THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Essays on the Missing Middle

Pham Thu Hien
BA, MIntDeSt(Adv)

A thesis submitted for the degree of Doctor of Philosophy at
The University of Queensland in 2016
School of Economics

Abstract

The missing middle refers to the empirical fact that most employment in developing countries is located in either small-sized or large-sized firms. In this thesis,we investigated the missing-middle phenomenon in a developing country, Vietnam. The first chapter consists of an introduction. The second chapter provides an overall picture of related literature. The third chapter focuses on the development of a theoretical model designed to capture the effect of corruption on the missing-middle formulation. The fourth chapter provides empirical evidence on the existence of a distinct distribution of firm size in Vietnam and some insights into the underlying mechanism of the phenomenon, focusing on corruption and productivity. The fifth chapter focuses on the industry efficiency score. In general, the fourth chapter is empirically oriented, whereas the third and fifth chapters are theoretically oriented. In particular, in the third chapter, we develop a theoretical model to capture an effect of corruption in the development of firms. The model incorporates the increasing costs of large-sized firms, compared to smaller ones. This model shows the firm size distribution (FSD) normally observed in developing economies along a continuum of infinitesimal firms in the market. In the fourth chapter, we discuss the empirical relationship between firm size and production efficiency, inefficiencies associated with scale, and the relationship between firm size and the likelihood of paying bribes, using firm-level data from Vietnam. Our analysis indicates that middle-sized firms production efficiencies tend to be lower than the efficiencies in small-sized or large-sized firms in most manufacturing industries. In addition, the least-efficient firm size differs across sectors, indicating the necessity of sector-by-sector analyses to study the underlying mechanisms of the missing-middle phenomenon. Moreover, our results show that as firm size increases, the likelihood of paying bribes also increases. This finding, together with our productivity analysis showing that small-sized firms produce more efficiently compared to middle- or large-sized firms, indicates that corruption can be one of the reasons for firms to decide not to grow. We analyzed production efficiency at the aggregate level and provide an approximation proposition that simplifies the computation and interpretation of various industry inefficiency measures proposed in the literature. Moreover, we extended the proposed framework to the meta-technology approach. We found a difference between the group-frontiers and the meta-frontiers. In addition, the meta-technology efficiency (MTE) scores were highest for small-sized firms, in particular, firms with fewer than 50 employees. This finding confirms the differences in the characteristics of firms operating environment across size groups.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

No publications.

Publications included in this thesis

No publications included.

Contributions by others to the thesis

An extract of Chapter 4 of the thesis has been submitted to the Centre for Efficiency and Productivity Analysis with Dr Shino Takayama co-authors.

An extract of Chapter 5 of the thesis has been submitted to the Centre for Efficiency and Productivity Analysis with Dr Peyrache Antonio co-authors.

Statement of parts of the thesis submitted to qualify for the award of another degree

None.

Acknowledgments

No one walks alone in the PhD journey. Those who joined me, helped me, and inspired me along the path have made the last four years an unforgettable experience. This acknowledgment conveys my deep gratitude and appreciation to the many people who made this thesis possible. The Ph.D. journey would not have been the same without you.

First,I would like to express my deep gratitude to the University of Queensland for providing me with the opportunity to pursue a postgraduate education. The University of Queensland, and the School of Economics, in particular, has been an engaging and inspiring environment and has meant a lot to me and to the completion of this thesis.

I am equally indebted to my principle advisor, **Dr. Shino Takayama**. It has been an honor to be her Ph.D. student. She has taught me how good research is conducted. I appreciate all her contributions of time, ideas to make my Ph.D. experience productive and stimulating Her dedication and tough evaluation of my work enabled me to raise the bar and encouraged me to pursue higher targets. I am also thankful for the excellent example she has provided as a successful woman. I look forward to continuing our joint work in the future.

Further, I would like to express deep feelings of gratitude to my second advisor, **Dr. Antonio Payrache**, for his valuable guidance, especially during the final stages of this research program.

I also gratefully acknowledge other members of my Ph.D. committee, **Dr. Le Thanh Quang and Dr. Yuichiro Waki**, for their time and valuable feedback on this thesis. It would have been quite impossible to shape it into the final thesis without their help and effort.

I owe deep gratitude to **Prof. Andy McLennan, Prof. Valentin Zelenyuk**, and **Prof. Flavio Menezes** for valuable comments on my research project. Their assessments, comments, and suggestions were significant to improving this thesis. I would also thank **Fabrizio Carmignani, Begona Dominguez, Jeff Kline, Do Won Kwak, Hideyuki Mizobuchi, Rodney Strachan, Satoshi Tanaka, Haishan Yuan**, and other staff members in the School of Economics, University of Queensland, and all my Ph.D. fellows for their encouragement and support.

Special thanks go to the General Statistic Office of Vietnam and Center Institute for Economic Management for giving me permission to use the micro data of Vietnamese enterprises.

On a personal note, I must also express my deep respect and appreciation for my parents, **Pham Van Thao** and **Pham Thi Thuy Hang**, my sisters **Pham Thi Kieu Phuong**, **Pham Thi Phuong**

Anh. I do not know how to thank you for supporting me in all possible ways over the last four years, for encouraging me to pursue my dream, for sharing all my burdens, and for loving me.

I especially thank my husband **Nguyen Hai Anh** for the years of loving support. You have always been my source of motivation, my strength to overcome all challanges. It is to you that I owe my best achievements as a researcher. Also, many thanks to my children, **Nguyen Phuong Linh - Gau** and **Nguyen Son - Charlie** for being always beside me. You have brightened my evenings after each exhausting day and provided me with fresh motivation to complete my thesis. I only hope you will the good fortune to meet as many good people as I have had over the course of my studies in the University of Queensland.

This book is dedicated to you!

Keywords

firm size distribution, missing middle, productivity, corruption, efficiency, aggregation, data envelopment analysis, free disposal hull.

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 140102, Macroeconomic Theory, 33% ;

ANZSRC code: 140209, Industry Economics and Industrial Organisation, 33%;

ANZSRC code: 140302, Econometric and Statistical Methods, 34%.

Fields of Research (FoR) Classification

FoR code: 1401. Economic Theory, 33%

FoR code: 1402, Applied Economics, 33%

FoR code: 1403, Econometrics, 34%

Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Background

Firm size distribution (FSD) has been a particular concern of economists for nearly a century. On entering markets, firms face a selection process that causes less efficient firms to decrease their size and eventually exit, while more efficient firms start to grow. Survivors will then choose to allocate resources to maximize their profit in the given macro environment and within a certain size group. In other words, the accumulation of the underlying dynamics that results from a firm's decision regarding entry, development, or exit evolves into a certain firm size distribution. FSD is, therefore, an endogenous choice for firms, and an analysis of FSD could provide a possible interpretation of the behavior of firms in the market, given their opportunities, constraints, and efficiency levels. FSD could also contribute an explanation to the industrial concentration of an economy, which in turn represents the level of competition in the market. If small firms grow at a high rate, the increase in number of firms will lead to an improvement in market competitiveness. Conversely, a larger proportion of large firms may create monopolistic and oligopolistic markets in which a few giant firms employ a large number of employees and exercise significant market power. Consequently, small changes in the distribution of these firms can have considerable macroeconomic impact on labor policies, labor market evolution, and job creation and destruction.

Recently, there has been an increasing interest in FSD analysis. In recent studies, many important features of the FSD of developed economies have been confirmed. First, empirical studies have provided evidence of the existence of a relatively stationary distribution of firm size under various economic scenarios. Further, different measures of size do not alter the distribution in these scenarios. Such a distribution exhibits a robust right skew over time, regardless of changes in diversified regulatory environments, waves of mergers and acquisitions, and surges of firm entry and exit. More importantly, some significant FSD laws in developed economies have also been identified. One is Gibrats law, also known as the law of proportionate effect. According to this law, the probability of a given proportional change in size during a specified period is the same for all firms in an industry, regardless of the initial size at the beginning of the period. As a result, firm size can be explained purely by its idiosyncratic accumulation of multiplicative random shocks. In the literature, this law has provided a benchmark for both empirical and theoretical studies of FSD in developed economies. Another law is Zipf's law, in which the frequency of a certain size firm is inversely proportional to its rank in the frequency table.

The FSDs of developing economies are characterized with quite different features. That is, while FSDs of developed countries are characterized by the Pareto law or log-normal distribution, the size structure of firms in developing countries tends to deviate from such a distribution. In particular, a concentration of infinitesimal firms co-exits with a handful of large-sized firms and an under-representation of middle-sized firms. This phenomenon in developing countries is called the "missing middle". The missing middle was first documented in Liedholm and Mead (1987). In addition, Tybout (2000) found a large spike in the size distribution for the small-sized category, which drops off quickly in the middle-sized category among the poorest countries. (See also Little et al. (1987) for South Korea and Steel and Webster (1992) for Ghana). The presence of the missing-middle phenomenon and how it is defined have been extensively discussed in the literature (see Hsieh and Olken, 2014, Tybout, 2014a,b).

Because the missing middle is more evident in developing nations, we imagine there are some reasons that particularly apply to developing nations, but not so much to developed nations, likely because the business environment varies across nations, and particularly between developing nations and developed nations. In particular, we emphasize the mis-allocation of resources among firms as a source of differences in aggregate productivity, which may lead to such a FSD. One possible explanation is that this mis-allocation takes place because of government policies or other barriers that distort the way resources are allocated across heterogeneous production units.

1.2 Research questions and Objectives

Given the background and overview of the research problem, we identified an imminent need to conduct an analysis on the existence of the missing middle in a developing country as well as to explore the fundamental driving forces by which it is generated. The central research question addressed in this thesis, therefore, was to examine the missing middle phenomenon in a developing country.

To answer the above research question, the thesis focused on (a) developing a theoretical framework that could provide a plausible explanation for the missing-middle phenomenon; (b) empirically examining the two factors affecting FSD indicated in the theoretical model(corruption and productivity) using the micro data of Vietnam; and (c) providing further insights into the calculation of productivity at the aggregate level to explore the linkage between this measure and the missing middle.

Tybout (2014b) claimed that the missing middle is at least partly policy-induced and stated that when policies are imperfectly enforced, producers tend to stay small in order to avoid different constraints, such as financial constraints, human resource limitations, corruption, or regulatory burdens. In developing countries, business constraints are different across firm size.The middle-sized group tend to face the largest burdens. Another factor affecting the FSD is the firm's productivity level. Although it is difficult to measure how well a firm is managed, production efficiency can be evidence of the quality an organization's management. As indicated in Tybout (2000), in developing countries, firms perform poorly in several respects: "(1) markets tolerate inefficient firms, so cross-firm productivity dispersion is high; (2) small groups of entrenched oligopolists exploit monopoly power in product markets; and (3) many small firms are unable or unwilling to grow, so important scale economies go unexploited.".

Intuitively, the missing middle exists because there is a burden or barrier against entering a market segment, particularly for a middle-sized firm. At the same time, an element favors larger size firms. Hence, the phenomenon might exist in a developing country because of the combination of these two factors. Further, these two factors could be more striking in a developing country than in a developed country because of environmental factors, production technology, or the efficiency of market structures.

To understand these aspects further, first,we devised a theoretical framework that could plausibly explain the formulation of the missing middle in the firm size distribution of a developing country. Second, we thoroughly investigated the missing middle phenomenon in Vietnam. Vietnam is an interesting case because it is a transitional, developing country. Over the last two decades, since the emergence of the Doi moi (Renovation) process, the country has moved from bureaucratic, centralized management based on state subsidies toward a multi-stakeholder, market-oriented economy and has experienced a period of continuous high economic growth. However, Vietnam has yet to become a tiger in Asia as expected. Investigating the firm size distribution of Vietnam could shed light on the underlying mechanism as well as on factors that result in the mis-allocation of resources and hinder firm growth. As in other developing countries, studies on FSD in Vietnam are scarce. The few studies conducted have been limited to empirical studies only and none provides a comprehensive picture of the FSD or presents the characteristics of the FSD of Vietnam. Thus, there was an opportunity to investigate the FSD of Vietnam, from both empirical and theoretical perspectives.

In general, the main objectives of the thesis were (a) to provide a tractable dynamic model of the missing-middle formulation within a canonical framework from the literature, focusing on productivity and corruption; (b) to draw a comprehensive picture of the FSD of Vietnam as well as to report empirical evidence of the missing-middle phenomenon in Vietnam; (c) to study possible effects of corruption and productivity on the FSD of Vietnam; and (d) to further investigate aggregate-level productivity measures and locate a possible linkage with the missing middle.

1.3 Data

In this section, we present an overview of the two datasets used in the thesis[1].

The first dataset is *the enterprise census*. In Vietnam, the census has been conducted annually since 2000 and covers all registered firms in Vietnam[2]. In this paper, we focused on the nine-year period from 2000 to 2008. We excluded recent years (2009 to present) because the Vietnamese economy has suffered from the effects of the global financial crisis and because a 30% reduction in the corporate income tax rate for qualifying entities was implemented in the

[1]More detailed descriptions can be found in the data section in each chapter.

[2]In the enterprise census, an enterprise is defined as "an economic unit that independently keeps a business account and acquires its own legal status."

fourth quarter of 2008 and all of 2009[3]. In addition, small-sized and middle-sized firms involved in labor-intensive production and processing activities may have benefited from the subsequent tax reduction under Decree 60/2011. In addition, we excluded inconsistent data from our sample, such as observations that were recorded twice for the same firm in the same year, those with negative or zero revenue values, or those with an implausibly large number of employees. We included 14 sectors of the manufacturing industry that had at least 200 observations for each year, and the number of observations ranged from approximately 3,350 to 29,700 per industry. The information from this dataset was used in investigating the firm size distribution of Vietnam as well as in productivity calculations.

The second dataset was *the small and medium manufacturing enterprise survey*. The survey was conducted by the Central Institute for Economic Management (CIEM) in collaboration with the Vietnamese Institute of Labor Science and Social Affairs (ILSSA), the Department of Economics (DoE) of the University of Copenhagen, and UNU-WIDER with the Royal Embassy of Denmark in Vietnam. The survey has been conducted every two years since 2005. Over 2,500 enterprises across ten provinces have been surveyed. Surviving firms are interviewed each survey year (tracer survey). In comparison to the first dataset, the survey provides comprehensive information on all forms of SMEs including private firms, collectives, partnerships, private limited enterprises and joint stock enterprises. Joint ventures were excluded from the sample because of the high degree of governmental and foreign involvement in ownership structures. The dataset excludes all state-owned firms as well as firms in other sectors such as the service sector. More importantly, this survey contains information on both formal and informal firms. Our initial sample contained 9567 observations, which included all firms in the SME survey covering four years: 2005, 2007, 2009, and 2011. We excluded all firms for which we did not have information on firm size, wages and total assets. We also discarded firms for which we did not have information on firm age. The final sample consisted of 8421 non-state manufacturing firms in ten provinces of Vietnam. Employees in our analysis included both regular full-time and casual employees. The information of this dataset was later used to investigate the effects of corruption on the formation of the FSD in Vietnam.

[3]See Circular No. 03/2009/TT-BTC published in January 2009 by Vietnam's Ministry of Finance.

1.4 The Significance and Contributions of the Thesis

The Significance of the Thesis

FSD analysis is a multifaceted topic that has attracted the interest of scholars since at least the beginning of the 20^{th} century. The FSD formulation represents the dynamics of the economy because simply knowing the size or growth of individual firms may not facilitate in understanding of the status and development of the dynamic process of a country as a whole. Previously, theorists have proposed that individual firm growth resulting from competitive pressures and the economic environment seems to have marginal effect on the FSD formation. Based on Gibrat's models, individual firms are normally assumed to grow at a random rate based on Gaussian models. Recent research, however, has shown that firm growth is a much more complex process and that FSD and growth distribution do not seem to be purely dynamic. Most firms remain in the same size group in the distribution for a long time while few others grow rapidly. Such size stagnancy should be investigated to provide a comprehensive understanding of firm dynamics and growth. Thus, in this thesis, we explored the opportunity to provide some insights into the issue.

