Shanghai before 2008, such difference becomes smaller after 2008. Past studies have attributed this productivity difference to such factors as R&D investment and technological progress (Zhang et al., 2011), density of economic activity (Rizov and Zhang, 2013; Cai, Wang and Du, 2002), and economic structures (Li and Haynes, 2011; Chan and Zhang, 1999). Although these factors certainly influence the productivity level, this chapter offers an alternative mechanism that explains this observed productivity difference. More specifically, I argue that the productivity difference stems from two cities' Hukou allocation system (China's household registration system). I view Hukou as work-related benefits and build a theoretical model, where Beijing firms are entitled to allocate Hukou to their employees and can utilize it as a device for economizing on labor costs. In contrast, these Hukou are allocated to workers directly from the city authorities in Shanghai. The key mechanism of the model is that some Beijing firms find it optimal to retain mismatched workers, because Beijing's Hukou allocation policy allows them to set wages to a lower level and this advantage outweighs a loss in output due to lower productivity. Because of this firm's optimal behavior, the GDP per worker, the turnover rate and the wage are lower in Beijing than in Shanghai. This prediction is consistent with the statistics on these variables reported in Figure 4.1.

The main contribution of this chapter is to show that Hukou system plays a significant role in affecting regional productivity and therefore regional development. The fundamental lesson from this study is that the level of regional productivity can be low due to inefficiency arising from the retention of mismatched workers when Hukou are allocated to workers through their employer rather than being allocated to them directly. This insight can be applied to a wide range of economic issues concerning work-related benefits. This is especially valuable for immigration policy, because a work permit in some countries can be regarded as a work-related benefit.

This chapter is organized as follows. Section 4.2 describes the main features of the Hukou allocation system. Section 4.3 provides a simple model where a firm's optimal behavior is derived. Section 4.4 discusses three main implications from the model. Section 4.5 presents the concluding argument.

4.2 Background

In China, the Hukou allocation policy has a significant impact on people's standard of living because the quality of social services (e.g., education and medical services) that a person can potentially receive depends on where his or her Hukou is established, rather than where he or she lives. The details of the Hukou system have been extensively reviewed by many researches (e.g., Chan and Zhang, 1999), and a feature relevant to this analysis is as follows: While the benefits attached to a Hukou are so attractive that workers may sacrifice a part of their compensation in order to acquire the Hukou, firms may use an assignment of Hukou strategically to attract and retain desirable workers. Thus, the Hukou allocation system likely affects regional economic activities.

The most notable difference in the Hukou allocation policy between Beijing and Shanghai is the process of allocating a new Hukou. The Shanghai city authorities allocate Hukou to part of workers directly after they become a permanent employee. On the other hand, a fixed amount of Hukou is first allocated to a firm in Beijing. After this, it is the firm that determines whether a worker establishes a Hukou in Beijing by providing this valuable Hukou to worker. This feature of the Beijing's Hukou allocation policy allows Beijing firms to set lower wages in exchange for giving Hukou to their employees. Unlike other rewards, a firm cannot retract Hukou from its employees once they are already assigned to them, since the Hukou are their employees' property, even after they quit the firm.

Hukou system is a powerful household registration institution that affects many fundamental aspects of life for hundreds of millions of Chinese as well as China's economic development. Past studies have focused on examining its roles in obstructing migration (Goldstein and Goldstein, 1991; Chan and Yang, 1996), industrialization and urbanization (Chan, 1992; Cheng and Selden, 1994), and inequity of social welfare (Liu, 2005; Afridi et al., 2012). Recently, researchers have been increasingly aware of its broader ramification on Chinese society and regional economy. Yang, Xu, and Xiang (2003) pointed out that the Hukou system slowed down economic growth in developed regions because it prevents lower cost labor force from moving to more developed regions from poorer areas. Ma (1999) discussed that different regions have an uneven performance on economic development even though spatial patterns of interprovincial migration are very similar. Ma (1996) showed that, compared with intra-provincial migration, inter-provincial migration is a more prominent factor of population redistribution and economic development at the national and regional levels, and suggested that Hukou system has a negative effect on regional development because it prevents interprovincial migration. However, most of the past studies on this topic maintain an assumption that the Hukou system is the same through all regions. In this chapter, I relax this assumption and focus on the difference in Hukou allocation system between Beijing and Shanghai, and show that different Hukou allocation systems can change a path of economic development. This result is consistent with the finding from Ma (1999).

This analysis is also related to research on the relationship between labor market institutions and firm productivity. Ishida (2005) investigated how a unique management style of lifetime employment in Japanese labor market influences firm performances, and discussed that firms may not be better off under such lifetime employment. Besley and Ghatak (2008) argued that firms can increase their output by using status rewards such as job title or a medal. This research is also related to a line of research on immigration policy for economic growth (Friedberg and Hunt, 1995; Ruhs, 2008), since immigration permission in some country can be regarded as work-related benefit similar to Hukou.

I construct a theoretical model in which different Hukou allocation systems generate differences in productivity among regional economies. A key insight from this theoretical model is that a mismatch between firms and workers arising from a Hukou allocation system can be a source of regional productivity differences. Since the effects of Hukou system on regional economic growth have been usually examined from the viewpoint of rural-to-urban migration, this analysis shed a new light on this topic by offering an alternative explanation for regional productivity differences resulting from different Hukou systems.

4.3 Model

4.3.1 Environment

There are two cities, City B (Beijing) and City S (Shanghai), in the economy. While Hukou are allocated to workers directly from the city authorities in City S, these Hukou are allocated to workers through firms in City B and firms may use this advantage strategically to raise their profits. I consider a two-period model with a unit mass of risk-neutral firms and of risk-neutral workers in each city, each firm indexed by *i*.

In both periods, each firm employs at most one worker to produce output. The amount of output the firm produces only depends on a matching quality between the firm and a worker. For simplicity, the matching quality is either a good matching ($m = 1$) or a poor matching ($m = 0$). The production function of firm i is given by

$$Q(m_i) = x_{m_i} \quad \text{for} \quad m_i = 0,1 \ ,$$

where $x_1 > x_0$ is assumed.