In particular, the thesis focused on the FSD of developing countries. Despite the long history of FSD analysis in developed economies, few researchers have examined the FSDs and their distinct features in developing economies. These studies provide evidence of the missing middle phenomenon. The phenomenon indicates the absence of a robust and dynamic SME sector in many developing countries, particularly in the least developed countries. This structural imbalance has arisen despite various SME promotion programs in these countries. According to these studies, the cost of suboptimal size - that is, a firm that is too large or too small - can be significant[4]. Thus, the analysis of this phenomenon may illuminate ways to tackle sources of inefficiency and enhance economic development in developing countries.

No adequate explanation of the FSD in developing markets exists in the literature. Most researchers have considered small domestic market and poor infrastructure to be major problems that determine the concentration of small-sized firms and dis-economies of scale in developing countries. However, the underlying assumption in most studies is that firms respond identically to country-wide shocks. Little is known about the firm level performance as well as about the

[4]According to Riahi-Belkaoui, 1994, up to 25% of the cost that a large manufacturing firm has to pay are organizational slack due to communication problems, bureaucratic inefficiencies and other dis-economies of scale.

dynamics within one industry. Consequently, policy analysis in many important areas (for example, corporate governance) in developing economies rests on shaky foundations. We have neither a complete picture of the FSDs of developing countries nor a thorough understanding of the dynamics of firms behavior that can plausibly explain the FSD formulation. Therefore, a need exists for an analysis that addresses the specific problems encountered by small firms that leads to the lack of domestic enterprises in the middle range.

Finally, productivity and efficiency represent manifestations of the business environment in developing nations. Recent literature has documented empirically how distortions of business environments can affect aggregate productivity though mis-allocations of resources from more productive to less productive firms. Tybout (2000) stated, “given the fundamental importance of efficiency and productivity growth, improvements in the way that we measure these concepts should be a priority”. In developing countries, firms that are more efficient commonly have “too little” output or employment allocated to them because of various distortions in their economies; thus, the gap in aggregate productivity is a sign of such mis-allocation of resources. Analysis on productivity at the aggregate level, thus, can shed light on the mis-allocation of resources, which may possibly explain the missing middle. The thesis contributes to the literature in this respect.

Major Contributions of the Thesis

In the thesis, we filled in the gap in the literature by providing evidence of the existence of the missing middle and by offering a comprehensive analysis of factors that may constitute a plausible explanation of the phenomenon. The major contributions of the thesis involve empirical findings, theoretical modeling and statistical estimation theory.

The first contribution is in the theoretical perspective. In the thesis, we have developed a theoretical framework whereby corruption acts as the barrier that prevent small firms from growing to a bigger size group. In particular, the model describes the process of market selection, technology accumulation and fixed cost as the main determinations of the FSD formulation. The model has extended the powerful insights of continuous-time mathematics into environments where firms have heterogeneous technology and dynamically decide to enter or exit the market while facing various financial constraints. Thus, the innovation of this study was to apply continuous-time mathematics to describe the dynamic decision making of firms and FSD within a model that included the constraints that firms face. It is clear from the current literature that there is an urgent need for better ways to model a firm’s dynamic decision making. The thesis is a contribution towards this undertaking.

The second contribution involves the empirical perspective. Using the micro data of manufacturing firms in Vietnam, we draw a comprehensive picture of the FSD of a developing country. We provide evidence against uni-modality of the FSD and show the evidence of the missing-middle phenomenon in Vietnam. We also examined the underlined mechanism behind the missing middle by looking at the relationship between the phenomenon and factors indicated in the theoretical framework, such as corruption, efficiency, and returns to scale. First, we used a unique firm-level dataset, which included both informal and formal companies, to examine what characteristics of firms are the important determinants of bribery in a developing nation. We found that as firm size increases, the likelihood of paying bribes also increases. In addition, the formality of firms also matters in the sense that the likelihood of paying bribes is higher for formal firms, which tend to be larger. In addition, we compared our results to the extant literature by examining the missing middle at the sectoral level of a developing country. Our analysis shows that middle-sized firms production efficiencies tend to be lower than the efficiencies of small-sized or large-sized firms in most manufacturing industries. Further, the least efficient firm size differed across sectors, which indicates the necessity of sector-by-sector analyses to study the underlying mechanisms of the missing-middle phenomenon. To the best of our knowledge, we do not know of any other studies that link the missing middle and production efficiencies across different firm sizes in various sectors of a developing country. Finally, we investigated the inefficiencies associated with scale across different sectors and among different size groups in each sector. We found that large-sized firms in most sectors tended to suffer from inefficiencies associated with scale; however,the possibility of increasing returns to scale may exist in part of the production functions in which many small firms operate. Our analysis indicates that one of the problems, in Vietnam, at least, is how to activate some resources that these large-sized firms hold but tend to only partially utilize in production.

The third contribution of the thesis involves the statistical perspective. The thesis used the polynomial-time algorithm developed bySoleimani-Damaneh and Reshadi (2007) to identify inefficiency associated with scale under non-convex technology. Relaxing the convexity assumption of the production set becomes critical in an investigation of local non-convexities as a technological explanation for the missing middle. In addition, we looked at industry efficiency, which is compatible with the entry and exit of firms. We devised an approximation theorem that includes the different specifications introduced in the literature to the same basic optimization program to calculate for efficiency score. This program is linear and can be solved using standard simplex methods. This technique introduced advantages in terms of both computation and interpretation. The computational advantage is obvious. The interpretational advantage comes from the fact that we needed to discuss a simple basic model to which we could refer for any type of aggregate analysis. As all industry inefficiencies are approximately equal regardless of the underlying technology, what seems to matter at the industry level is the efficiency decomposition. In the thesis, we also discuss efficiency decomposition whereby industry efficiency can be decomposed into the sum of aggregate technical efficiency and re-allocative efficiency.

1.5 Limitations of the Thesis

The research was confined to the analysis and investigation of the missing-middle phenomenon. The literature has indicated a variety of factors that might affect the FSD in general and the missing middle, in particular. Many factors, including financial constraints, regulatory framework, and property rights, can affect the formation of FSDs in developing countries. However, to keep the thesis within manageable proportions for rigorous investigation, only two factors are discussed in this thesis:corruption and productivity. In addition, because of time and resource constraints, the theoretical model developed in the thesis is limited with restricted assumptions.

The limitations of the thesis will be discussed in further detail in the conclusion chapter.

1.6 Thesis Outline

The thesis follows the doctoral thesis structure suggested by the University of Queensland. The thesis begins with a brief outline of the firm size distribution formulation and the missing-middle phenomenon. In this section, the overall outline and organization pattern of the thesis are discussed. The thesis contains six chapters. The chapters are summarized in the following paragraphs.

Chapter 1: Introduction. In the first chapter,we explore the research background, research questions, identify the research objectives,describe the data, and summarize the significance and contributions of the thesis. The limitations and the general outline of the thesis are also presented.

Introduction. The chapter explores the research background, research questions, identifies the research objectives and describes data, summarizes the significance and contributions of the thesis. The limitations and the general outline of the thesis are also presented here.

Chapter 2: Literature Review. The chapter is divided into two parts. The first part is devoted to identifying major concepts in the thesis. The second part focuses on the related literature. The chapter shows that FSD is a complex issue that requires an integrated analysis of various factors.

Chapter 3: Theoretical model. In this chapter we develop a theoretical framework to capture the effect of corruption on the development of firms. The model captures the increasing costs of large-sized firm groups, compared to the costs of the smaller ones, which leads to the concentration of a continuum of infinitesimal firms in the market. In particular, the chapter extends the model presented in Luttmer (2007) to a framework that could be applied to a developing economy. We used the model to describe firm dynamics - entry, exit, and growth, which, on the one hand, being consistent with empirical evidence in developing countries, while, on the other hand, allowing for heterogeneity in the characteristics of firms.

Chapter 4: Empirical Analysis - Corruption and Productivity. In this chapter, we investigated empirically the relationship between firm size and production efficiency and the relationship between firm size and the likelihood of paying bribes using firm-level data from Vietnam. Our analysis indicates that middle-sized firms' production efficiencies tend to be lower than the efficiencies of small-sized or large-sized firms in most of the manufacturing industries. In addition, as firm size increases, the likelihood of paying bribes also increases.

Chapter 5: Productivity at the industry level. This chapter provides analyses of the total inefficiency at the industry level. We provide an approximation theorem that basically report all the different specification of efficiency scores in the aggregate level introduced in the literature to the same basic optimization program. This program is linear and can be solved using standard simplex methods. This technique introduces advantages in terms of both computation and in terms of interpretation.

Chapter 6: Conclusion. This chapter provides the summary and implication of the thesis. The objective of the chapter is to consolidate the answers to the research questions. In the chapter, we synthesize all major findings from Chapters 2 through 5 and present the research implication to researchers and policy makers. Further details regarding major contributions and limitations of the thesis are also discussed. Finally, future research directions are suggested.

CHAPTER 2

LITERATURE REVIEW

2.1 Theoretical Concepts

2.1.1 The definition of a firm

The diversification of theories that model FSD largely results from the complexity and multiplicity of the definition of a firm. Such complexity gives rise to different approaches to modeling FSD. Therefore, it is worth presenting a brief summary of the definition of a firm and its "efficient boundary" in the literature.

How may a firm be defined? For centuries, scholars have asked and answered the question. First, in 1776, Adam Smith, in *Wealth of Nations* defined a firm in terms of the division of labor. Later in the 1890s, Marshall defined a firm as a profit-maximization production function. This approach remains the neo-classical view of a firm to this day.

Such a view, however, cannot explain why firm sizes vary even when operating in the same industry where they face similar optimization problems. It also does not explain why certain economic activities are carried out by the market while others are organized under the hierarchic structure within a firm. A firm is still a "black box" transforming inputs into outputs. Alternatively, a firm is just a profit-maximizing (or cost-minimizing) mechanism, operating in the exogenous business environment that lies beyond its control.

Recent approaches to defining a firm are presented in the following section.

The first approach to defining a firm is associated with the *transaction cost theory*, in which a firm is considered an internal process that replaces the external price mechanism. In 1937, Coase, in his article "The Nature of the Firm", suggested that the existence of a firm is fundamentally a matter of contracting costs. According to Coase, the main reason for firm establishment is to escape from the costs associated with using the market mechanism such as costs in identifying what parties to deal with, establishing terms and conditions, conducting negotiations, and concluding, monitoring and enforcing contracts. These costs can be avoided by binding employees under employment contracts and having them follow orders from managers. The upper limit of firm size occurs where internal costs within a firm, which rise due to diminishing returns to management equal costs of making the transaction in the open market.

Later, Williamson (1975) suggested that a firm governance structure is influenced by the degree of uncertainty, the frequency of transactions and the degree to which durable, transaction-specific investments are required to realize the least-cost supply. With transactions that occur frequently, long-term arrangements should be established, especially when one party needs a particular resource and has to make especially large investments in the transaction. Because they are more vulnerable to opportunistic manipulation by the second party, the investment party will try to arrange coverage for such an event. Firm formation, in this case, can solve a problem, which is well-documented in the literature as the "hold-up" problem.

Although the transaction-cost approach has not been formalized into theoretical models, it offers an "empirical success story", which has been demonstrated by many studies (see Masten, 2000, Boerner and Macher, 2002, Saussier, 2000). Transaction cost theory, however, has been criticized for failing to consider the demand side. According to the theory, a firm is considered simply "a nexus of contract" and the trade off between external and internal coordination costs determines the optimal boundaries of the firm.

The second approach to defining a firm follows the *property rights theory*. According to the theory, a firm is defined in terms of the ownership of property rights associated with it (see Grossman and Hart, 1986, Hart and Moore, 1990, Holmstrom and Milgrom, 1991). Proponents of this approach are aligned with transaction cost theory but further develop the definition by taking into consideration both costs and benefits of integration. Thus, a firm is defined as a set of assets under common ownership and its boundaries are determined by the optimal allocation of residual rights of physical assets (Hart and Moore, 1990).

The third approach to defining a firm is *a knowledge-based* view of the firm in which firms are considered groups of capacity (see Penrose, 1959, Richardson, 1972). In the view, the knowledge, learning, and capacity development aspects of a firm are major elements. A firm is considered a collection of productive resources including tangible assets and human capital (not just the resources themselves, but also the services they render). The administrative and managerial authority over the use of such resources defines the efficient boundaries of a firm (Penrose, 1959). Richardson (1972) further expanded the discussion by defining a firm as the coordination of capacities among industrial systems. In Richardson's setting, a firm is a network and a firm's boundaries are influenced by varying degrees of specific production activities which are underpinned by firm-specific capacities.

The fourth approach to defining a firm follows the *evolutionary theory*. A firm, according to the theory, is defined based on bounded rationality and routines. In other words, a firm is an entity that processes, stores, and produces knowledge in a set of routines. According to Nelson and Winter (1982), "a firm can be understood in terms of a hierarchy of practiced organizational routines, which defines lower order bureaucratization skills and higher order decision procedures for choosing what is to be done at the lower level".

Evolutionary theory, therefore, not only provides a definition of a firm but also sheds light on the explanation of firms' heterogeneity (firms are heterogeneous when they have different sets of routines that are either tacit or that incur high costs to transfer). The theory also stimulates the investigation of firms' dynamics through examining processes of searching and selection and the transformation of routines into activities. The theory, however, does not provide clear insight to deal with conflicts within firms, for example, capital - labor conflicts, corporate governance, and the issue of entrepreneurship (Witt, 1998).

Finally, there are other less influential approaches to defining a firm, such as *incentive-system theory* (Holmstrom and Milgrom, 1991, Holmstrom, 1999) and *adaptation theory*[1] (Klein, 2000a; Klein and Murphy, 1988,).

From the preceding discussion, it is clear, the definition of a firm has evolved over time. Instead of being considered merely a black box, a firm can be understood in a much wider perspective, encompassing a more ecological view in which firms interact with other agents and have their own internal functions.

Nevertheless, only partial answers concerning the nature of the firm, its boundaries, and its

[1]A further detailed discussion on these two approaches can be found in Gibbons (2005).

internal organization are available. Why a firm chooses a certain type of organization or makes a certain decision regarding mergers and acquisitions (M&A) or outsourcing is, largely unknown.

In this thesis, we define a firm as a collection of assets including both physical and human capital. This definition is an extension of the firm definition in property rights theory: our definition includes labor as a part of the firm operation process.

2.1.2 The Measure for Firm Size

Firms are heterogeneous, differing along in various dimensions including firm size. The choice of measure used to determined firm size can have a significant impact on the analysis of firm size distribution. In this section, we identify and discuss one plausible measure of firm size.

In the literature, many possible ways to measure firm size have been presented. The size of a firm can be proxied by financial or stock market value , sales or revenue (Basu,1997), total assets (...), value added production (Basu,), total number of employees or even the economic capital including debt plus equity.

From an economic point of view, in general, one would expect those measures to be proportional to each other in the same business model. Kirchhoff and Norton (1995) used Standard & Poor's COMPUSTAT data to compare three measures of firm size (number of employees, assets and total sales) and obtained similar results over a seven-year period, showing that if the distribution holds for one of the variables, it also holds for the others.

There are, however, obvious differences in the ability of variables to capture the internal processes of a firm. For instance, sales can be influenced by marketing or financial decisions and are vulnerable to inflation and exchange rate fluctuations. The total asset variable does not capture all aspects of size dynamics because many firms have substantial intangible assets that are difficult to measure and rarely appear on balance sheets. The total assets variable also does not take into account a firm's liabilities. The revenue measure, in contrast, can capture fluctuations in price and quantities sold in the market.

Among firm size measures, the number of a firm's employees is the most widely used indicator of organizational complexity and the most suitable measure for distribution analysis (Penrose, 1959). The number of employees measure reflects internal processes involving all of a firm's adaptations to changes, such as organizational restructuring and seasonal operational activities.

Measuring size by the number of employees is also preferable because it is not sensitive to inflation and exchange rates and can be considered a direct indicator of domestic firm size. Further, this measure can directly link to macro-economic issues such as job creation, employment and unemployment.

However, we acknowledge certain drawbacks associated with using the number of employees as a measure of firm size. The measure is not able to capture improvement in quality, such as the increase of productivity, the degree of integration and other make-or-buy decisions. In addition, according to our measure, the number of employees is limited to only domestic laborers. In other words, in multi-national enterprises, workers working in branches abroad are not counted according to our measure.