Although the production function is identical across firms, they are heterogeneous with respect to the ability to draw a matching quality. Let q denote this ability and the variable q is assumed to be distributed uniformly over the unit interval between 0 and 1:

$$f(q) = 1 \quad \text{for} \quad q \in [0,1] \ .$$

A matching quality is determined randomly after hiring a worker. Specifically, firm *i*'s probability of drawing a given matching quality is given by

$$\text{prob}(m_i = d|q_i) = q_i^d (1 - q_i)^{1-d} \quad \text{for} \quad d = 0,1 \ .$$

The probability of drawing a good matching quality increases with firm's ability *q*.

4.3.2 Employment Contract

In the first period, firms in both cities offer an employment contract that specifies wages for both periods. The contract offered by firm in City B also includes whether provide Hukou to the worker. The matching quality is unknown to both the firm and the worker at the beginning of the first period, but it is revealed completely at the end of the first period. At the beginning of the second period, the firm may dismiss its worker. To ease the explanation, I assume that renegotiation about the second period wage is never allowed after knowing the matching quality. I do not discount time.

4.3.2.1 Employment Contract in City S

In City S, a firm offers an employment contract so as to maximize its expected profit subject to a worker's participation constraint. Formally, firm's maximization problem can be written as

$$\max_{w_i} \sum_{t=1}^{2} \left[q_i (x_1 - w_{it}) + \left(1 - q_i\right)(x_0 - w_{it}) \right]$$

$$\text{s.t.} \quad w_{it} \geq \bar{u} \quad \text{for} \quad t = 1,2 \qquad .$$

The optimal wage is simply given by

$$w_S \equiv w_{it} = \bar{u} \quad \text{for} \quad t = 1,2 \ .$$

4.3.2.2 Employment Contract in City B

Assume that α fraction of the firms in City B obtain the right to entitle Hukou to their employee. For simplicity, Hukou are given to firms only in the first period. The wage setting is the same as the firms in City S, if a firm does not obtain this right. When a firm in City B obtains that right, its employment contract is determined by solving the following maximization problem:

$$\max_{w,A} \sum_{t=1}^{2} \left[q_i (x_1 - w_{it}) + \left(1 - q_i\right)(x_0 - w_{it}) \right]$$

$$\text{s.t.} \quad w_{it} + A_{it} h \geq \bar{u} \quad \text{for} \quad t = 1,2 \qquad ,$$

where A is a variable that equal to 1 if the Hukou is entitled and 0 otherwise, and h captures the level of benefit from obtaining a Hukou. Regardless of the firm’s decision on providing the Hukou to its worker, the optimal wage is set to

$$w_B(A_{it}) \equiv w_{it} = \bar{u} - A_{it} h \quad \text{for} \quad t = 1,2 \ .$$

Then turn to the decision on *A*. By setting $A_{i1} = A_{i2} = 1$, the firm can save on the cost of both periods’ wages, if the firm decides to retain the worker in the second period. In addition, it can save on the cost of the first period’s wage, even when the firm dismisses the worker at the beginning of the second period. On the other hand, the firm is able to leave the entitlement of Hukou to the second period and set $A_{i1} = 0$ and $A_{i2} = 1$, but this can only save the second period’s wage. Therefore, the firm’s profit is maximized when both A_{i1} and A_{i2} are set to 1. By combining this result with the equation above, the optimal wage is given by

$$w_B \equiv w_{it} = \bar{u} - h \quad \text{for} \quad t = 1,2 \ .$$

4.3.3 Decision on the Dismissal of a Worker

At the beginning of the second period, each firm knows the quality of match with the worker employed in the first period and makes a decision on whether to dismiss her. When the firm decides to retain the worker in the second period, she carries out the second period production. When the firm chooses to dismiss her, it hires a new worker who is drawn randomly from the pool of workers. Therefore, there are four possible firm behaviors in the second period:

I. After knowing a good match between the firm and the worker, the firm retains the worker and receives output x_1.

II. After knowing a good match, the firm dismisses the worker and hires a new worker.

III. After knowing a poor match, the firm retains the worker and receives output x_0.

IV. After knowing a poor match, dismisses the worker and hires a new worker.

The firm's optimal strategy with respect to retention of workers can be solved by using backward induction with the optimal wage showed in Section 4.3.2.

4.3.3.1 Dismissal Decision in City S

To consider an optimal dismissal decision of a firm in City S, suppose that a good matching quality is realized in the first period. Taking the labor contract as given, the second period's profit is $x_1 - \bar{u}$ if the firm retains the well-matched worker, and the expected profit is $q_i(x_1 - \bar{u}) + (1 - q_i)(x_0 - \bar{u})$ if it dismisses her and hires a new worker. Thus, the firm's optimal decision in this case is to retain the well-matched worker, because $x_1 > x_0$ and $q_i \in [0,1]$. Next, consider the case where matching quality was poor in the first period. The second period profit is $x_0 - \bar{u}$ when retaining the worker, and the expected profit is $q_i(x_1 - \bar{u}) + (1 - q_i)(x_0 - \bar{u})$ when dismissing her.

Hence, the firm in City S has no incentive to keep a poorly-matched worker.

Let D_s be a dummy variable that indicates 1 if a firm in City S dismisses its worker at the beginning of the second period and 0 otherwise. The firm's optimal decision rule is simply expressed as

$$D_S(m_i) = 1 - m_i \ . \tag{4-1}$$

4.3.3.2 Dismissal Decision in City B

In this section, I consider an optimal dismissal decision by a firm in City B. First, consider the case where the firm does not have the right to entitle Hukou to its employee. This case is identical to firms in City S. The optimal decision for this type of firm (type NH) is given by

$$D_{B,NH}(m_i) = 1 - m_i \ . \tag{4-2}$$

Next, consider the case where the firm has the right to entitle Hukou to its employee and uses it strategically to increase its profit. When a match is good, the profit in the second period is $x_1 - \bar{u} + h$ if the firm retains the well-matched worker, and is $q_i(x_1 - \bar{u}) + (1 - q_i)(x_0 - \bar{u})$ if dismisses the worker and hires a new worker. It is easy to see that the firm's optimal decision in this case is to retain the worker.