Despite these disadvantages, however, the number of employees is, however, still the most accessible and appropriate measure for analyzing the distribution of firm size

2.2 Firm Size Distribution Literature

2.2.1 Specific Features of Firm Size Distribution

Empirical evidence from different databases has confirmed that the firm density distribution is leptokurtic, stable, and highly skewed to the right. This old and intriguing observation is clearly a robust pattern confirmed by many scholars (Coad, 2009, Cabral and Mata, 2003).

One of the most influential findings in the FSD literature is Gibrat's law or the law of Proportionate Effect[2]. According to the law, the probability of a given proportional change in size during a specified period is the same for all firms in a given industry - regardless of their size at the beginning of the period (Mansfield, 1962). Firm size, therefore, is dependent on random shocks that are independent of each other and of the initial size of the firm.

Gibrat's law predicts that firm growth is a purely random process, independent of firm size. In other words, "the probability of a given proportionate change in size during a specified period is the same for all firms in a given industry, regardless of their size at the beginning of the period" (Mansfield, 1962). As a result, the market concentrates as large firms increase their power in

[2]The details of Gibrat's law are presented in the Appendix.

the market and as the FSD converges to a log-normal distribution over time.

Much research effort has focused on determining the validity of Gibrat's law. Empirical studies of Gibrat's law are difficult to compare because there are significant differences in samples used and methodologies applied[3]. In a survey on Gibrat's law, Santarelli et al. (2006)[4] summarized the difficulty of drawing a clear conclusion about whether Gibrat's law is valid or can be rejected. In general, recent evidence has shown that Gibrat's law tends to hold when using data with a large enough number of firms (see Geroski and Gugler, 2004, Hart and Oulton, 1996), and the law tends to be rejected when all firms are included because of the more rapid growth rate of small, young firms[5].

Regarding the shape of the FSD, the literature has provided evidence supporting both the log-normal and Pareto distributions.

First, many studies have indicated the shape of FSD can be approximated by the log-normal distribution[6]. Quandt (1966) provided evidence of the superiority of the log-normal distribution over all three versions of the Pareto distribution. More recently, Ishikawa (2007) showed that the log-normal distribution can be deduced from the non-Gibrat's law. Other studies, however, have provided evidence of the deviation of FSD from the log-normal distribution, especially in the right tail and in disaggregated datasets. For example, Cefis et al. (2008) showed that the log-normal distribution fits FSD well, except for the upper tail, when using the dataset of the Netherlands. In the same vein, Marsili (2005) reported that deviations from the log-normal distribution in different sectors were witnessed even when the law held at the aggregate level. One possible explanation, according to Dosi (2007), is that such a FSD shape may be the result of statistical aggregation effects alone, rather than the result of any mechanism with deeper economic meanings.

[3]The methodology used to test Gibrat's law also varies over time from the simple Ordinary Least Square (OLS) technique (Mansfield, 1962) to Augmented Dickey-Fuller test (Geroski et al., 2003), Unit Root test (Oliveire and Fortunato, 2006), General Method of Moments (Oliveire and Fortunato, 2004) or Heckman's equation (Calvo, 2006).

[4]The survey included 60 empirical studies. According to the results, the law can be tested using either: (a) all firms in a given industry, including those that have already exited during the studied period; (b) firms that survived over the entire period; (c) firms that are large enough to overcome the minimum efficient scale of a given industry. Moreover, these three populations can all be tested using either static analysis or dynamic analysis.

[5]Many studies have shown significant negative correlation between firm growth and size (Evan, 1987, Mansfield, 1962) and between firm growth and age (Evan, 1987). Average cost increases for sizes below some minimum levels but is quite stable for sizes above it (Mansfield, 1962).

[6]The log normal distribution is concave to the origin on the log-log scale . This implies the existence of more middle-sized firms and fewer large firms in the FSD than would be predicted by the Pareto distribution.

Another set of studies has focused on the empirical test of whether FSD is Pareto distributed[7] Axtell (2001) found that the size distribution of US firms is well approximated by the Pareto distribution. Axtell's conclusion was supported by many other studies (see Fujiwara, 2004, Hart and Prais, 1956). More recently, Cirillo and Husler (2009) found the size distribution followed the Pareto law in 40% of the largest Italian firms, and Marsili (2005) found the Pareto distribution in the upper tail of the FSD in the Netherlands.

Another feature of the FSD in many developed countries is Zipf's Law[8]. Zipf's law is evidenced in several economies including the United State (Axtell, 2001), France (Giovanni and Levchenko, 2008), European countries (Fujiwara, 2004) and Japan (Okuyama, 1999). In general, Zipf's law tends to fit well with size distributions that exclude small firms from the dataset[9].

2.2.2 Influential factors

Given particular features of the FSD, various factors may be attributed to differences in firm size evolution and thus play a role in FSD formulation.

The literature includes several important factors that affect firm size and growth. For example, Storey (1994) specified three groups of firm growth determinants: (a) entrepreneurial resource factors (such as education, management experience, training, age, etc.); (b) firm-related factors (such as product, sector, legal form, location, size, ownership, etc.); and (c) strategy-related factors (such as workforce training, technological sophistication, market adjustment, competition, exporting, etc.). (see Storey, 1994, for a detailed discussion on firm growth determinants)

[7] In 1896, Pareto was the first to find this systematic distribution of personal incomes. Later, a series of studies showed Pareto Law is consistent over time, across countries and across size distributions as diverse as individuals, cities and firms. Compared to the log-normal distribution, the Pareto distribution puts more weight on the lower tail (there are more small firms than predicted by the log-normal distribution) and the upper tail decays much less rapidly (Coad, 2009). If FSD follows the Pareto law, its coefficient (the slope of the Pareto curve) can be considered to represent the degree of concentration of larger firms and the structure depends on the interaction among firms rather than external effects or individual firm behaviors (Krugman, 1996).

[8] Zipf's Law appears when the Pareto distribution exponent equals unity. Zipf's law implies that the frequency of the variable is inversely proportional to its rank. That is, the n^{th} most common event appears roughly $1/n$ times as often as the most common one. Zipf's law gives rise to a flatter tail distribution, implying a few large firms account for a major share of overall economic activity. Further discussion on power law, Pareto distribution and Zipf's law can be found in the Appendix.

[9] Zipf's law can be tested by obtaining the tail index and comparing it with the benchmark value of unity using various econometric methodologies. These include OLS on the log-log scale (Axtell, 2001), the $Rank - 1/2$ estimator (Gabaix and Ibragimov, 2011), and the Maximum Likelihood technique (Hill, 1975).

Among these factors, firm-related factors, such as the decomposition of output or the sector within which a firm operated were important in predicting the size of a firm. Many researchers have also stressed the importance of a negative correlation between growth and size. (see Sleuwaegen and Goedhuys, 2002, Liedholm and Mead, 1987). According to Liedholm and Mead (1987), small, young firms are more likely to grow. In addition, estimation results in McPherson (1996) indicated that age is the variable with the most predictive power for employment growth.

In this review, however, we focus on other factors that have recently added to the literature and strongly influence modeling FSD in the economy.

2.2.2.1 technology

We focus on two proxies for the technology factor: productivity and economies of scale.

With regard to productivity, a natural assumption is that firms with the highest productivity will grow while firms with the least productivity will decrease in size and eventually exit. The relationship between firm size and productivity in the literature, however, encompasses many facets. On the one hand, many researchers have found a positive correlation between increase in productivity and the concentration of employment in large firms. For example, Bartelsman and Doms (2000) studied the distribution and growth of productivity of developed countries. Bartelsman and Doms reported that productivity levels were widely dispersed and highly persistent across sizes and noted the large productivity gap between the largest and smallest firms. Leung et al. (2008) investigated the US - Canada difference in the distribution of employment over firm size and confirmed that a larger average size supports higher productivity at both the plant and firm level. Diaz and Sanchez (2008) used a stochastic frontier model to investigate small-sized and medium-sized manufacturing enterprises in Spain and found that these firms tended to be less efficient than their large-sized peers.

On the other hand, other studies have provided evidence that small firms are sources of dynamism and productivity growth in many OECD countries.

The second proxy for technology involves economies of scale and dis-economies of scale. The effects of self-reinforcing mechanisms on industrial structures introduced by economies of scale is an important cause of nonlinear interactions in economic systems (Arthur, 1988). When there are economies of scale, the firm can choose different optimal investment strategies (Wagener, 2003) which result in different industrial structure (Gustavasson, 2002). In fact, when FSD is described by the number of firms economies of scale can be a crucial factor in addition to other factors such as entry and exit that determines the FSD (see Baumol et al., 1983, 1988). Crosato et al. (2009) investigated the data from a micro-survey in Italy, using nonparametric estimates, and found that firms in the concave part of the Zipf plot of Italian firm distribution overwhelmingly experienced increasing returns to scale, while firms in the linear part were mainly characterized by constant returns to scale.

2.2.2.2 Mergers and Acquisition

External growth through integration and disintegration is another important channel of firm development. In the literature, many studies have been dedicated to analyzing the contribution of mergers and acquisitions (hereafter M&A) to FSD development. M&A activities, in these studies, tend to shift the FSD toward larger sizes, leading to a departure from the log-normal distribution.

Ijiri and Simon (1974) showed that M&As can contribute to a deviation of FSD from the Pareto distribution because external growth is more constrained by firm size (and less random) compared to internal growth. Ijiti and Simon provided empirical evidence that the growth rate of firms resulting from M&A activities does not follow Gibrat's law. Similarly, Singh (1975) concluded that bigger firms usually acquire small firms while large-scale actors tend to merge among themselves, creating a size-strong segment in the upper tail of FSD. In addition, Hannan and Freeman (1977) found the effect of mergers was significant and suggested that without the M&A effect, small firms would have grown much faster than large ones. More recently, using data from manufacturing firms in the Netherlands, Cefis et al. (2008) documented such a shift in FSD. They showed that the effects of M&A were not uniform; in fact, middle-sized firms increased in number, outweighing the increase in the upper tail group, and therefore, the market concentration of the economy decreased. In other words, as long as M&A are more effective than internal growth in achieving size than internally growth and are sustainable, we can expect higher growth rates of larger firms.

M&A, in sum, are the main contributors to the variance in firms' growth because they are the mechanisms behind most rapidly growing cases; thus, the effect of M&A challenges the validity of Gibrat's law and the Pareto shape of the FSD.

2.2.2.3 Policy-related factors

Government intervention can have profound impacts on firms' activities. For example, policies that distort trade patterns can result in a "premature" shift of resources to products that require more capital-intensive production arrangements. Further, policies that discriminate against small- and medium-sized firms in terms of credit allocation can have negative impacts on the development of these firms. Guner et al. (2008) examined the costs associated with size-dependent policies that distorted production scale. They found that the effects of this class of policies are large as such a policy lead to sizable reductions in output per firm and a significant increase in the number of establishments. On the other hand, such an intervention can result in an increase or decrease in productivity depending on whether the firms are efficient. In the bad case, firms who receive external supports may artificially increase their efficiency, relative to other firms and thus, may replace efficient firms that do not receive any external supports. As a result, such a policy can cause inefficient firms to supplant efficient firms in the market and this may result in the increase in market concentration.

2.2.2.4 Stochastic factor

Another finding in the empirical literature related to firm size and firm growth is the high degree of variance left unexplained by firm growth and size regression analyses. According to Coad (2009), the levels of R-squared from most regressions on firm growth are surprisingly low (often approximately 5%). These findings indicate that randomness[10] makes a significant contribution to firm growth and size determination.".

In the literature, there are many studies that provided evidence supporting of predictions made by Gibrats law of proportional effect in which all firms, irrespective of size, grow by some random draw from the distribution of growth rates.

Later, researchers extended the idea by stating that firm size follows a random walk. Geroski (1999) noted: (a) there is significant contribution of randomness to firm size growth; (b) the

[10]Geroski (1999) defines randomness as "an event with an unknown form or a known fact whose date of occurrence is undetermined

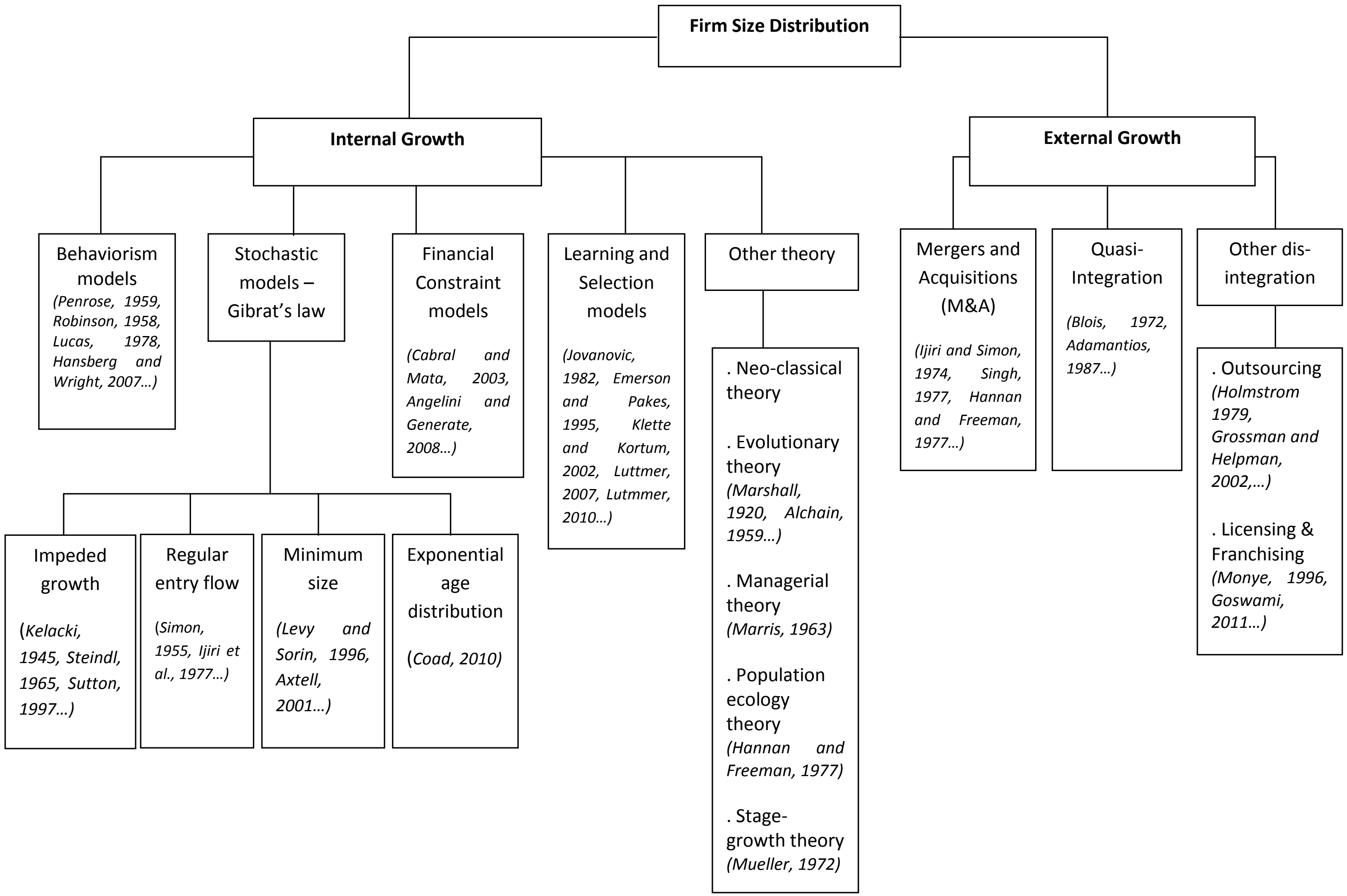
effects of such a shock can be permanent to firms with growth patterns showing some noise rather than following a deterministic trend; and (c) firm growth is idiosyncratic in that, the growth rates of any two randomly chosen firms are uncorrelated. This latter fact is somewhat interesting. We normally expect to see a proportional change in firm growth rates and any variations of the economic environment.

Recently, with the development of economic theories, the principle of randomness seems to be more evident and serves as a better explanation of FSD in modern economies where turbulent competition and rapid technological development are dominant characteristics. As a result, many stochastic studies have recently published in the FSD literature.