When a poor matching quality is realized, the firm's optimal decision is not straightforward and depends on the firm's ability *q*. The second period profit is $x_0 - \bar{u} + h$ if the firm retains the worker, and the expected profit is $q_i(x_1 - \bar{u}) + (1 - q_i)(x_0 - \bar{u})$ if it dismisses her. Comparing these two profits yields

$$\begin{cases} x_0 - \bar{u} + h \geq q_i x_1 + (1 - q_i)x_0 - \bar{u} & \text{if } q_i \leq h/(x_1 - x_0) \\ x_0 - \bar{u} + h < q_i x_1 + (1 - q_i)x_0 - \bar{u} & \text{if } q_i > h/(x_1 - x_0) \end{cases} .$$

In other words, the firm dismisses its worker only when the matching quality is poor and the firm's ability to draw a well-matched worker is sufficiently high. The optimal decision of such firms (type WH) can be expressed as

$$D_{B,WH}(m_i, q_i) = (1 - m_i)qd(q_i) \ , \quad (4\text{-}3)$$

where $qd(q_i) = 1$ if $1 \geq q_i > h/(x_1 - x_0)$ and $qd(q_i) = 0$ if $0 \leq q_i \leq \frac{h}{(x_1 - x_0)}$.

4.4 Analysis

4.4.1 Job Turnover

I first compare the dismissal rate of the two cities. A dismissal rate is defined as the probability of dismissing workers at the beginning of the second period. As described in equation (4-1), firms in City S dismiss all poorly-matched workers. The dismissal rate in City S is

$$D_S \equiv \int_0^1 \Pr[D_S(m) = 1] f(q)dq = \int_0^1 (1 - q) f(q)dq = \frac{1}{2} \ . \quad (4\text{-}4)$$

Similarly, the dismissal rate of City B firms that do not possess the right to entitle Hukou is given by

$$D_{B,NH} \equiv \int_0^1 \Pr\left[D_{B,NH}(m) = 1\right] f(q)dq = \frac{1}{2} \ .$$

The dismissal rate of City B firms that possess the right to entitle Hukou is

$$D_{B,WH} \equiv \int_0^1 \Pr\left[D_{B,WH}(m, q) = 1\right] f(q)dq = \int_{\tilde{q}}^1 (1 - q) f(q)dq = \frac{1}{2} - \tilde{q}\left(1 - \frac{1}{2}\tilde{q}\right) \ ,$$

where $\tilde{q} \equiv h/(x_1 - x_0)$. The dismissal rate in City B is now given by

$$D_B \equiv \alpha D_{B,WH} + (1 - \alpha) D_{B,NH} = \frac{1}{2} - \alpha\tilde{q}\left(1 - \frac{1}{2}\tilde{q}\right) \ . \quad (4\text{-}5)$$

Equations (4-4) and (4-5) imply that $D_S \geq D_B$ because $\alpha\tilde{q}\left(1 - \frac{1}{2}\tilde{q}\right) \geq 0$. Therefore, I have the following:

Proposition 4-1. *The dismissal rate of City B is lower than that of City S.*

The optimal behavior of firms in City S does not depend on the firm's ability *q*, and they are able to dismiss a poorly-matched worker regardless of this ability. In contrast, the optimal behavior of City B firms depends on this ability. Those City B firms that use the Hukou strategically to lower wages do not dismiss a poorly-matched worker when their ability to draw a well-matched worker is low. Some inefficient matches between workers and firms are preserved under the allocation system where Hukou are given to firms directly.

4.4.2 Productivity

Assume that the population is the same in both cities, and then the total output measures their productivity. Since there is no difference in the first period total output between the two cities, I can only compare the second period total output.

In City S, for given q, the fraction q of firms drew a good match and produce x_1 in the second period. The fraction $1-q$ of firms hire a new worker in the second period and their expected output is $qx_1+(1-q)x_0$. Therefore, the total expected output of City S in the second period is given by

$$EQ_S=\int_0^1 qx_1 f(q)dq+\int_0^1(1-q)[qx_1+(1-q)x_0]\,f(q)dq=\frac{2}{3}x_1+\frac{1}{3}x_0 \quad . \qquad (4\text{-}6)$$

Next, consider the output level of City B. First, for firms that do not have the right to entitle Hukou to their employee, the case is identical to firms in City S, I have

$$EQ_{B,NH}=\frac{2}{3}x_1+\frac{1}{3}x_0 \quad .$$

Then turn attention to firms that have the right to entitle Hukou to their employee. For given q, the fraction q of firms drew a good match in the first period and produce x_1 in the second period. On the other hand, the fraction $1-q$ of firms drew a poor match in the first period. Those firms that satisfy $\tilde{q}\le q\le 1$ hire a new worker in the second period and their expected output is $qx_1+(1-q)x_0$. Those firms that satisfy

$0 \leq q \leq \tilde{q}$ continue to employ the poorly-matched worker and produce x_0 in the second period. The total expected output is

$$EQ_{B,WH} = \int_0^1 qx_1 f(q)dq + \int_{\tilde{q}}^1 (1-q)[qx_1 + (1-q)x_0]\, f(q)dq + \int_0^{\tilde{q}} (1-q)\, x_0 f(q)d$$

$$= \left(\frac{2}{3} - \frac{1}{2}\tilde{q}^2 + \frac{1}{3}\tilde{q}^3\right) x_1 + \left(\frac{1}{3} + \frac{1}{2}\tilde{q}^2 - \frac{1}{3}\tilde{q}^3\right) x_0 \ . \qquad (4\text{-}7)$$

Finally, the total expected output of City B is given by

$$EQ_B = \alpha EQ_{B,WH} + (1-\alpha)EQ_{B,NH} = \frac{2}{3}x_1 + \frac{1}{3}x_0 + \alpha\left(\frac{1}{3}\tilde{q} - \frac{1}{2}\right)\tilde{q}^2(x_1 - x_0) \ . \qquad (4\text{-}8)$$

Since $0 \leq \tilde{q} \leq 1$, the third term in equation (4-8) is negative. I have the following implication:

Proposition 4-2. *The total output of City B is smaller than that of City S.*

Some firms in City B retain a worker even if a poor matching quality between the firm and worker is realized. This inefficient match lowers the aggregate productivity of City B, and the third term in equation (4-8) captures a loss of total output due to the Hukou allocation system. Note that the third term in equation (4-8) can be written as $\alpha\left[\frac{1}{3}\cdot\frac{h}{(x_1-x_0)} - \frac{1}{2}\right]\left[\frac{h^2}{(x_1-x_0)}\right]$. Loss of total output increases with α and h. Since the main source of the distortion is the Hukou allocation system, the loss of total output increases when more firms can strategically use the advantage to construct their labor contract. A higher level of utility associated with Hukou allows more firms to retain a poorly-matched worker, which also results in a larger loss of total output. On the other hand, the loss of total output decreases with $x_1 - x_0$. Retaining a poorly-matched worker forgoes a possibility of hiring a well-matched worker and producing high output level x_1, whereas its benefit comes from keeping a low wages. When a gap between x_1 and x_0 is large, the cost outweighs the benefit, and firms avoid retaining a poorly-matched worker.

4.4.3 Profit

This section compares firm profits of both cities. In doing so, it is assumed that the law of large numbers is satisfied. Since labor is the only factor of production in this model, firm profit is simply defined as the total output minus the total wage payment through two periods.

In City S, the fraction q of firms drew a good match and the total expected output is $\int_0^1 q x_1 f(q)dq + \int_0^1 (1-q)x_0\, f(q)dq = \frac{1}{2}x_1 + \frac{1}{2}x_0$ in the first period, and the total expected output in the second period is $\frac{2}{3}x_1 + \frac{1}{3}x_0$ as described in equation (4-6). Taking the optimal wage $\bar{u}$ in both periods as given, the total profit of firms in City S is

$$\Pi_S = \left(\frac{1}{2}x_1 + \frac{1}{2}x_0 - \bar{u}\right) + \left(\frac{2}{3}x_1 + \frac{1}{3}x_0 - \bar{u}\right) = \frac{7}{6}x_1 + \frac{5}{6}x_0 - 2\bar{u}\ . \qquad (4\text{-}9)$$

Next, in City B, the total profit of firms that do not possess the right is identical to firms in City S, and then I have

$$\Pi_{B,NH} = \frac{7}{6}x_1 + \frac{5}{6}x_0 - 2\bar{u}.$$

On the other hand, for firms in City B that possess the right, the total expected output in the first period is $\frac{1}{2}x_1 + \frac{1}{2}x_0$ similarly, and the total expected output in the second period is $\left(\frac{2}{3} - \frac{1}{2}\tilde{q}^2 + \frac{1}{3}\tilde{q}^3\right)x_1 + \left(\frac{1}{3} + \frac{1}{2}\tilde{q}^2 - \frac{1}{3}\tilde{q}^3\right)x_0$ as given in equation (4-7). Since firms have the right to entitle Hukou to their employee, the total wage in the first period is $\bar{u} - h$. In the second period, the q fraction firms that drew a good match in the first period keep the well-matched worker at wage rate $\bar{u} - h$. The fraction $1 - q$ of firms drew a poor match in the first period, those firms that satisfy $0 \le q \le \tilde{q}$ continue to employ the poorly-matched worker and their wage payment is still $\bar{u} - h$, but those firms that satisfy $\tilde{q} \leq q \leq 1$ hire a new worker in the second period and their wage payment becomes $\bar{u}$. Then the total wage in the second period is $\int_0^1 q(\bar{u}-h)f(q)dq + \int_0^{\tilde{q}}(1-q)(\bar{u}-h)\,f(q)dq + \int_{\tilde{q}}^1(1-q)\bar{u}\,f(q)dq = \bar{u} - \left(\frac{1}{2}+\tilde{q}-\frac{1}{2}\tilde{q}^2\right)h$. The total profit of firms that possess the right is

$$\Pi_{B,WH} = \left[\frac{1}{2}x_1 + \frac{1}{2}x_0 - (\bar{u}-h)\right] + \left[\left(\frac{2}{3}-\frac{1}{2}\tilde{q}^2+\frac{1}{3}\tilde{q}^3\right)x_1 + \left(\frac{1}{3}+\frac{1}{2}\tilde{q}^2-\frac{1}{3}\tilde{q}^3\right)x_0 - \bar{u} + \left(\frac{1}{2}+\tilde{q}-\frac{1}{2}\tilde{q}^2\right)h\right]$$

$$= \frac{7}{6}x_1 + \frac{5}{6}x_0 - 2\bar{u} + \left(\frac{1}{3}\tilde{q}-\frac{1}{2}\right)\tilde{q}^2(x_1-x_0) + \left(\frac{3}{2}+\tilde{q}-\frac{1}{2}\tilde{q}^2\right)h \; .$$

Then, the total profit of firms in City B is

$$\Pi_B = \alpha\Pi_{B,WH} + (1-\alpha)\Pi_{B,NH}$$

$$= \frac{7}{6}x_1 + \frac{5}{6}x_0 - 2\bar{u} + \alpha\left[\left(\frac{1}{3}\tilde{q}-\frac{1}{2}\right)\tilde{q}^2(x_1-x_0) + \left(\frac{3}{2}+\tilde{q}-\frac{1}{2}\tilde{q}^2\right)h\right] \; . \quad (4\text{-}10)$$

Note that the forth term in equation (4-10) is positive because it can be written as $\left(\frac{1}{3}\tilde{q}-\frac{1}{2}\right)\tilde{q}h + \left(\frac{3}{2}+\tilde{q}-\frac{1}{2}\tilde{q}^2\right)h$ and $\left(\frac{1}{3}\tilde{q}-\frac{1}{2}\right)\tilde{q} \geq -1$ and $\frac{3}{2}+\tilde{q}-\frac{1}{2}\tilde{q}^2 \geq \frac{3}{2}$. This generates the following implication:

Proposition 4-3. *The total profit of City B is larger than that of City S.*

While the first term in the square bracket of equation (4-10) reflects a productivity loss, its second term is a profit gain arising from wage saving. Proposition 4-3 says that this profit gain exceeds the productivity loss, and it confirms that a lower level of productivity is justified at the firm level by larger profits.