2.2.3 Different Theoretical Approaches to modeling FSD

As mentioned in the preceding section, FSD formulation can be attributed to various factors. As a result, many theories have been developed to explain the shape of FSD. This section is an overview of related theories. Table 1 provides a summary of all related theories.

Table 1: Summary of Theoretical Model of the FSD of Developed Countries



2.2.3.1 Stochastic Models

The stochastic model approach was developed simultaneously with other theories relating to FSD such as *neo-classical* and *behaviorism theory*. The major objectives of the stochastic models are (a) to prove the existence and persistence of stochastic factors, and (b) to investigate the concentration level of firms that gives rise to a certain FSD.

Gibrat was the first and probably the most influential scholar in this school of thought. Gibrat (1931) documented the Law of Proportionate Effect, also known as the Gibrat's law in which stochastic factors are used to explain the log-normal distribution of the FSD, showing increasing mean and variance among a fixed number of firms. Gibrat's law has become the most influential hypothesis and is used extensively in later stochastic models of FSD. However, in Gibrat's model, there is no steady state because the probability of the existence of a firm size in a certain interval approaches zero in any given size interval, a generation of stochastic models has been developed, incorporating different assumptions to impose the steady state into Gibrat's model[11]. These models generate either the log-normal or Pareto distribution or similar types of distribution (e.g., Yule distribution, extended Katz distribution, Waring distribution). More recently, Coad (2010) proposed an alternative approach to prove that FSD is stationary by integrating the log-normal distribution of a Gibrat-type growth process over the exponential distribution of firms' age. Such an approach also results in the stationary Pareto FSD.

The stochastic approach, however, has been criticized in the literature because of its assumption that firm growth follows a random walk and that exogenous shocks to the growth rate of a firm and thus random elements determine the fate of the firm.

2.2.3.2 Learning and Selection Models

In contrast to stochastic models, researchers developed learning and selection models by using economic shocks to explain productivity growth, capital accumulation, and the skew of FSD. This approach emphasizes the learning capacity of firms and their ability to conduct innovations. Selection, then, occurs when firms realize their actual productivity on entering the market. The

[11] Gerrit (2005) summarized these studies and found three additional assumptions in these models: (a) the negative correlation of the average growth rate of a firm and its size (impeded growth) (Kalecki, 1945, Steindl, 1965, Suton, 1977); (b) the inclusion a constant stream of small firms entering into the market at the minimum firm size (Simon, 1955, 1960). The relatively regular flow of newly created firms; (c) the existence of a minimum firm size boundary below which firms cannot decline (Axtell, 2001).

main characteristic of these models is that they simultaneously take into consideration firm dynamics and levels of efficiency, through which the chances of firms' survival are determined.

Jovanovic (1982) assumed that firms enter a market without knowing their true productivity. Later, firms get to know their capacities through the production process and over time, the less efficient firms realize their inefficiency, produce less, and eventually exit. Small young firms tend to have less experience and respond more sensitively to market signals. Therefore, smaller, younger firms have higher probabilities of failure and more variable growth rates. The effect of selection on the evolution of the industry, however, eventually vanishes because technology shocks do not have permanent impacts.

Klette and Kortum (2002), following the same trend, developed a general equilibrium model of heterogeneous firms in which the stochastic process of innovation was the key element for firm heterogeneity and simultaneous entry and exit. In Klette and Kortum's model, the "Schumpeterian force of creative destruction" shows how firms' innovations occur at their competitors' expense. This model predicts the exit of firms and shows FSD is driven by the expansion of commodity varieties.

Under the same flow of destructive creation, Luttmer (2010) developed a model to capture the thick-tailed FSD with high entry and exit rates and relatively young ages of large firms. In Luttmer's model, firm growth stems from the creation of new products of both incumbents and new entrants. Relatively young large firms emerge from firms that have better blueprints and implement these blueprints more quickly. Their growth rates, however, are predicted to eventually slow to a low development phase at some random time. This results in the Pareto-type tail of FSD.

2.2.3.3 Behaviorism Models

Studies on behaviorism models have emphasized the knowledge assets and processes of coordination within a firm. For example, Penrose (1959), one of the leading protagonists of the school, stated that firm growth is led by the internal learning-by-doing process of managers and thus is constrained by the limitation of available managerial attention. In such a setting, fast-growing firms accumulate much higher operating costs, compared to those of lower-growing competitors. This situation is commonly known in the literature as the "Penrose effect"[12].Robinson (1958), Garicano and Hansberg (2004), on the other hand, relied on the rigidity of the managerial talent factor to explain FSD in the economy. They argued that the shape of FSD depends more on the distribution of managerial talents. Hansberg and Wright (2007) also related the industrial FSD to the intensity of industry-specific human capital. The limitation of human capital resource leads to more diminishing returns to human capital, which increases the scale dependence of firm size. As a result, industries such as the manufacturing sector, with a small share of human capital, tend to have much thinner tails. This approach, however, has been criticized for a lacks of empirical support for the existence of a link between the knowledge structure and firm growth (Carrizosa and Agusti, 2006).

In addition, many related theories have provided a part of the explanation behind the emergence of a certain firm size. These include other firm-growth theories, including classical economic theory and theories on external growth of firms (integration versus outsourcing, franchising, and licensing).

2.3 The Missing Middle and the Firm Size Distribution in Developing Economies

As mentioned, the major part of the literature on FSD has been based on studies conducted in developed economies. It is a well-known fact that profound differences in the business environment exit between developed and developing economies.

Given such differences, a sharp empirical contrast also exists between FSDs in developed and developing economies. While the size distribution in developed countries can be approximated by either the log-normal distribution or the Pareto law or a combination of the two, the developing economies' FSD is characterized by the missing middle where middle-sized firms account for the smallest proportion of the total number of firms.

Liedholm and Mead (1987) documented the missing middle first by presenting evidence to support the claim that most employment in developing countries is located in either small or

[12]Another hypothesis proposed by Penrose is that a firm is considered to be the accumulation of resources (i.e. brand names, internal knowledge of technology, trade contracts, machinery, etc.) and is bounded by the administrative capacity. All administrative skills and other inimitable knowledge such as working routines define the competencies as well as the size of a firm. The uniqueness of such competency results in the heterogeneity of firms in the market.

large-sized firms.

Later, Tybout (2000) reexamined the phenomenon. He found that in the context of the poor performance of manufacturing industries, FSDs of developing countries tend to exhibit dualism in which many micro firms coexist with a handful of modern large-scale firms and a missing proportion of moderate-sized firms. Tybout argued that strong business regulation can be the underlying reason behind the disproportionate presence of small entrepreneurs.

Similar evidence of the phenomenon was documented in Little et al. (1987) for South Korea and Steel and Webster (1992) for Ghana in the 1980s. More recently, Alfaro et al. (2008) reexamined the Melitz (2003) model using data from 79 economies. In the study, Alfaro et al. generated the bimodal distribution and specified the level of distortion that formulated such a distribution.

Recently, Hsieh and Olken (2014) studied data from India, Indonesia and Mexico and presented the following results: (a) many small firms exists but not so many middle-sized or large-sized firms, and in this sense, there is no missing middle, (b) large firms rather than small firms seemingly incur the large fixed costs, and (c) there is no evidence of discontinuities in regulatory obstacles across different firm sizes.

In response to the findings by Hsieh and Olken (2014), Tybout (2014a,b) suggested that a better test of the missing middle is to ask whether the share of middle-sized firms, as opposed to small- or large-sized firms, is smaller than the share that one would observe in an undistorted economy. One sign would be a concentration of small firms.

In general, studies on FSDs of developing countries have provided evidence that firm size distributions are more skewed to the right compared to the FSDs of developed economies. The high level of concentration of micro firms in developing countries, in conjunction with the existence of some large-sized firms indicates the presence of an underlying mechanism.

The missing middle phenomenon in developing countries, is emerging as exciting research focus for industrial organization scholars. The interest in missing middle research shows that forces, other than those traditionally mentioned in the economic literature, might be at work. Researchers have proposed various explanations, mainly examining the differences in the business environment between developed and developing countries. In this section, we focus on some potential factors that have been explored in the literature. The following section summarizes some influential factors found among the scarce literature that could explain the dual FSDs in developing economies.

2.3.1 Financial constraint

Many surveys carried out in developing countries have shown that access to finance is considered the most serious problem for SMEs. A low level of financial development skews the FSD by increasing the relative share of small firms. Cabral and Mata (2003) found that firms with limited financial capacity are forced to choose sub-optimal sizes, the optimal size can only be reached if financial constraints are removed. In other words, the evolution of the size distribution is determined by eliminating firms' financial constraints.

However, upon closer inspection, access to credit is not usually a major constraint to micro firms in developing countries because most SME credit facilitation programs are targeted at the micro-finance level. Business people can also easily borrow small amounts from friends and relatives. The medium-size firms, seeking to expand their operations, on the other hand, face a severe lack of access to medium- and long-term credit because commercial banks, hampered by limited resources, may choose to lend to large firms only, to reduce risk and save costs per unit loaned.

As a result, many authors have suggested that the lack of financing is one of the most important factor responsible for the missing middle (see Chowdhury et al., 2013, Gibbons, 2011, Schiffer and Weder, 2001). Marchiavello (2010) developed an incomplete contract model to explaining the missing middle. In his model, firms produce goods that can be used by customers or sold in the spot market. Both firms and customers, then, must borrow from external investors. Under circumstances of poor contract enforcement, firms may choose to integrate to maximize the returns in vertically related projects. Here, financial constraints act as a barrier to the creation and expansion of firms, resulting in an increase of vertical integrations and a surge in the number of large firms, hence, the emergence of the "missing middle" in the economy. More recently, Khan (2014) concluded that the major cause for the SME missing middle in Bangladesh was a continuing lack of access to financing.

2.3.2 Regulatory burdens

An emerging body of empirical evidence shows that various tax or regulatory policies can have significant impacts on firm size. In developing countries, the burdens of business regulations are dense and unpredictable (Brunetti et al., 1997). It is also not unusual for governments in developing countries to provide size-dependent policies, for example giving special incentives or allowing special tax breaks to large-sized projects or firms. In addition, labor market regulations often vary by firm size.

Given these facts, many authors have suggested that the missing middle emerges, in part, because taxes and regulations are enforced only among formal-sector firms. Dhammika et al. (2010) provided a framework in which the government imposes differentiated tax levels on firms when the government faces fixed administrative tax collection costs. As a result, the government finds it optimal to exempt small firms from taxation even though some firms refrain from growing beyond the cut-off level to avoid paying taxes. Giovanni and Goyette (2009) developed a model in which tax distortion, particularly, uneven tax auditing, in association with financial constraints, explain the distortion of the FSD. Cai and Manish (2013) also extended the Lucas (1978) mode to investigate the effects of size-dependent labor regulations in India. Cai and Manish showed that, the missing middle in Indian manufacturing industries were resulted from size-dependent regulations.

2.3.3 Corruption

Of concern is, another explanation for the missing middle, corruption. Many studies have shown that burdens of bribery can reshape the patterns of firm's growth in developing countries, resulting in a U-shaped distribution of firm size.

In the context of high-level corruption, large-scale firms may enjoy advantages by lobbying civil servants. This is not only because large firms can effectively maintain contact with officials because of their long-term relationships but also because costs of dealing with regulation burdens may be fixed while the pay-offs increase with the scale of operation. As a result, micro-firms and small firms are reluctant to join the formal sector, fearing regulatory and corruption burdens. In such an environment, middle-sized firms, which are big enough to be visible to officials but too small to enjoy economies of scales, are the most vulnerable to these constraints.

Rausch (1991) presented a a Lucas (1978)-type model that generated the missing middle in FSD attributable to entrepreneurs' decision to operate in, either the informal or the formal sector. In the informal sector, firms avoided regulations (e.g. minimum wage) and rent-seeking activities. However, in order to avoid interacting with the government, they had to remain below a certain size threshold. Thus, those firms that decided to formally register always operated well above the threshold because it never payed to be just large enough to suffer regulations. Emerson (2001) also illustrated such an explanation using the Cournot-type oligopoly model. In Emerson's model, the number of formal firms was endogenized as a function of rent-seeking activities, which in turn depended on the level of corruption in the country. The higher the degree of corruption, the fewer opportunities emerge for larger firms. Giovanni and Goyette (2009) suggested the existence of the missing middle was based on higher governmental rent in Uganda.

2.3.4 Some other factors

Other factors noted in the literature may indicate the missing middle in firm size distribution. One plausible explanation is that limitation to infrastructure access can segment markets and insulate local producers in developing countries, resulting in weak competition and smaller average firm size. Eifert (2007) developed a framework in which a low provision of intermediate inputs such as electricity motivated incumbent firms to bargain with the electricity supplier for preferential treatment. Under weak conditions they colluded together to extract surplus from consumers, resulting in even greater market concentration, lower total output and higher prices.

In another explanation for the missing middle, Dasgupta (2010) developed a model that could generate a dualism by considering agents' optimal choice between the traditional and modern sector. Dasgupta concluded that the dual structure was therefore not stationary but dynamic, and further that such a dual distribution would eventually converge to the uni-modal distribution as the economy grew.

Marchiavello (2010) explored the relationship between the missing middle and both the financial constraint and vertical integration. Marchiavello developed a model in which vertical integration was more likely among intermediate levels. Contract enforcement can help to reduce vertical integration once there is no barriers in financial market. In addition, the missing middle tend to arise in industries that favored vertical integration.

Bakhtiari (2013), in contrast, relied on the outsourcing decision of firms to explain the narrowing-down pattern of the firms in the middle of the distribution. According to Bakhtiari, small firms tend to choose outsourcing to avoid extra overhead costs and risks associated with large-scale production. The model developed by Bakjtiari can also accommodate outsourcing decisions of large-scale as long as their sizes are not too large. In contrast to Dasgupta's model, Bakjtiari 's model predicted the long-term bimodal distribution of firms in the economy. This explanation, however, is open to criticism because it fails to explain why the phenomenon happens only in developing countries because same forces exist in developed economies.

2.4 Corruption and Firm Size Distribution

Corruption is one of the most pervasive obstacles to economic and social development. As mentioned in the previous section, that has been considered as one possible explanation for the missing middle in many developing countries. Since in the later chapters of the thesis, we analyzed the effect of corruption on firm size distribution, this section is, therefore, devoted to providing an overview on the corruption situation as well as its effect on firm size distribution.

Bureaucratic corruption is pervasive throughout the world. The relationship between corruption and economic growth has been broadly studied in the literature. Many studies provided empirical evidence on the negative impact of corruption on economic growth and development Knack and Keefer (see 1995), M'eon and Sekkat (see 2005). Other studies also investigated the principal transmission mechanism through which corruption reduces investment and hence, hampers economic growth Mo (see 2001a). However, most of these papers use indexes on the perception of corruption. Few studies, however, investigate the issues using micro-firm data. McArthur and Teal (2004), which examine the effect of corruption on the performance of African firms with data from the Africa Competitiveness Report 2000/1. They find that firms paying bribes experiences a 20% reduction in output per worker and that firms in countries with higher corruption are 70% less efficient. Svensson and Fisman (2007), on the other hand, showed that an increase of 1% in the bribe rate can decrease firms' growth by 3%.

The incidence of corruption, however, tend to vary considerably across countries, conditional on other social and economic factors. Neeman et al. (2008) found the existence of negative relationship between corruption and growth only for countries with a high degree of financial openness. Aidt et al. (2008) also provided evidence that in countries with high quality of political institutions, corruption tended to be more harmful to growth. Such a finding forms the basis of what Wedeman (2002) termed the ”East Asian paradox”: many countries in East and South-east Asia such as China, Indonesia, South Korea and Thailand, have grown remarkably well in spite of high levels of corruption. The general idea is that corruption facilitates beneficial trades that would otherwise not have taken place. In doing so, it promotes efficiency by allowing individuals in the private sector to correct pre-existing government failures of various sorts. Whether these countries are affected by corruption and whether corruption does have negative impact on economic development? One possible answer may lie in the effect of corruption on the firm size distribution.