4.5 Conclusion of Chapter 4

I provided an analysis that helps understand the effect of the Hukou allocation system on productivity, and offered an alternative explanation for an observed difference in firm productivity between Beijing and Shanghai. This theoretical analysis indicated that the Beijing's Hukou allocation system results in keeping an inefficient match between workers and firms, which, as argued before, is a source of lower productivity in Beijing. This chapter demonstrated that the way of allocating Hukou affects not only worker's welfare, but also regional development.

Note that for simplicity I assumed that all workers are identical and wage renegotiation is not allowed. In a more realistic model, however, workers would possess different level of ability and also have the right of wage renegotiation. This may affect some conclusions of this analysis because firms and workers likely act so as to influence bargaining power and it may in turn change the nature of the mismatch problem I discussed above. Nonetheless, the model in this chapter captures the fundamental relationship between Hukou allocation system and productivity, and provides meaningful implications for economic development policies.

Chapter 5.

Conclusion

In this dissertation, I investigated the effects of new entry, resources misallocation, and the Hukou policy on productivity in order to measure real technical improvement more precisely by distinguishing physical productivity and revenue productivity. More specifically, I proposed a theoretical approach to examine the entry effect on productivity by controlling for both supply and demand factors, and tried to separate the physical productivity improvement from the revenue productivity change. I also used firm-level data to examine how observed regional productivity differences are due to resource misallocation. The analysis of this dissertation is expected to improve our understanding of sources and mechanisms that determine productivity and to have important policy implications for regional development. The proposed methodology is also expected to help measure technical efficiency improvement more precisely.

In Chapter 2, I examined how new entry influences incumbent firm's measured productivity when the highest quality product is introduced to the market by incorporating both demand and supply factors into a single analytical framework. First I built a theoretical model by extending the model of Johnson and Myatt (2003) where both consumers and firms take product quality into account when they decide their optimal behaviors. The extended model allowed me to link physical productivity and revenue productivity under several types of new entry. The key insight form this analysis is that incumbent firm's revenue productivity can be affected by both business stealing effects and technical improvement effects. The theoretical predictions generated from this analysis need to be examined against data in order to be more convincing. One of the future tasks is to examine the relationship between revenue productivity and physical productivity in this analysis by using real data.

In Chapter 3, I used firm-level Chinese manufacturing data to investigate how input market distortions affect the aggregate productivity differences between two major cities in China, Beijing and Shanghai. In this empirical analysis, I employed an extended version of Hsieh and Klenow (2009) approach and an alternative approach developed from Midrigan and Xu (2014) to estimate productivity losses from resource distortions. This empirical analysis revealed that the aggregate productivity level is lower in Beijing than that in Shanghai, and the input market distortions, especially the capital misallocation is more severe in Beijing than that in Shanghai. A limitation of this empirical analysis is that a source of misallocation cannot be identified because of data limitations. However, the findings from this study still can provide important policy implications for economic development.

In Chapter 4, I offered a possible mechanism that generates a regional productivity difference between Beijing and Shanghai through labor market misallocation. I constructed a theoretical model to explain a possible role of Hukou allocation system (a unique household registration policy in China) in influencing regional productivity through firm's strategic behaviors with respect to the retention of workers. The theoretical analysis showed that the level of regional productivity can be low due to inefficient labor allocation arising from Hukou system, and that the way of allocating Hukou is a source of labor market distortions and productivity differences across regions. In this model, I assumed that all workers are identical and wage renegotiation is not allowed for simplicity, but a more realistic analysis needs to be done by relaxing these assumptions.

Although this dissertation research has some limitations, it shed a light on several important issues of the productivity determinants and made it clear that productivity level is influenced significantly by new entry, resource misallocation, and economic policies.

References

Abbott, T.A., 1992, “Price Dispersion in U.S Manufacturing Implications for the Aggregation of Products and Firms,” *US Census Bureau Working Paper 92-3*.

Afridi, F., Li, S.X. and Ren, Y., 2012, “Social Identity and Inequality: The Impact of China's Hukou System,” *IZA Discussion Paper*, No. 6417.

Aghion, P., Bloom, N., Blundell, R., Griffith, R. and Howitt, P., 2005, “Competition and Innovation: An Inverted U Relationship,” *Quarterly Journal of Economics,* 120(2), pp. 701-728.

Amaral, P.S. and Quintin, E., 2010, “Limited Enforcement, Financial Intermediation, and Economic Development,” *International Economic Review*, 51, pp. 785-811.

Arnold, J., Brys, B., Heady, C., Johansson, A., Schwellnus, C. and Vartia, L., 2011, “Tax Policy for Economic Recovery and Growth,” *The Economic Journal*, 121, pp. 59-80.

Asplund, M. and Nocke, V., 2006, “Firm Turn- over in Imperfectly Competitive Markets,” *Review of Economic Studies*, 73(2), pp. 295-327.

Atkeson, A., Khan, A. and Ohanian, L.E., 1996, “Are Data on Industry Evolution and Gross Job Turnover Relevant For Macroeconomics?” *Carnegie-Rochester Conference Series on Public Policy*, 44, pp. 216-250.

Aw, B.Y., Chen, X. and Roberts, M.J., 2001, “Firm-Level Evidence on Productivity Differentials and Turnover in Taiwanese Manufacturing,” *Journal of Development Economics*, 66(1), pp. 51-86.

Baily, M.N., Hulten, C. and Campbell, D., 1992, “Productivity Dynamics in Manufacturing Plants,” *Brooking Papers: Microeconomics*, pp. 187-249.