The effect of corruption on firms can be classified into different categories. First, corruption may change the effective marginal tax rate faced by firms. Svensson (2003) estimated that that each USD 1.00 in firm profits per employee leads to about USD 0.004 in additional bribes paid, for a marginal bribe rate of 0.4 percent on profits.Romer (1994) also viewed corruption to be the tax that hinders the entry of new technology and new goods. Second, many studies showed that argued that the uncertainty with corruption makes it more costly than an equivalently-sized tax. In addition, Sequeira and Djankov (2010) suggested that about 46 percent of South African firms located in regions in which overland costs to the port of Maputo are 57 percent lower go the long way around to Durban to avoid higher bribe payments. This represents another type of corruption distortion: changes in firms' optimal choice of production to avoid corruption.

In the mid-1980s, the sociologist Hernando de Soto organized a very interesting social experiment in Peru (de Soto, 1990). He was concerned about the enormous size of the informal sector of the economy wanted to understand why. He conducted a simple experiment by completing one task: obtain all the permits and approvals needed to start a small two-sewing machine garment factory in a Lima town legally and record the time and effort required in doing this. The result of the experiment was striking. It took about 300 days of 6 hours per day or 32 times the monthly minimum wage to complete the task. It is now easy to see why most firms prefer to stay in the informal sector in such a corrupted market.

In a related literature on informality, Johnson et al. (2000) provided evidence that firms in Russia and Ukraine tend to hide their activities as compared to managers from Poland, Slovakia and Romania because they face higher tax rates, worse bureaucratic corruption, a greater incidence of mafia protection and have less faith in their court system. Ingram et al. (2007) showed that formality is negatively correlated with the rate of taxation and corruption using data from the

World Bank on firms from six African countries which include Uganda. Dabla-Norris et al. (2008), using the World Business Environment Survey data, showed that informality is negatively related to firm size and that, although heavier regulations may conduct to higher informality, this needs not be so in countries with strong rule of law.

In the literature, research that directly investigates on the impacts of corruption on different firm size groups has been scarce[13]. Meanwhile, size has been proven to be a significant factor in firm growth and performance. On the one hand, studies on large companies, or SMEs, and their effect on growth produced contrasting findings. On the other hand, there have been few comparative studies that provide information about all three types of companies at the firm level. Furthermore, the heterogeneous effect of corruption on firm growth based on size remains largely unexamined.

In this thesis, we focus on analyzing the impact of corruption as a barrier for the development of firms. In particular, firms in the market may be forced to pay bribes to officials as civil servants have some discretionary powers to enforce rules that can affect firms' business. Cumbersome officials can defer granting licenses or permits, slow down the operation process, restrict access of credit or create bureaucratic delays until receiving bribes. Officials, in this sense, act as price discriminators who try to extract the highest possible bribes subject to the constraint of being got caught or the possibility of firm exit. The distortion caused by such activities can be considered an extra tax on profits of enterprises. In fact, small firms face two contradicting forces: on the one hand, firms want to grow to take advantage of economies of scale and better access to finance, on the other hand, small firms can save costs as they become invisible to officials and thus can escape from giving bribes. In the context of heavy regulatory and corruption burdens in developing countries, small firms may prefer to remain small until there exists large enough productivity shocks so that the first effect overwhelms the second one.

[13] Among studies on the impact of corruption on firm size, Beck et al. (2005) provided evidence that financial, legal, and corruption constraints depended on firm size, and that smallest firms were most constrained in 54 countries in the World Business Environment Survey (WBES). Athanasouli et al. (2012) provided evidence that small and medium firms display a higher engagement in corrupt practices. However, their performance is less correlated with corruption than that of large firms

2.5 Conclusion

Taken as a whole, the literature on the size distribution falls into one of two categories. One strand of statistically based research focuses on the distributional properties of firm size. The other strand consists of recent research integrating the size distribution of firms into standard economic theory. However, the dearth of empirical research persists.

Despite researchers' various attempts in the last century, theories of modeling FSD have not yet reached a convergence on the factors that cause the missing middle. Such a diversification in the literature confirms the complexity of the issue. This chapter provided a summary of the controversial debate.

The most important findings of this chapter are summarized as follows:

- The definition of firm is multiple and complex;
- The FSD evolves from a complex process that is affected by both internal and external factors and firm size may follow a random walk;
- The FSD of developed economies can be approximated by either the Pareto or log-normal distributions or a combination of the two, and in many cases, can be approximated by a straight line with downward slope in accordance with Zipf's law;
- Many theoretical models have been developed to capture the mechanism behind the FSD formation, but as yet no complete and satisfactory explanation exists for the emergence of the FSD;
- A significant difference exist between FSDs of developed and developing countries. The FSD in developing countries exhibits the missing middle, whereby a small number of giant firms co-exists with many small firms and the medium-sized firms are under-represented. Such a dual structure has been attributed to distinct characteristics of the business environment in developing countries such as regulatory framework, access to finance, and corruption levels, among others;
- Many factors have been identified to explain the missing middle. However, no unified agreement as been reached on the factors that can satisfactorily explain empirical findings.

Given the preceding summary and discussion, the thesis focused on examining the “missing middle” phenomenon in a developing country - Vietnam.

Vietnam was a relevant case to study because Vietnam is a transitional developing country. Over the last two decades, with the emergence of the “Doi moi” (Renovation) process, Vietnam has experienced a period of continuously high economic growth. However, to date, Vietnam has failed to become a tiger in Asia as expected. One of the possible explanations for this failure is the under-representation of medium-sized firms, a segment commonly contains the most dynamic and growth-oriented firms. Serious consideration, therefore, should be given to clarify the mechanism behind such a bias in FSD because this under-representation of middle-sized firms underpins the mis-allocation of resources and hinders firm growth.

As in the case for other developing countries, studies on the FSD in Vietnam are scarce. Related studies, if any, have been limited to empirical studies only. None gives a comprehensive picture of the FSD or tests for the validity of Zipf’s law in the Vietnamese situation. Therefore, this research was a timely and essential opportunity to investigate the FSD in Vietnam, both empirically and theoretically.

CHAPTER 3

THE MISSING MIDDLE - A PHENOMENON OF FIXED COST

3.1 Introduction

As noted in the previous chapter, firm density is leptokurtic, stable, and highly skewed to the right. Moreover, the FSDs of developed countries are commonly characterized by either the log-normal or Pareto distribution and tend to follow Zipf's Law. Firms in developing countries, however, operate in different business environments characterized by dense burdens of business regulations, corruption and financial restrictions. This environment in developing countries leads to the deviation of FSDs from the log-normal, Pareto distribution or Zipf's law. Interestingly, in many developing countries most employment is located in either small-sized or large-sized firms, resulting in the well-studied empirical fact that the distribution of firm size in developing countries tends to exhibit a missing middle.

The missing middle phenomenon was first documented in Liedholm and Mead (1987). In addition, Tybout (2000) found a large spike in the size distribution for the small-sized category, which dropped off quickly in the middle-sized category among the poorest countries. Tybout argued that strong business regulations could be a reason for the existence of too many small firms[1].

[1] See also Little et al. (1987) for South Korea and Steel and Webster (1992) for Ghana. The presence of the missing middle phenomenon and how we define it have been intensively discussed in the literature (see Hsieh and Olken, 2014, Tybout, 2014a,b).

Given the missing middle is more evident in developing nations (see Tybout, 2000), one would expect to find some explanation that particularly applies to developing nations, but not so much to developed nations. In this study, our objective was to identify possible reasons for the missing middle phenomenon, which is more evident in developing countries than in developed countries. The difference in our approach from the approach documented in previous literature[2] is that rather than looking at taxes or regulations themselves, we studied corruption as a manifestation of the business environment in developing nations.

No matter how we define government corruption, it is higher in developing nations, compared to developed nations. According to the Transparency International Corruption Perception Index report, most developing countries are at the bottom of the corruption list (e.g., Indonesia, Tanzania), in contrast, developed economies such as New Zealand, and Denmark are in the top 20.[3]

In this chapter, our main goal is to develop and present a theoretical model that could capture the impact of corruption on FSD formulation in developing countries.

As stated in (Kwok and Tadesse, 2006), “Corruption is generally defined as the abuse of public power for private gain (Alvaro, 2006)...[or] an arrangement that involves an exchange between two parties, the “demander” and the “supplier”, which has an influence on the allocation of resources and involves the use or abuse of public or collective responsibility for private ends” .

Two major approaches to modeling corruption have been discussed in the literature. The first is the agent-principle model type in which information asymmetry and the problem of enforcement cause the principal (the government) to hold the agent fully accountable for its actions (See Brandt and Svendsen, 2013, Groenendijk, 1997). The second approach is the “resource allocation” model type in which corruption is considered a production cost (See Ernesto and Rossi, 2007, Tullock, 1993). The first model type is normally used to explain the grand corruption type or legislative corruption (corruption that occurs at the highest levels of government requiring significant subversion of the political, legal and economic systems). The second model type seems to best explain petty corruption (corruption that occurs at a smaller scale within

[2]A more detailed discussion on different factors that have impacts on FSD formulation can be found in the literature chapter

[3]Evidence exists that corruption levels are significantly higher for developing and transitional countries (e.g., Nowak, 2001). Some studies (see He, 2000, CIEM, 2006)[4] suggest that the high level of corruption in these countries is perhaps due to “(i) abuse of power by public officials, (ii) arbitrary decisions related to policies and administration, (iii) weak accountability of officials and government agencies, and (iv) weak state implementation and monitoring.”.

established social frameworks and governing norms).

Between the two types of corruption, the second one is particularly common in developing countries and in situations in which public servants are significantly underpaid. As Jain (2005) stated, “corruption can be analogous to a tax, ...[and] such change in costs will cause a shift of the market equilibrium”. To function in the market, Firms may be forced to pay bribes to officials because civil servants have some discretionary powers to enforce rules that can affect firms’ operation. Officials, in this sense, act as price discriminators who try to extract the highest possible bribes subject to the constraint of being caught or the possibility of firm exit. Small firms, facing such coercion, may refrain from getting bigger to escape the rent-seeking activities of civil servants.

In this chapter, we follow the “resource allocation” model type to assess the impacts of petty corruption on the firm development process and on the FSD formulation of the Vietnamese economy. The model captures the increasing costs of middle-sized firms, compared to the costs of small-sized and large-sized firms. This increasing cost for middle-sized firms leads to the concentration of a continuum of infinitesimal firms in the market as well as a considerable number of large-scale firms. In particular, we extend the model presented in Luttmer (2007). Luttmer noted that the model is more applicable to a developed economy such as the one in the United State. With our model, in contrast, we seek to develop a framework that could be applied to a developing economy. Economic growth in the model is generated from firm-specific preference, technology shocks, and the selective survival mechanism of successful firms. Our model integrates the firm dynamics model based on Gibrat’s law and the theory of corruption to generate a general equilibrium and balanced growth that is consistent with salient features of the particular FSD of Vietnamese economy.

In this combined framework, productivity drives firm growth. Market penetration costs introduce important non-homogeneities into the process. The fixed cost, which is assumed to be fixed in Luttmer’s model, is assumed to be a function of size in our model. This assumption captures the empirical fact that corruption and other regulatory burdens tend to increase for medium-sized firms in developing countries. A further discussion of fixed cost is included to emphasize the difference between the two models.

In addition, in the model, the analyzed economy is a monopolistic competition market. Time is assumed to be continuous.

On the consumption side, following Dixit-Stiglitz, we assume that there is a continuum of infinitely-lived consumers and each of them supplies inelastically one unit of labor at every point of time. Consumers have a "taste for variety" in the sense that they prefer to consume a diversified bundle of goods. In the economy, we consider a representative consumer with preferences over rates of dynastic consumption. The consumer maximizes the utility derived from the composited good made up by a continuum of differentiated goods. Her preference over these commodities are additively separable. The consumer will consume the same amount of commodities of the same type and with same prices. The consumer faces a standard present value budget constraint in which income consists of claims from firms and labor wages.

On the production side, firms are monopolistic competitors. There are many producers, and each has a certain market power but to a certain extent, the free entry condition limits firms' profit opportunities. Firms produce differentiated goods using linear, constant return-to-scale technologies. Firms have heterogeneous productivities that evolve stochastically over time. Potential entrepreneurs must pay certain costs to enter the market whereas incumbents must pay a fixed cost per unit of time to continue. All costs are measured by labor.

Active firms in the market are subject to firm-specific, permanent shocks to both productivity and demand for their differentiated commodities. In particular, the states of consumers' tastes and firms' productivities are assumed to follow a geometric Brownian motion which has a mean and variance growth rate independent of size. Later, these processes translate into the selection mechanism of firm distribution of productivity over time, which is crucial for the development of the stationary FSD in the economy.

Technologies available to entrants are assumed to develop at an exogenous rate, which later determines the economic growth rate. A stationary FSD arises as productivity and demand shocks improve the productivity of existing firms at an average rate that is not too high compared to the rates encountered by potential entrants. One possibility is that the logarithm of firm size follows the Pareto distribution.

The general equilibrium in the economy, then, is characterized by the following sequences: (a) consumers maximize their inter-temporal utility by choosing a series of dynastic consumptions of a composite good, subject to the inter-temporal budget constraint (b) firms maximize their profits taking the price index as given; (c) a closed-form stationary distribution of firm size results; and; (d) labor and goods markets clear.

Related to the costs of firms, a proportional increase in both entry and fixed cost lowers the output level, thus reducing the number of firms and the variety of goods. These events lower the average size and profitability of firms and also slow down the selection mechanism by which productivity improves overtime. These changes, however, do not affect the shape of the size distribution. On the other hand, a fluctuation in fixed cost alone can change the shape of the FSD given that such a variation is not uniform across different size groups. As fixed cost increases, the value of the firm decreases and the exit barrier increases. This situation also leads to a decrease in the size density for the middle-sized group. In other words, the increase in fixed cost implies a smaller number of firms in the middle-sized group, which is empirically the case in many developing countries.

The importance of the findings rests on the asymmetry of the growth opportunities for small firms and large firms in a developing market. This study is one of the few to to provide evidence for the inverse relationship between growth and the initial firm size (i.e. violations of Gibrats law). Evidence for this violation is cited as justification for differential treatment for small businesses, an issue that raises attention in both academic and policy circles. Further, the detailed calculation of the model, which is not available in Luttmer's paper is provided in the Appendix.

The structure of the chapter is as follows. The second section is devoted to the set up of the model, including the description of consumers and firms in the economy and the decision of firm entry and exit. In the second section, we report on the firm size distribution. The third section focuses on the competitive equilibrium and balanced growth path. The fourth section provides further discussion on fixed cost. Conclusions are presented in the last section.

3.2 Model set up

3.2.1 Preferences

Let us assume that the economy is populated by a mass L of identical individuals who work, consume and own firms[5]. The economy can be represented with a single representative consumer who is an aggregate of many consumers with distinct individual preferences for each va-

[5] Although labor force only comprises of a part in total population, in developing countries, where the population tend to be young, labor force contribute a major proportion of the total population. Moreover, in long-term, there is a strong correlation between the changes in population and total labor supply.

riety of goods and maximizes the utility derived from the dynastic consumption stream $\{C_t\}_{t\geq 0}$ of the continuum of differentiated goods. The utility is given by[6]:

$$(\mathrm{E}[\int_0^\infty \rho \exp(-\rho t)[C_t \exp(-\eta t)]^{1-\gamma} \mathrm{d}t])^{\frac{1}{1-\gamma}}, \tag{1}$$

where the discount rate ρ is assumed to be positive and the population growth rate η is assumed to be non-negative. The population size at time t is $N_t = H\exp(\eta t)$ where H is the initial population. The population is equal to the total supply of labor in the market.

Let $c_t(u,p)$ denote the consumption of a product type u at price p. At some $\beta \in \{0,1\}$, consumption for the composite goods is a constant elasticity of substitution (CES) aggregator of the set of differentiated goods which are available at

$$C_t = \int u^{1-\beta} c_t^{\beta}(u,p) \mathrm{d}M_t(u,p), \tag{2}$$

where M_t is the measure of commodities that are available at time t, defined on the set of commodity types u and price p and $\mathrm{d}M_t$ is the density of particular goods of type u at price p. The level of $c_t(u,p)$ is chosen to minimize the cost to acquire C_t, which implies, at the equilibrium:

$$pc_t(u,p) = P_t(uC_t)^{1-\beta} c_t^{\beta}(u,p). \tag{3}$$

Then, the price index is:

$$P_t = [\int up^{\frac{-\beta}{1-\beta}} \mathrm{d}M_t(u,p)]^{\frac{-(1-\beta)}{\beta}}. \tag{4}$$

and the demand for the good type u is:

$$c_t(u,p) = [\frac{P_t}{p}]^{\frac{1}{1-\beta}}.[uC_t]. \tag{5}$$

Along the balanced growth path, per capita consumption and real wages grow at a common rate of κ. In particular, $C_t \exp(-\eta t) = C\exp(\kappa t)$ and $w_t = w\exp(\kappa t)$, respectively. The interest rate is also assumed to be constant at $r = \rho + \gamma\kappa$.

[6]This type of utility function allows us to model many different consumers' preferences. In particular such a function can be used to model various characteristics such as preference towards variety, substitutability versus complementarity, and multiplicity of price equilibria, just by adjusting γ and C_t. Preferences over commodities are additively separable with weights that define the type of commodities

The following assumption is adequate to ensure that the present value of aggregate consumption, labor and utility are finite.

Assumption 1. *The growth rate η and κ satisfy $\eta \geq 0$ and $\rho + \gamma\kappa > \kappa + \eta$. where $\rho + \gamma\kappa = r$ and r is the interest rate in unit of the composited good*[7].

3.2.2 Production

A firm is distinguished by its unique access to technology to produce particular differentiated goods. Assume that at time t, each firm of age (a) has a productivity level of $z_{t,a}$. The firm that was set up at time t uses $L_{t,a}$ units of labor to produce $z_{t,a}L_{t,a}$ units of the commodity of quality $u_{t,a}$. Given the price $p_{t,a}$, the revenue of the firm is $R_{t,a} = p_{t,a}z_{t,a}L_{t,a}/P_t$. Substituting into equation (3), the revenue can be re-written as:

$$R_{t,a} = C_{t+a}^{1-\beta}(Z_{t,a}L_{t,a})^{\beta}, \tag{7}$$

where $Z_{t,a} = (u_{t,a}^{1-\beta} z_{t,a}^{\beta})^{1/\beta}$.

We can see that $Z_{t,a}$ combines the state of preferences and technology and represents the productivity level of the firm. In this model, we simply refer $Z_{t,a}$ to productivity. $Z_{t,a}$ is assumed to evolve independently across firms according to:

$$Z_{t,a} = Z\exp(\theta_E t + \theta_I a + \sigma_Z W_{t.a}), \tag{8}$$

where θ_E and θ_I denote the productivity growth rate of new entrants and the productivity growth of incumbents, respectively, Z is the initial condition and $[W_{t,a}]_{a\geq 0}$ is a standard Brownian motion[8]. The Brownian Motion assumption arises because the continuous time limit of the

[7] The optimal consumption growth satisfies the Euler condition:

$$\begin{aligned} r &= \rho + \gamma(\tfrac{DC_t}{C_t} - \eta) \\ &= \rho + \gamma[(\kappa + \eta) - \eta] \\ &= \rho + \gamma\kappa, \end{aligned} \tag{6}$$

where r is the constant interest rate in unit of the composited good.

[8] Brownian motion was first observed by a English botanist, Robert Brown who realized the haphazard zigzag movement of microscopic pollen grains suspending in a drop of water. A satisfactory explanation for such phenomenon was given by Einstein in 1905 but it was not until 1923 that Nobert Wiener provided a clear picture of the Brownian Motion stochastic process by setting the modern mathematical foundation for such a process. Today, Brownian Motion and its generations and extensions occur in various areas of pure and applied science including

firm growth rate is assumed to be a discrete random walk. The formula implies the geometric Brownian motion form of productivity[9]:

$$\frac{\mathrm{d}Z_{t,a}}{Z_{t,a}} = \theta_E \mathrm{d}t + \theta_I \mathrm{d}a + \sigma_Z \mathrm{d}W_{t,a}. \tag{9}$$

The assumption of geometric Brownian motion implies that the expected growth rate of firms' productivity is similar for all incumbent firms and is size independent. This, together with the assumptions on production technology and Dixit-Stiglitz demand specification, causes the identical expected growth rates of sales per consumer across all incumbent firms. The specification also guarantees the high persistence of firm-level shocks and implies the validity of Gibrat's law, in which the productivity growth rate is independent of the current firm size. Moreover, as we will show later in this chapter, the geometric Brownian motion also leads to the Pareto distribution of firm size along the balanced growth path.

In the market, firms first choose variable labor to maximize profit $\pi_{t,a} = R_{t,a} - w_{t+a}L_{t,a}$, where firm revenue is given by Eq. (7). Taking the first derivative and rearranging, the optimal choice for this profit maximization problem is:

$$\begin{bmatrix} R_{t,a} \\ w_{t+a}L_{t,a} \end{bmatrix} = \begin{bmatrix} 1 \\ \beta \end{bmatrix} (\frac{\beta Z_{t,a}}{w_{t+a}})^{\beta/(1-\beta)} C_{t+a}. \tag{10}$$

This, together with Eq. (7) and (8) implies that along the balance growth path, the revenue and labor of a firm are independent of calendar time.

In addition, we also assume that in each time period, an incumbent with size s must pay fixed cost $\lambda_F(s)$ units of labor to maintain operation. In this chapter, we assume that the fixed cost not only covers the lump-sum taxes/fees as well as any direct and indirect expenditures to implement the regulations that each firm must pay to the government but also includes the amount of bribes paid to officials[10]. In other words, fixed cost refers to the level of corruption and regulatory burdens faced by each firm in the market.In contrast to Luttmer (2007)'s setting, fixed cost in our model is a function of size[11].

economics, communication, biology, physics, etc. (Taylor and Karlin, 1998)

[9]Such process can evolve in case that both $u_{t,a}$ and $z_{t,a}$ are geometric Brownian motions.

[10]The fixed cost can represent the cost to preserve acquired information (see Luttmer, 2007) or the cost of a managerial fixed factor (see **?**).

[11]A detailed description and analysis on fixed cost is provided in Section 3.5.

Then, along the balanced growth path, the revenue net fixed and variable cost can be written as[12]:

$$R_{t,a} - w_{t+a}(L_{t,a} + \lambda_F(s_a)) = w_{t+a}(\exp(s_a) - \lambda_F(s_a)), \tag{11}$$

where:

$$s_a \;=\; S[Z] + \frac{\beta}{1-\beta}[\ln(\frac{Z_{t,a}}{Z_{t,0}}) - \theta_E a]; \tag{12}$$

$$\exp(S[Z]) \;=\; (1-\beta)\cdot\frac{C}{w}\cdot\left(\frac{\beta Z}{w}\right)^{\beta/(1-\beta)}. \tag{13}$$

We see that, $S[Z]$ relates the de-trended productivity of a firm to its size and thus $\exp(S[Z])$ represents the size of firm with productivity of $Z(\exp(\theta_E t))$ at time t and Z is the de-trended initial productivity of the firm.

Moreover, because both revenue and variable cost are proportional to $w_{t+a}.\exp(s_a)$, we can consider s_a a measure of firm size, relative to wage rate, and s_a evolves with age according to

$$\mathrm{d}s_a = \mu \mathrm{d}a + \sigma \mathrm{d}W_{t,a}, \tag{14}$$

where

$$\begin{bmatrix} \mu \\ \sigma \end{bmatrix} = \frac{\beta}{1-\beta}\begin{bmatrix} \theta_I - \theta_E \\ \sigma_Z \end{bmatrix}. \tag{15}$$

μ and σ in Eq. (15) represents the trend and variance of firm size. As long as incumbents have higher productivity growth rate than do new firms, firm size will have a positive growth trend. The variance of firm size moves in the same direction with the change in the variance of productivity (σ_Z). Moreover, the trend and variance of productivity shocks tend to be magnified when goods are close substitutes.

In the market, the aggregate supply of labor grows at the rate of η and incumbents must pay a fixed cost $\lambda_F(s)$ to operate in the market. The number of firms in the market also grows at the rate of η at the balanced growth path, and entry and exit generate a time-invariant distribution of labor input and productivity.

In addition, considering Eq. (7), because the per capita consumption growth rate is κ, an aggre-

[12]The detailed calculation can be found in the Appendix.

gate consumption growth rate is $\kappa + \eta$. Moreover, because the number of firms grows at the rate of η, the growth rate of per capita consumption must equal the growth rate of average revenues per firm. Taking the logarithm of Eq. (7) in both sides and rearranging terms yields:

$$\kappa = \theta_E + (\frac{1-\beta}{\beta})\eta.^{13} \tag{16}$$

3.2.3 Entry and Exit Decision

Exit Decision

Introducing idiosyncratic shocks in the model implies that firms have non-constant operational profits and as a result, there is an endogenous exit margin. On the one hand, incumbent firms must decide at each period whether to remain in the industry. The exit decision is now equivalent to asking if the firm's optimal value is non-positive. On the other hand, if there are shocks and endogenous exits, firms are more eager to enter because they are not stuck with a particular productivity level.

Readers should note that the existence of fixed cost guarantees the minimum size of firms because firms with not enough productivity are not able to cover the fixed cost and will eventually exit the market. An additional assumption is needed:

Assumption 2. *The fixed cost satisfies* $\infty > \lambda_F(s) \geq \bar{\lambda}_F > 0$.

Assumption 2 ensures a minimum firm size below which the firm will exit because it cannot cover the fixed cost to operate in the market. The assumption also ensures that the fixed cost does not grow without bound.

Because the incumbent with size s must pay fixed cost $\lambda_F(s)$ units of labor every time period to remain in the market, the value of a firm at time t with initial productivity $Z\exp(\theta_E t)$ is:

$$V_t[Z] = \max_{L,\tau} E_t[\int_0^\tau \exp(-ra)(R_{t,a} - w_{t+a}[L_{t,a} + \lambda_F(s_a)])\mathrm{d}a], \tag{17}$$

where revenue and productivity are defined in Eq. (7) and (8), respectively; τ is the stopping

[13] An exercise to measure the magnitudes of impact of the increase in variety, technology progress and selection to firm growth can be done using equation 16. The detail of the exercise can be found in the Appendix.

time and the process is martingale (i.e. production and exit decision of a firm depends on available information).

In addition, because $w_a = \exp(\kappa a)$, when $t = 0$, substituting from Eq. (11) we have:

$$\begin{aligned} \exp(-ra)(R_{0,a} - w_a[L_{0,a} + \lambda_F(s_a)]) &= \exp(-(r-\kappa)a)\exp(-\kappa a)(\exp(s_a) - \lambda_F(s_a) \\ &= \exp(-(r-\kappa)a)(\exp(s_a) - \lambda_F(s_a)). \end{aligned}$$

Then, the value of a firm of size s can be written as:

$$V(s) = \max_{\tau} E[\int_0^{\tau} \exp[-(r-\kappa)a](\exp(s) - \lambda_F(s_a))\mathrm{d}a | s_0 = s], \tag{18}$$

where $V(s)$ is the value of a firm size s.

To ensure that the firm value is finite, we need the following assumption:

Assumption 3. *The preference and technology parameters satisfy* $\rho + \gamma\kappa > \kappa + \mu + \frac{1}{2}\sigma^2$.

Assumption 3 implies that the interest rate is greater than $\kappa + \mu + \frac{1}{2}\sigma^2$. This, together with the assumption of finite fixed cost, ensures that the revenue of firm is finite. Thus, Assumptions 2 and 3 are adequate to guarantee that the value of the operating firm is finite.

The firm solves the problem of choosing the optimal stopping time such that a reward function is: $f(t, x) = \exp(-\bar{\rho}t) \cdot \phi(x)$ where $\bar{\rho} = r - \kappa$ and $\phi(x) = \exp(s) - \lambda_F(s_a)$.

The solution of such a problem must satisfy the Hamilton-Jacobi-Bellman equation:

$$(r-\kappa)V(s) = \begin{cases} (\exp(s) - \lambda_F(s)) + V'(s)\mu + \frac{1}{2}V''(s)\sigma^2 & \text{if } s \geq b \\ 0 & \text{otherwise} \end{cases}. \tag{19}$$

with the following boundary conditions:

- Value matching condition: $V(b) = 0$. The firm value equals zero at the exit barrier b (i.e. the firm exits at b);
- Smooth pasting condition: $V'(b) = 0$. The value function is differentiable at b;

- Firms' Value Condition: $V(s) < \frac{\exp(s)}{r-[\kappa+\mu+\sigma^2/2]}$. The value function cannot exceed the value of the firm operating without fixed cost or $V(s)$ must lie below $\exp(s)/[r-(\kappa+\mu+\sigma^2/2)]$.

The first equation of (19) is a linear, second order non-homogeneous differential equation of the type:

$$ay''(s) + by'(s) + cy(s) = g(s). \tag{20}$$

Solving the equation, we get the result shown in the following Proposition

Proposition 3.1. *Given Assumption 2 and Assumption 3, a firm's exit decision is dependent on firm's value and the exit threshold, which are respectively,*

$$\begin{aligned} V(s) &= \tfrac{1}{r-\kappa}(\tfrac{\xi}{1+\xi})(\exp(s-b)-1-\tfrac{1-\exp(-\xi(s-b))}{\xi})\textit{ for } s \geq b. \\ \exp(b) &= (\tfrac{\xi}{1+\xi})(1-\tfrac{\mu+\sigma^2/2}{r-\kappa}). \end{aligned} \tag{21}$$

Proof. The detailed calculation of these equations can be found in the Appendix. □

0 Because $\xi > 0$ and b is well defined (Assumptions 1 and 3), the value of a firm increases in the interval (b, ∞), which depends on the μ and σ. In other words, the value of the firm increases as the difference between productivity growth rates of incumbents and new firms increases and decreases as the variance of productivity shocks increases. Because new entrants have very high productivity growth rate, compared to incumbents, the value of the incumbent firm will be very close to the exit barrier b.

Entry Decision

In accordance with Luttmer (2007), we assume that firms enter the market as a result of a drawn from the initial productivity level Z, which follows the exogenous J distribution. A firm entering at time t will have initial productivity of $Z\exp(\theta_E t)$ and the initial size of $S[Z]$. Potential entrepreneurs who want to enter the market must pay an entry cost of λ_E which is linear at the entry rate. The zero profit condition implies

$$\lambda_E = \int V(S[Z])dJ(Z). \tag{22}$$

Because $S[Z]$ is proportional to λ_E, at the equilibrium, if the entry cost λ_E is very high, the initial size and productivity of firms in the market must also be high.

The only condition needed to ensure the value of entry is finite is the following assumption:

Assumption 4. *The initial productivity distribution satisfies* $\int Z^{\frac{\beta}{1-\beta}} dJ(Z) < \infty$.

3.3 The Stationary Distribution

Considering a market containing numerous firms, the assumption of stochastic growth rate of productivity guarantees that the probability distributions for individual firm size is the same as the cross-section size distribution in the economy.

Along the balanced growth path, firms enter and exit at a constant rate, giving rise to a time-invariant distribution of firm size so that the aggregate measure of firms (M) expands at the rate η. Given that η is non-negative, the following assumption ensures that the mean of firm size distribution is finite.

Assumption 5. *The productivity parameters satisfy* $\eta > \mu + \frac{1}{2}\sigma^2$.

Given that $\mu + \frac{1}{2}\sigma^2$ is the drift of size variable, then, Assumption 5 implies that the size of an incumbent does not grow faster than the population growth rate. If η equals zero, μ must be negative. Thus, with a positive population growth rate, the mean of productivity growth is always positive.

Consider the measure of a firm, defined on a set of all possible age a and size s. This measure grows at the rate of η. Let the density of the measure at time t be $m(a, s)I \exp(\eta t)$ where $I \exp(\eta t)$ is the number of new firms attempting to enter per unit of time. Then, at each point of time, there is a measure of new firms entering the market with the initial size of x. Upon entering, each firm faces productivity shocks that cause firm size to evolve according to the stochastic Brownian Motion process in Eq.(14). In other words, for each age level, a corresponding density exists that describes how firms reach a certain level of productivity and thus a certain size.