Baldwin, J.R. and Gu, W., 2006, “Plant Turnover and Productivity Growth in Canadian Manufacturing,” *Industrial and Corporate Change*, 15(3), pp. 417-465.

Bartelsman, E.J., Haltiwanger, J. and Scarpetta, S., 2004, “Microeconomic Evidence of Creative Destruction in Industrial and Developing Countries,” *IZA Discussion Papers* No.1374.

Becker, B. and Milbourn, T., 2011, "How Did Increased Competition Affect Credit Ratings?" *Journal of Financial Economics*, 101(3), pp. 493-514.

Bellone, G., Smirne, C., Mauri, F.A., Tonel, E., Carbone, A., Buffolino, A., Dughera, L., Robecchi, A., Pirisi, M. and Emanuelli, G., 2006, "Cytokine Expression Profile in Human Pancreatic Carcinoma Cells and in Surgical Specimens: Implications for Survival," *Cancer Immunol Immunother*, 55(6), pp. 684-98.

Bennett, V.M., Pierce, L., Snyder, J.A. and Toffel, M.W., 2013, "Customer-Driven Misconduct: How Competition Corrupts Business Practices," *Management Science*, 59(8), pp. 1725-1742.

Bernard, A.B., Redding, S. and Schott, P., 2007, "Comparative Advantage and Heterogeneous Firms," *Review of Economic Studies*, 74(1), pp. 31-66.

Besley, T. and Ghatak, M., 2008, "Status Incentives," *American Economic Review*, 98, pp. 206-211.

Beveren, I.V., 2012, "Total Factor Productivity Estimation: A Practical Review," *Journal of Economic Surveys*, 26(1), pp. 98-128.

Bijlsma, M., Koning, P., Shestalova, V. and Aouragh, A., 2010, "The Effect of Competition on Process and Outcome Quality of Hospital Care: An Empirical Analysis for the Netherlands," *CPB Discussion Paper No. 157*.

Bloom, N. and Reenen, J.V., 2010, "Why Do Management Practices Differ across Firms and Countries?" *The Journal of Economic Perspectives*, 24(1), pp. 203-224.

Blundell, R., Griffith, R. and Van Reenen, J., 1995, "Dynamic Count Data Models of Technological Innovation," *The Economic Journal*, 105(492), pp. 333-344.

Boyreau-Debray, G. and Wei, S.J., 2004, "Can China Grow Faster? A Diagnosis on the Fragmentation of the Domestic Capital Market," *IMF Working Papers*, International Monetary Fund, 04/76.

Brandt, L. and Rawski, T.G., 2008, "China's Great Economic Transformation," Cambridge and New York: Cambridge University Press.

Brandt, L., Van Biesebroeck, J. and Zhang, Y., 2012, "Creative Accounting or Creative Destruction? Firm-level Productivity Growth in Chinese Manufacturing," *Journal of Development Economics*, 97(2), pp. 339-351.

Broda, C. and Weinstein, D.E., 2006, "Globalization and the Gains from Variety," *Quarterly Journal of Economics*, 121(5), pp. 541-585.

Cai, F., Wang, D. and Du, Y., 2002, "Regional Disparity and Economic Growth in China: The Impact of Labor Market Distortions," *China Economic Review*, 13, pp. 197-212.

Cameron, G., 2003, "Why did UK Manufacturing Productivity Growth Slowdown in the 1970s and Speed Up in the 1980s?" *Economica*, 70(1), pp. 121-141.

Chan, K.W., 1992, "Economic Growth Strategy and Urbanization Policies in China, 1949–82," *International Journal of Urban and Regional Research*, 16, pp. 275-305.

Chan, K.W. and Yang, Y., 1996, "Inter-provincial Migration in China in the post-1949 Era: Types, Spatial Patterns, and Comparisons," *Seattle, WA: Seattle Population Research Center Working Paper*, No. 96-14.

Chan, K.W. and Zhang, L., 1999, "The Hukou System and Rural-urban Migration: Processes and Changes," *The China Quarterly*, 160, pp. 818-855.

Chari, V.V., Kehoe, P.J. and McGrattan, E.C., 2007, "Business Cycle Accounting," *Econometrica*, 75(3), pp. 781-836.

Cheng, T. and Selden, M., 1994, "The Origin and Social Consequences of China's Hukou System," *The China Quarterly*, 139, pp. 644-668.

Ciccone, A. and Hall, R., 1996, "Productivity and the Density of Economic Activity," *American Economic Review*, 86, pp. 54-70.

Correa, J.A. and Ornaghi, C., 2014, "Competition & Innovation: Evidence from U.S. Patent and Productivity Data," *The Journal of Industrial Economics*, 62(2), pp. 258-285.

Deltas, G., Stengos, T. and Zacharias, E., 2011, "Product Line Pricing in a Vertically Differentiated Oligopoly," *Canadian Journal of Economics*, 44(3), pp. 907-929.

Deng, P. and Jefferson, G., 2011, "Explaining Spatial Convergence of China's Industrial Productivity," *Oxford Bulletin of Economics and Statistics*, 73, pp. 818-832.

Diewert, W.E. and Nakamura, A.O., 2007, "The Measurement of Productivity for Nations," *Handbook of Econometrics*, Vol. 6A, Chapter 66, Elsevier.

Disney, R., Haskel, J. and Heden, Y., 2003, "Restructuring and Productivity Growth in UK Manufacturing," *Economic Journal*, 113(489), pp. 666-694.

Dunn, A., 2008, "Do Low-quality Products Affect High-quality Entry? Multiproduct Firms and Nonstop Entry in Airline Markets," *International Journal of Industrial Organization*, 26(5), pp. 1074-1089.

Eckel, C. and Neary, J.P., 2010, "Multi-Product Firms and Flexible Manufacturing in the Global Economy," *Review of Economic Studies*, 77(1), pp. 188-217.

Ericson, R. and Pakes, A., 1995, "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies*, 62(1), pp. 53-82.