Given an initial distribution, the density, then, must satisfy the partial differential equation - the

Kolmogorov Forward equation that is given by:

$$D_a m(a,s) = -\eta m(a,s) - \mu D_a m(a,s) + \frac{1}{2}\sigma^2 D_{ss} m(a,s)\ \forall a > 0 \text{ and } s > b, \tag{23}$$

subject to the following boundary conditions:

- First boundary condition:

  $$\lim_{a \downarrow 0} \int_b^s m(a,x)\mathrm{d}x = G(s) - G(b), \tag{24}$$

  where G denotes the size distribution among entrants when age goes to zero. The lower bound of the size distribution is achieved when age goes to zero and the distribution resembles the size distribution of new entrants. Such a distribution follows the productivity distribution J among new entrants via $J(Z) = G(S[Z])$.

- Second boundary condition: $\forall a > 0$,

  $$m(a,b) = 0. \tag{25}$$

  A firm must exit when its size reduces to the threshold b and none should enter at a size below b.

Because firms must pay at least a fixed cost $\lambda_F(s)$ to exist, the measure of firms has to be finite in the equilibrium. The solution to Eq.(23), subject to the two boundary conditions, is given by:

$$m(a,s) = \int_b^\infty \exp(-\eta a)\psi(a,s|x)dG(x), \tag{26}$$

where

$$\psi(a,s|x) = \frac{1}{\sigma\sqrt{a}}[\phi(\frac{s-x-\mu a}{\sigma\sqrt{a}}) - \exp[-\mu(x-b)/(\sigma^2)]\phi(\frac{s+x-2b-\mu a}{\sigma\sqrt{a}})]. \tag{27}$$

Detailed calculation can be found in the Appendix.

Note that $m(a,s)$ reduces to $\exp(-\eta a)\psi(a,s|x)$ as G is replaced by a distribution concentrated at x and $\exp(-\eta a)\psi(a,s|x)$ is the density of firm size and age among firms with initial size x.

In general, $m(a, s)$ represents the stationary density of the set of incumbents operating in the equilibrium over different levels of ages and sizes. At the equilibrium, a firm is characterized by a certain size and age. The firm enters the market with the initial productivity level Z at size x and evolves following the realization of Eq.(14). The firm exits as its size falls below b and will be replaced by a cohort of new firms entering at x. The density of each age level is described by $m(a, s)$. The marginal density of $m(a, s)$ with respect to s captures the size density of all firms, independent of the age level.

Integrating $m(a, s)$ over a gives us the density $\pi(s|x)$. Moreover, to compute the probability density of firm size, we need to integrate $\pi(s|x)$ over s to find the corresponding normalizing constants.

The resulted density is provided in the following Proposition:

Proposition 3.2. *$\forall x > b$ and $a > 0$, the firm size density given the initial size x, is the weighted average of the density $\pi(s|x)$ and can be determined by integrating $m(a, s|x)$ over all age a. In particular:*

$$\begin{aligned} \pi(s|x) &= [\tfrac{\min(\exp[(\alpha+\alpha_*)(s-b)],\exp[(\alpha+\alpha_*)(x-b)])-1}{\exp[\alpha(s-b)]}] \times [\tfrac{\alpha+\alpha_*}{\alpha\alpha_*} \cdot (\exp(\alpha_*(x-b)) - 1)]^{-1} \\ &= (\tfrac{\exp[\alpha_*(x-b)-1]}{\alpha_*} \tfrac{\exp[\alpha(s-b)]}{\alpha})^{-1} \min[\tfrac{\exp[[\alpha+\alpha_*](s-b)]-1}{\alpha+\alpha_*}, \tfrac{\exp[(\alpha+\alpha_*)(x-b)]-1}{\alpha+\alpha_*}]. \end{aligned} \tag{28}$$

where $\alpha = \frac{-\mu}{\sigma^2} + \sqrt{(\frac{\mu}{\sigma^2})^2 + \frac{\eta}{\sigma^2/2}}$ and $\alpha_* = \frac{\mu}{\sigma^2} + \sqrt{(\frac{\mu}{\sigma^2})^2 + \frac{\eta}{\sigma^2/2}}$.

Proof. Detailed calculation can be found in the Appendix. □

For any $\alpha > 0$ and $\alpha_* \geq 0$, the above firm size density, given initial size x, is a well-defined density.

Moreover, because we have:

$$\pi(s|x) \propto \exp[-\alpha(s-b)][\min(\exp[(\alpha+\alpha_*)(s-b)], \exp[(\alpha+\alpha_*)(x-b)]) - 1], \tag{29}$$

the size distribution follows a Pareto density with the tail probability of the form $\exp\left[-\alpha(s-b)\right]$. In other words, given that $s \geq x$, the density is just approximated by a Pareto density with tail probabilities of the form $\exp[-\alpha(s-x)]$ and α becomes the tail index of the conditional size distribution $\pi(s|x)$. Moreover, as long as new entrants enter with the same level of productivity, the initial size distribution G will be just a point mass at x and $\pi(s|x)$ represents the firm size density.

In addition, in an extreme case, if population growth goes to zero, the limiting distribution can be obtained by letting x go to b. Then, the profitability process of firms becomes a Brownian motion with a negative drift and a reflecting barrier at b and the resulting distribution of firm size is a Pareto distribution with mean $\exp(b).\frac{\alpha}{\alpha-1}$, given $\exp(s) \geq \exp(b)$.

3.4 Equilibrium and the Balanced Growth Path

3.4.1 The Balanced Growth Path

Definition 1. *On the Balanced Growth Path, per capita consumption (C), per firm revenue (V) and real wages (w), grow at a constant rate of $\eta + \kappa$. Firms enter and exit at constant aggregate rates in such a way that the aggregate measure of firms expands at a constant rate of η. The distribution of size shifts to the right with constant steps and its shape is time invariant. The interest rate (r), firm profit (π), and labor distribution ($m(s)$) and other characteristics of the firm's cross-sectional distribution are constant.*

Proposition 3.3. *Under Assumptions 1 through 5, a balanced growth path exists for the economy.*

Proof. From the setting of the model, we have the following:

- $P_t = P$;
- $N_t = H \exp(\eta t)$;
- $r_t = \rho + \gamma\kappa$;
- $C_t = w_t = w \exp(\kappa t)$, where $\kappa = \theta_E + \frac{1-\beta}{\beta}\eta$;
- $Z_{t,0} = Z \exp(\theta_E t)$;
- $R_{t,a} = C_{t+a}^{1-\beta}(Z_{t,a}L_{t,a})$ also grows at rate of κ.

Given the value of r and κ, we can find the value of the firm through Eq. (21). This, together with the zero profit condition determines $S(Z)$[14]. Thus, the stationary FSD of the economy is described as in the previous section.

The next step is to determine I, C and w.

Assume that the amount of labor assigned to set up a new firm, the amount of labor assigned to cover the fixed cost of an incumbent and the amount of labor assigned to production are, respectively, $L_E \exp \eta t, L_F \exp \eta t, L \exp \eta t$.

Depending on the profit maximization condition as specified in Eq. (11) - (13), after simplifying, we have:

$$\begin{bmatrix} L_E \\ L_F \\ L \end{bmatrix} = \begin{bmatrix} \lambda_E \\ \lambda_{F(s)} \int_b^\infty m(s)ds \\ \frac{I}{1-\beta} \int_b^\infty [\beta \exp(s) - \lambda_F(s)(\beta - 1)]m(s)ds \end{bmatrix} I,^{15} \tag{30}$$

The detailed calculation can be found in the Appendix.

Noting that from the labor market clearing condition, we have:

$$H = L_E + L_F + L.$$

Then, we can determine I, which is the rate that firms attempt to enter the market.

Second, from the firm profit maximization condition, specified in (11) - (13), we have

$$\frac{R_{t,a}}{w_{t+a}} = L_{t,a} + \exp(s_a). \tag{31}$$

Since the total aggregate output Y equals the sum of firm revenues, the aggregate output,

[14]The firm's productivity level is related to size through $S(Z)$

[15]The detailed calculation can be found in the Appendix.

$Y. \exp{(\kappa + eta)t}$ can be calculated as:

$$\begin{aligned} \frac{Y}{w} &= \frac{I}{1-\beta}\int_b^\infty [\beta \exp(s) - \lambda_F(s)(\beta-1)]m(s)ds + I\int_b^\infty [\exp(s)m(s)ds. \\ &= I.[\frac{1}{1-\beta}\int_b^\infty \exp(s)m(s)d(s) - (\beta-1)\int_b^\infty \lambda_F(s)m(s)ds]. \end{aligned} \tag{32}$$

This, together with the goods market clearing condition ($C = Y$) enables us to find the ratio $\frac{C}{w}$ and thus C and w[16].

□

3.4.2 Competitive Equilibrium

Definition 2. *The competitive market equilibrium for the economy is trajectory of a set of prices* $\{p_i\}$*, a set of decision rules* $\{C, L_i, E, b\}$ *and the distribution of firms* $\{m(s)\}$*, where* $L_i = L_E, L_F, L$ *and* E *represent the decision of a firm to be active in the market* $(E = 1)$ *or to exit* $(E = 0)$*), such that:*

- *The representative consumer choose the sequences* $\{C_t\}$ *of the composite goods to maximize his or her utility given the inter-temporal budget constraint, as shown in Eq.* (6)*;*
- *All active firms, at every age level, choose price* p *and labor* L *to maximize their profits in equation (11), (12) and (13), taking the price index as given;*
- *Firms at age* a *make the decision* E *given the evolution of productivity following the Brownian Motion in Eq.* (8)*. In particular, firms enter the market if the productivity level of the new entrants is such that the initial value of firms is greater than* b*, given that the free-entry condition is satisfied in Eq.* (22) *and exit once the firms' values fall below* b*;*
- *The stationary distribution of firms evolves according to equation (28) given* b *and* I*.*
- *The goods market clears as consumption equals total income;*
- *The labor market clears as* $H = L_E + L_F + L$*.*

[16] $C/w/w^{\beta/1-\beta}$ can be determined using the zero profit condition.

3.5 Further discussion on Fixed cost

In the preceding section, we developed and presented a stationary distribution that follows a Pareto distribution. The underlying stochastic structure is such that the probability distribution of individual firm size can be interpreted as the cross-sectional size distribution for the whole continuum of firms. The derived distribution can be considered the cross-sectional distribution of firm size in the economy. Such a distribution is developed with size measured by $\exp(s)$. In other words, $\exp(s)$ is the size of any firms with productivity $Z\exp(\theta_E t)$ at time t given that the initial size is greater than exit barrier b. As long as firm sizes of incumbents are larger than the initial sizes, and as long as fixed cost does not change across firm size groups, the firm size density exhibits a Pareto density. Moreover, if firms with much larger sizes than the exit barrier b exist, then there will be fewer small firms than anticipated by Pareto distribution.

As mentioned in the previous chapter, the business environment is significantly different between developed and developing economies. In a transitional, developing country, we expect to see regulatory burdens and corruption costs differ across different firm size groups. In particular, medium-sized firms, in many cases, bear the largest burden compared to their small and large peers. To capture such characteristics in the model, we modified the firm fixed cost (λ_F) to be a function of firm size; thus, fixed cost in this model captures the burden of corruption as well as regulatory cost. Readers should note that even though the corruption tends to be different across size groups, firms with similar sizes tend to have similar corruption costs. In particular, the costs only tend to change dramatically once firms develop beyond certain sizes (i.e, firms becoming large enough so they cannot be operating in informal sector). Therefore, it is plausible to assume corruption as a part of fixed costs of firms.

In the literature, corruption and regulatory burdens have been modeled as either a fixed or a variable cost. In this paper, we assumed these cost are a part of fixed costs to firms.

First, contrast to variable cost, the fixed corruption cost represents the contractual relationship between the private interest and the government agent (Talvitie, 2012) in which the exchange of bribes and favors occurs repeatedly over times. This is in line with the characteristics of corruption in Asia[17]. Second, the regulatory burden is normally considered compulsory for

[17] In Asia, a stable government and informal institutions (i.e. informal norms, rules or customs of the society) facilitate the long-term relationship between enterprises and governmental officials. Both agents in the corruption transaction tend to cooperate to maximize their long-term benefits and refrain from short-term opportunism. It

firms to operate in the market. These regulations include not only official taxes and fees but also other direct expenses on compliance, including paperwork, or employee training as well as indirect costs such as the value of time diverted to regulatory compliance by the owner, manager, or employee. Moreover, as described in the previous chapter, the bribe values are not the same across different firm size groups. In particular, the incidence of paying bribes is much higher for formal or larger-sized firms. Therefore, it is plausible to assume that the fixed cost that each incumbent has to bear is a function of firm size. In particular, medium-sized firms must pay a larger fixed cost, compared to small and large-sized firms.

Referring to the description of firms in previous section, from Eq. (21) we can see that the value of the firm and thus the exit barrier b are dependent on the level of the fixed cost. As the fixed cost increases, the value of the firm $V(s)$ and the exit barrier (b) decreases. Thus, there exists a non-linear relationship between firm growth and the initial firm size. This implies the rejection of Gibrat's law where, the growth rate of a given firm is independent of its size at the beginning of the period examined.

The increase in exit due to higher fixed cost also leads to a decrease in the size density $\pi(s|x)$ for medium-sized firm groups. In other word, the increase in fixed cost implies a smaller number of firms in the medium sized firm group, in accordance with empirical fact in many developing countries reported on in the literature. Moreover, because firms are identified by differentiated goods that they produce, a decrease in the number of firms also implies a fall in the number of differentiated goods, holding other variables constant.

As a result, in a developing country, as fixed cost differs across firm sizes with medium-sized firms paying larger amounts, compared to small- and large-sized firms, we may expect the more frequent appearance of small and large-sized firms in the firm size distribution, compared to medium-sized firms . Moreover, as fixed costs or the costs for corruption and regulations grow, a concentration of small-sized firms in the distribution also grows, leaving a smaller proportion of the medium-sized group.

is not uncommon in Asia that firms develop the relationship with civil servants by paying the periodic bribes to maintain the long-term relationship. This is very similar to the repeated prisoners' dilemma where both parties agree to cooperate after repeated iterations.

3.6 Concluding Remarks

Empirical practice has shown that the FSDs in developing countries deviate from the patterns observed in developed economies. In this chapter, a simple modification of the model developed by Luttmer (2007) enabled us to explain such a diversion of the FSD in a developing country. In the model, we considered corruption to be the fixed cost that each incumbent firm must pay to continue operating in the market. In accordance with empirical fact, the fixed cost in our model was assumed to be a function of size and was higher for medium-sized firms. As a result, the FSD of the economy deviates from the Pareto distribution as evidenced by the concentration of small firms, normally observed in developing economies.

This study has addressed a limited area of research within the domains of corruption, firm size, and the distribution of firms. However, we note that the evolution of FSD in different industries is presumably more complex than characterized by our study. The literature has documented numerous empirical evidence on the complexity of forces that can affect the FSD. These forces include, but not limit to, the growth of demand, cost and technology shocks, regulatory framework and financial burdens (See Mazzucato,1999) for a survey of the forces of growth in certain stages of the industry life cycle). However, analysis may provide a theoretical rationale and some empirical evidence to highlight some underlying mechanisms that possibly account for the typical distribution of firm size in developing countries and help to explain the fluctuation of firm size in the long run. Providing a theoretical model to generate this phenomenon would be a promising future research project. In addition, because in the presented model, the decision of bribery was assumed to be exogenous to firms, endogenizing the bribery decision of firms in the interaction with governmental officials and political elites is also an another possible future study.

CHAPTER 4

THE MISSING MIDDLE - PRODUCTIVITY AND CORRUPTION: AN EMPIRICAL ANALYSIS

4.1 Introduction

In the preceding chapter, we developed and presented a theoretical model that captured the impact of corruption on firm growth in a developing economy. We showed that corruption and other regulatory burdens result in an increase in fixed cost for middle-sized firms, leading to the decrease in the size density for middle-sized group. This model explains the missing middle in the FSD of developing countries.