Eslava, M., Haltiwanger, J., Kugler, A. and Kugler, M., 2004, "The Effects of Structural Reforms on Productivity and Profitability Enhancing Reallocation: Evidence from Colombia," *Journal of Development Economics*, 75(2), pp. 333-371.

Eslava, M., Haltiwanger, J., Kugler, A. and Kugler, M., 2013, "Trade Reforms and Market Selection: Evidence from Manufacturing Plants in Colombia," *Review of Economic Dynamics*, 16, pp. 135-158.

Fan, Y. and Yang, C., 2016, "Competition, Product Proliferation and Welfare: A Study of the U.S. Smartphone Market," *NET Institute Working Paper*, No.14-14.

Feenstra, R.C. and Ma, H., 2008, "Optimal Choice of Product Scope for Multiproduct Firms under Monopolistic Competition," The Organization of Firms in a Global Economy, edited by Helpman, E., Marin, D. and Verdier, T., Harvard University Press,

Feenstra, R., Li, Z. and Yu, M., 2014, "Exports and Credit Constraints under Incomplete Information: Theory and Evidence from China," *The Review of Economics and Statistics*, 96(4), pp. 729-744.

Foster, L., Haltiwanger, J. and Krizan, C.J., 2001, "Aggregate Productivity Growth: Lessons from Microeconomic Evidence," *New Developments in Productivity analysis*, edited by Hulten, C.R., Dean, E.R. and Harper, M.J., University of Chicago Press, pp. 303-363.

Foster, L., Haltiwanger, J. and Krizan, C.J., 2006, "Market Selection, Reallocation, and Restructuring in the US Retail Trade Sector in the 1990s," *The Review of Economics and Statistics*, 88(4), pp. 748-758.

Foster, L., Haltiwanger, J. and Syverson, C., 2008, "Reallocation, Firm Turnover and Efficiency: Selection of Productivity or Profitability?" *American Economic Review*, 98(1), pp. 394-425.

Friedberg, M.R. and Hunt, J., 1995, "The Impact of Immigrants on Host Country Wages, Employment and Growth," *The Journal of Economic Perspectives*, 9, pp. 23-44.

Gal-Or, E., 1983, "Quality and Quantity Competition," *Bell Journal of Economics*, 14(2), pp. 590-600.

Gaynor, M. and Town, R.J., 2011, "Competition in Health Care Markets," *NBER Working Paper No. 17208*.

Geroski, P.A., 1990, "Innovation, Technological Opportunity, and Market Structure," *Oxford Economic Papers*, 42(3), pp. 586-602.

Geroski, P.A., 1995, "What Do We Know About Entry?" *International Journal of Industrial Organization*, 13(4), pp. 421-440.

Goldstein, S. and Goldstein, A., 1991, "Permanent and Temporary Migration Differentials in China," *Honolulu, Hawaii, East–West Center, Papers of the East–West Population Institute*, No. 117.

Gong, G. and Hu, G., 2016, "The Role of Returns to Scale in Measuring Frictions in Resource Allocation: Revisiting Misallocation and Manufacturing TFP in China," *Economics Letters*, 138, pp. 26-29.

Griffith, R., 2001, "Product Market Competition, Efficiency and Agency Costs: An Empirical Analysis," *IFS working paper*.

Griffith, R., Harrison, R. and Simpson, H., 2010, "Product Market Reform and Innovation in the EU," *Scandinavian Journal of Economics*, 112(2), pp. 389-415.

Griliches, Z. and Regev, H., 1995, "Productivity and Firm Turnover in Israeli Industry: 1979- 1988," *Journal of Econometrics*. 65(1), pp. 175-203.

Guner, N., Ventura, G. and Xu, Y., 2008, "Macroeconomic Implications of Size-dependent Policies," *Review of Economic Dynamics,* 11(4), pp. 721-744.

Haltiwanger, J., 1997, "Measuring and Analyzing Aggregate Fluctuations: The Importance of Building from Microeconomic Evidence," *Review of the Federal Reserve Bank of St. Louis*. 79(3), pp. 55-78.

Harris, R. and Li, Q.C., 2008, "Exporting, R&D, and Absorptive Capacity in UK Establishments," *Oxford Economic Papers*, 61(1), pp. 74-103.

Hendel, I. and Nevo, A., 2006, "Sales and Consumer Inventory," *RAND Journal of Economics,* 37(3), pp. 543-561.

Holz, C.A., 2009, "No Razor's Edge: Reexamining Alwyn Young's Evidence for Increasing Interprovincial Trade Barriers in China," *The Review of Economics and Statistics,* 91(3), pp. 599-616.

Hopenhayn, H.A., 1992, "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica,* 60(5), pp. 1127-1150.

Hsieh, C-T. and Klenow, P.J., 2009, "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 74(4), pp. 1403-1448.

Hsieh, C-T. and Klenow, P.J., 2014, "The Life Cycle of Plants in India and Mexico," *Quarterly Journal of Economics*, 129(3), pp. 1335-1384.

Ishida, J., 2005, "Life Employment as A Coordination Failure," *Japan and the World Economy,* 17, pp. 209-222.

Itoh, M., 1983, "Monopoly, Product Differentiation and Economic Welfare," *Journal of Economic Theory,* 31(1), pp. 88-104.

Johnson, J.P. and Myatt, D.P., 2003, "Multiproduct Quality Competition: Fighting Brands and Product Line Pruning," *American Economic Review*, 93(3), pp. 748-774.

Jovanovic, B., 1982, "Selection and the Evolution of Industry," *Econometrica*, 50(3), pp. 649-70.

Katayama, H., Lu, S. and Ttybout, J., 2003, "Why Plant-Level Productivity Studies Are Often Misleading, and an Alternative Approach to Interference," *National Bureau of Economic Research Working Paper No.9617*.

Kessler, D. and McClellan, M., 2000, "Is Hospital Competition Socially Wasteful?" *Quarterly Journal of Economics*, 115(2), pp. 577-615.

Klette, T.J. and Griliches, Z., 1996, "The Inconsistency of Common Scale Estimators when Output Prices Are Unobserved and Endogenous," J*ournal of Applied Econometrics*, 11(4), pp. 343-361.

Levinsohn, J. and Petrin, A., 2003, "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70(2), pp. 317-342.

Li, H. and Haynes, E.K., 2011, "Economic Structure and Regional Disparity in China: Beyond the Kuznets Transition," *International Regional Science Review*, 34, pp. 157-190.

Liu, Z., 2005, "Institution and Inequality: The Hukou System in China," *Journal of Comparative Economics*, 33, pp. 133-157.

Loecker, D., 2011, "Product Differentiation, Multi-Product Firms and Structural Estimation of Productivity," *Econometrica*, 79(5), pp. 1407-1451.

Ma, C.A. and Burgess, J.F., 1993, "Quality Competition, Welfare, and Regulation," *Journal of Economics*, 58(2), pp. 153-173.

Ma, L.J.C., 1996, "The Spatial Patterns of Interprovincial Rural-to-urban Migration in China, 1982–1987," *Chinese Environment and Development*, 7, pp. 73-102.

Ma, Z., 1999, "Temporary Migration and Regional Development in China," *Environment and Planning A*, 31, pp. 783-802.

Mairesse, J. and Jaumandreu, J., 2005, "Panel-Data Estimates of the Production Function and the Revenue Function: What Difference Does It Make?" *Scandinavian Journal of Economics*, 107(4), pp. 651-672.

Manez-Castillejo, J.A., 1999, "Price Competition and Price Dispersion in the UK Supermarkets," *Working paper*, Departamento de Economa Aplicada II Universitat de Valencia.

Martin, S., 1993, "Endogenous Firm Efficiency in a Cournot Principal Agent Model," *Journal of Economic Theory*, 59(2), pp. 278-283.

Matsa, D.A., 2011, "Competition and Product Quality in the Supermarket Industry," *The Quarterly Journal of Economics*, 126(3), pp. 1539-1591.

Mayer, T., Melitz, M.J. and Ottaviano, I.P., 2014, “Market Size, Competition, and the Product Mix of Exporters,” *American Economic Review*, 104(2), pp. 495-536.

McMaster, R., 1995, “Competitive Tendering in UK Health and Local Authorities: What Happens to the Quality of Services?” *Scottish Journal of Political Economy*, 42(4), pp. 409-427.

Melitz, M.J., 2000, “Estimating Firm-Level Productivity in Differentiated Product Industries,” Harvard, mimeo.

Melitz, M.J., 2003, “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” *Econometrica*, 71(6), pp. 1695-1725.

Midrigan, V. and Xu, D.Y., 2014, “Finance and Misallocation: Evidence from Plant-Level Data,” *American Economic Review*, 104(2), pp. 422-58.

Mussa, M. and Rosen, S., 1978, “Monopoly and Product Quality,” *Journal of Economic Theory*, 18, pp. 301-317.

Nishimura, M., Kusaba, M., Miyahara, K., Nishio, T., Iida, S., Imbe, T. and Sato, H., 2005, “New Rice Varieties with Low Levels of Easy-to-digest Protein, “LGC-Katsu” and “LGC-Jun”,” *Breeding Science*, 55(1), pp. 103-105.

Olley, G.S. and Pakes, A., 1996, “The Dynamics of Productivity in the Telecommunications Equipment Industry,” *Econometrica*, 64(6) pp. 1263-1297.

Pavcnik, N., 2002, “Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants,” *Review of Economic Studies*, 69(1), pp. 245-276.

Restuccia, D. and Rogerson, R., 2008, “Policy Distortions and Aggregate Productivity with Heterogeneous Plants,” *Review of Economic Dynamics*, 11(4), pp. 707-720.

Rizov, M. and Zhang, X., 2014, “Regional Disparities and Productivity in China: Evidence from Manufacturing Micro Data,” *Papers in Regional Science*, 93, pp. 321-339.

Rose, N.L., 1990, “Profitability and Product Quality: Economic Determinants of Airline Safety Performance,” *Journal of Political Economy*, 98(5), pp. 944-964.

Ruhs, M., 2008, "Economic Research and Labour Immigration Policy," *Oxford Review of Economic Policy*, 24, pp. 403-426.

Scarpetta, S., Hemmings, P., Tressel, T. and Woo, J., 2002, "The Role of Policy and Institutions for Productivity and Firm Dynamics: Evidence from Micro and Industry Data," *Economics Department Working Paper No. 329*, Paris: OECD.

Seim, K. and Viard, B., 2011, "The Effect of Market Structure on Celluar Technology Adoption and Pricing," *American Economic Journal: Microeconomics*, 3(2), pp. 221-251.

Smiley, R., 1988, "Empirical Evidence on Strategic Entry Deterrence," *International Journal of Industrial Organization*, 6(2), pp. 167-180.

Sohn, M.W. and Rathouz, P.J., 2003, "Competition among Hospitals and Quality of Care: Hospital-level Analysis," *Unpublished manuscript*, University of Chicago.

Syverson, C., 2004, "Market Structure and Productivity: A Concrete Example," *Journal of Political Economy*, 112(6), pp. 1181-1222.

Syverson, C., 2011, "What Determines Productivity?" *Journal of Economic Literature*, 49(2), pp. 326-365.

Trefler, D., 2004, "The Long and Short of the Canada–US Free Trade Agreement," *American Economic Review*, 94(4), pp. 870-895.

Yang, Y., Xu, Y. and Xiang, S., 2003, "Jiuye tidai yu laodongli liudong: yige xindi fenxi kuangjia [Employment Replacement and Labor Migration: A New Analytical Framework] ," *Jingji Yanjiu*, 8, pp. 70-75.

Young, A., 2000, "The Razor's Edge: Distortions and Incremental Reform in the People's Republic of China," *The Quarterly Journal of Economics*, 115(4), pp. 1091-1135.

Zhang, R., Sun, K., Delgado, M. and Kumbhakar, S., 2012, "Productivity in China's High Technology Industry: Regional Heterogeneity and R&D," *Technological Forecasting and Social Change*, 79, pp. 127-141.