In this chapter, we employed a more empirical perspective to investigate the phenomenon in the context of a particular developing country - Vietnam. Vietnam is a relevant case because it is a transitional, developing country. Over the last two decades, since the emergence of the "Doi moi" (Renovation) process, the country has moved from bureaucratic centralized management based on state subsidies toward a multi-stakeholder, market-oriented economy. The country has experienced a period of continuous high economic growth. However, Vietnam has yet to become a tiger in Asia as expected. Investigating the firm size distribution of Vietnam could shed light on the underlying mechanism of economic growth as well as on factors that cause the mis-allocation of resources in the economy.

In accordance with the theoretical framework, in this study, we focused on two factors: corruption and productivity. One may wonder why we studied these two aspects, because they may seem unrelated. The reason for our choice was that these two aspects might represent the distortions driving the mis-allocation of resources of firms, which may cause the missing middle phenomenon, if present in the economy.

In this paper, we argue that small firms may be reluctant to become formal to avoid a rise in expected corruption costs. They can then substitute capital for labor to scale-up production and wait for a productivity shock that will offset the cost of growing. Alternatively, a possible differential in technology or in productivity among firm-size groups may reconcile the patterns observed in the data. We wondered about the existence of a relationship between firm sizes and production efficiency, which could be affected by the management of production in firms. We anticipated that if firm size increased, management costs also increased, and that perhaps, this relationship was not linear, because the management of a production process would involve organizing various hierarchies and networks. Our analysis later in this chapter shows, middle-sized firms tend to be less efficient in production. Our result indicates that middle-sized firms may face some misalignment, or some management difficulty.

Business environments vary across nations. Although it is difficult to measure how good the business environment is or how well each firm is managed, production efficiency can be evidence of the quality of an organization's management efforts. We provide empirical evidence that available resources might be less efficiently used in developing nations, and this could be an important factor contributing to the missing middle phenomenon.

Indeed, in our empirical analysis, we do not attempt to establish a link between corruption and productivity; rather, we simply present evidence of potential productivity growth for small-sized firms in Vietnam and one possible explanation of why they *do not grow*. For this purpose, we used two data sets from Vietnam. The first data set that we used to measure productivity and efficiency was the Enterprise Census conducted by the General Statistics Office of Vietnam (GSO). The Enterprise Census contains information on all registered firms in Vietnam. Our data included capital, labor, and intermediate materials as inputs, and goods measured in monetary terms as the output. Table 6 lists the sectors we study in this paper.[1]. The second data set, the Small and Medium Enterprises (SME) survey, which we used to analyze the relationship between bribes and firm size, and obtain comprehensive information on all forms of small-sized

[1]These sector codes are based on the International Standard Industrial Classification (ISIC) codes. We chose these sectors because each sector based on the two-digit-level ISIC code has more than 200 observations per year; thus, we could ensure robustness of the nonparametric estimations.

and medium-sized firms[2].

First, we investigated whether any evidence exists of the missing middle phenomenon in Vietnam. As mentioned previously, the missing middle has been acknowledged and discussed in many papers since Liedholm and Mead (1987). Recently, Hsieh and Olken (2014) studied data from India, Indonesia and Mexico and found (a) there are many small firms but not so many middle-sized or large-sized firms, (b) large firms rather than small firms seemingly incur larger fixed costs, and (c) there is no evidence of discontinuities in regulatory obstacles across different firm sizes.

In response to the Hsieh and Olken (2014)'s findings, Tybout (2014a) (see also Tybout, 2014b) suggested that a better test of the missing middle was to ask whether the share of middle-sized firms, compared to the shares of small- or large-sized firms, was smaller than the share that one would observe in an undistorted economy. One sign would be a concentration of small firms.[3] In fact, we observed a concentration of small firms in Vietnamese data. In addition, we were interested in exploring the underlying mechanisms for the observed firm size distribution. Our data showed features consistent with the findings in Hsieh and Olken (2014) on three points as mentioned previously, and our tests did not reject bimodality in the case of Vietnam for this particular period.

Next, we studied production efficiency in each sector of the manufacturing industry by using the frontier approach within a nonparametric model. To the best of our knowledge, we do not know of any other studies that link the missing middle and production efficiencies across different firm sizes in various sectors of a developing country. The analysis of the underlying mechanisms for the observed firm size distribution has been less developed, perhaps because of the unavailability of detailed data.[4]

[2]The survey was published in "Characteristics of doing business in Vietnam" conducted by the Central Institute for Economic Management (CIEM) in collaboration with the Vietnamese Institute of Labor Science and Social Affairs (ILSSA), the Department of Economics (DoE) of the University of Copenhagen, and UNU-WIDER together with the Royal Embassy of Denmark, every two years since 2005. For details, see Rand and Tarp (2012).

[3]Furthermore, Tybout (2014a) indicated that it might be problematic to equate the missing middle with bimodality.

[4]Recently, there has been growing interest in the relationship between the firm size distribution and productivity in developed nations. Leung et al. (2008) investigated the US–Canada difference in the distribution of employment over firm size and confirmed that a larger average size supported higher productivity at both the plant and firm level. Crosato et al. (2009) investigated the data of a micro-survey in Italy and, using nonparametric estimates, found that firms in the concave part of the Zipf plot of the Italian firm distribution overwhelmingly experienced increasing returns to scale, while firms in the linear part were mainly characterized by constant returns to scale. Diaz and Sanchez (2008) use a stochastic frontier model to investigate small-sized and medium-sized

In recent years, researchers have developed many methods to estimate production efficiency under the frontier approach.[5] Comparing average products across different inputs is a complicated task. For example, Ray (2004) noted that measures of a firm's productivity that rely on a single input, disregarding other inputs, may fail to reflect total factor productivity. Considering this, we used the frontier approach in our analysis.

Our analysis shows that in most of the manufacturing industries that we studied, middle-sized firms' production efficiencies tend to be lower than those of small-sized or large-sized firms. Further, we showed that the level of minimum efficiency and the scale at which it occurs vary across industries.

Further, our work indicates that the large-sized firms may be unable to fully utilize their inputs. Hsieh and Olken (2014) claimed that the problem of economic development in low-income countries is how to relieve the differential constraints faced by large firms. Our empirical analysis supports this view in concrete terms. Our analysis on inefficiency associated with scale showed that the large-sized firm group includes most firms with lower average product and larger production scales, compared to firms with the highest average product in their categories. Our analysis indicated that one of the problems, in Vietnam, at least, is how to activate some resources that these large-sized firms hold but tend to utilize incompletely in production.[6]

Because our second method was developed for measuring returns to scale, our work relates to the literature regarding measuring returns to scale. Basu and Fernald (1997) used U.S. data to estimate returns to scale and found that a typical industry appeared to have constant or slightly decreasing returns to scale. Diewert et al. (2011) found a similar result using Japanese data for the period of 1964 to 1988, provided explanations of the time variation in returns to scale in light of economic occurrences and government policies at each time period. Using the French data from the nineteenth century, Doraszelski (2004) showed that returns to scale vary across industries and time. These researchers estimated returns to scale for the production function of the entire industry. Using the data of large U.S. banks in the late 1980s, Miller and Noulas (1996) conducted a data envelopment analysis and showed that larger and more profitable banks

manufacturing enterprises in Spain and found that these firms tend to be less efficient than their large-sized peers.

[5]The production frontier can be estimated using parametric estimate (deterministic or stochastic frontier analysis) or non-parametric one (data envelopment analysis or free disposal hull). For more details, see Kalirajan and Shand (1994).

[6]For example, in Vietnam, Sector 17 (Textiles) and Sector 21 (Paper and paper products) include many large-sized firms, previously or currently state-owned. As shown later in Figure 7.6, these industries exhibit this feature more visually.

had higher levels of technical efficiency, however, larger banks tended to be too focused on cost considerations. Recently, Feng and Zhang (2014) investigated the returns to scale of large banks in the United States over the period 1997 to 2010 by considering technological heterogeneity and showed that most firms' production levels exhibit constant returns to scale. Johannes (2005) provided evidence on the differences in the evolution of firm size and productivity distribution across nine sub-Saharan African countries and developed countries including the United States. Banerjee and Duflo (2005) provided a survey and analyzed the causes of lower total factor productivity (TFP), such as access to technology or human capital, in poor countries (which they call "macro puzzles").

The discussion in this chapter adds to the literature on firm size and productivity in developing countries. Using Vietnamese data that include small-sized firms, we show that most manufacturing industrial sectors displayed similar features and thus, we demonstrate how the result differs across different sectors. Specially, we show that productivity is different across different firm sizes and middle-sized firms have lower efficiencies of production.

We found that the level of minimum efficiency and the scale at which minimum efficiency occurs varied across industries, which indicated that an analysis of each industry was needed to study underlying mechanisms for the observed firm size distribution. In addition, interestingly, our data analysis of production costs and outputs indicated that middle-sized firms tend to have lower production efficiency even before paying bribes or other fees.

In addition, we conducted a logit regression to study the relationship between firm sizes and the likelihood of paying bribes. We used the Small and Medium Enterprises (SME) survey, which provides comprehensive information on all forms of small-sized and medium-sized firms[7] and analyzed the important factors involved in firms' bribing activities, particularly in terms of the relationship between firm size and the likelihood of paying bribes, by using a logit regression.

Our results regarding corruption indicated that as firm size increases, the likelihood of paying bribes also increases. The formality of a firm matters in the sense that the likelihood of paying bribes is higher for formal firms, which tend to be larger. We also analyzed the relationship between the likelihood of paying bribes and firm age. Our analysis showed that firms' ownership types and locations are important determinants of bribing activities. Interestingly, our analysis

[7] The survey was published in "Characteristics of doing business in Vietnam" conducted by the Central Institute for Economic Management (CIEM) in collaboration with the Vietnamese Institute of Labor Science and Social Affairs (ILSSA), the Department of Economics (DoE) of the University of Copenhagen, and UNU-WIDER together with the Royal Embassy of Denmark, every two years since 2005. For details, see Rand and Tarp (2012).

showed that firms often seem to pay bribes at the start of business; however, those that do are increasingly likely to pay bribes over time.

In the literature, some studies (e.g., Dutta et al., 2013, Mishra and Ray, 2013) have indicatedthat small firms may refrain from becoming formal to hide themselves from the rent-seeking activities of civil servants[8]. Our findings about informality and the likelihood of bribery are consistent with the findings of previous studies.

Although it is believed that bribing activities are common, the dynamics of bribing activities and firms' growth have not yet been well understood. Using data from Greece, Athanasouli et al. (2012) demonstrated that firm engagement in corruption is heterogeneous, and small and medium firms display a higher engagement in corrupt practices. Using enterprise data for the economies of Central and Eastern Europe and Central Asia, De et al. (2010) showed that bribes appear to have a negative impact on firm-level productivity. Using firm-level data from Turkey, Ayaydin and Hayaloglu (2014) studied the influence of corruption on firm growth. Wu (2009) conducted a regression analysis with World Business Environment survey, and found that Asian firms were more likely to bribe when faced with fierce market competition or corrupt court systems. Bai et al. (2013) used cross-industry heterogeneity in growth rates and a data set covering formal companies in Vietnam to test empirically whether growth leads to lower corruption and found that it does.

This chapter contributes to the literature in several aspects. First, we used a unique firm-level data set, consisting of both informal and formal companies to examine the characteristics of firms that are important determinants of bribing in a developing nation. Second, from our analysis, we provide evidence that firm size, formality, and firms' ownership type are important factors affecting the likelihood of bribes, while firm age negatively affects the likelihood of bribes. One may question which factor occurs first (i.e., whether corruption determines if a firm decides to grow to a different size or whether firm size leads to corruption). However, our productivity analysis, showing that small-sized firms produce efficiently compared to other size-groups, suggests that corruption could be one of the reasons for firms to decide *not to grow*.

The rest of the chapter is organized as follows. In the second section, we describe our first data set and provide a summary of Vietnam's firm size distribution and average products. The

[8] Using data from the World Business Environment Survey compiled by the World Bank for a large number of developing and developed countries, Dabla-Norris et al. (2008) studied the link between firm size and informality. Further, see Dabla-Norris et al. (2008), Dutta et al. (2013), Mishra and Ray (2013).

third section involves our methodology for measuring productivity and efficiency, and presents the analysis on production efficiency, and inefficiency associated with scale. The third section also provides the empirical analysis on bribes and firm size. The last section concludes the discussion.

4.2 Firm Size Distribution at a Glance in Vietnam

4.2.1 Census Data Description

In Vietnam, the Enterprise Census has been conducted annually since 2000. In the census, an enterprise is defined as "an economic unit that independently keeps a business account and acquires its own legal status." In this paper, we focused on the nine-year period from 2000 to 2008. We excluded recent years (from 2009 to present) because the Vietnamese economy has suffered from the effects of the global financial crisis, and because a 30% reduction in the corporate income tax rate for qualifying entities was implemented in the fourth quarter of 2008 and all of 2009[9]. In addition, small-sized and middle-sized firms involved in labor-intensive production and processing activities may have benefited from the subsequent tax reduction under Decree 60/2011. We also excluded inconsistent data from our sample, such as observations that were recorded twice for the same firm in the same year, those with negative or zero revenue values, or those with an implausibly large number of employees. We included 14 sectors of the manufacturing industry that had at least 200 observations for each year, and the number of observations ranged from approximately 3,350 to 29,700 per industry.

Table 6 lists the sectors we studied.[10] Table 8 provides the summary of the number of firms for each year in each selected manufacturing sector.

Insert Table 6 Insert Table 8

[9]See Circular No. 03/2009/TT-BTC published in January 2009 by Vietnam's Ministry of Finance.

[10]These sector codes are based on the International Standard Industrial Classification (ISIC) codes. We chose these sectors because each sector based on the two-digit level ISIC code had more than 200 observations per year, so that we can ensure robustness of the nonparametric estimations.

4.2.2 Summary

To convey the features of our data, we start with a discussion of the figures for the whole manufacturing industry. In the next sections, we present an analysis of each sector.[11] For the whole sample, the number of observations ranged from 9,143 (year 2000) to 27,918 (year 2008). The number of sectors was 27.

Figure 7.1, Figure 7.2 and Figure 7.3 summarize the firm-size distribution for 2000, 2004 and 2008. Figure 7.1 shows the distributions of firm size measured by number of employees in each firm. Figure 7.2 shows the distributions of employment share by firm size; specifically, showing the percentage of firms having the number of employees contained in each value of the x-axis relative to the total number of firms. Figure 7.3 shows the distribution ratios relative to capital and labor, as well as the distributions of revenue ratios. To save space, we do not present the figures for the other years, which are quite similar to the ones presented here.[12]

Insert Figure 7.1, Figure 7.2 and Figure 7.3

To make the patterns more visible, in Figure 7.1 and Figure 7.2, we classify firm size into subintervals of 0 to 200 employees (column 1), 10 to 200 employees (column 2), 20 to 200 employees (column 3), 50 to 200 employees (column 4) and 200 to 3000 employees (column 5). These two sets of histograms present the distribution of firm size in bins of 10 workers. In Figure 7.3, we take the maximums and the minimums of $\log(\frac{\text{Value added}}{\text{Capital}})$ (the first row), $\log(\frac{\text{Value added}}{\text{Capital}})$ (the second row) and $\log(\frac{\text{Revenue}}{\text{Material}})$ (the third row), for the entire period of 2000 to 2008, and truncate the intervals into 50 bins.

The key observations in the figures include the following:

- From Figure 7.1, there are a large number of small firms compared to the number of other sizes.

[11]Figure 7.1, Figure 7.2 and Figure 7.3 are similar with those in Hsieh and Olken (2014) for India, Indonesia and Mexico.

[12]Figures of Average Product and Firm Size (similar to Figure 3 in Hsieh and Olken (2014)) are also available upon request. In these figures, because we had many observations, we constructed the figures for each sector. In some sectors, our figures show similar features with the ones shown in Hsieh and Olken (2014), although in particular sectors such as Sectors 17 and 18, the relationship between $\log(\frac{valueadded}{worker})$ and size was U-shaped. In this chapter, to study the relationship between sizes and productivity, we assessed the the efficiency scores and inefficiency associated with scale for each sector of the manufacturing industry, instead of studying average products of the whole manufacturing industry.

- From Figure 7.2, the distribution of employment in each column is "flatter" compared with the distribution shown in Figure 7.1.
- From Figure 7.2, the employment share for very small firms increased over time (from 2000 to 2004 and from 2004 to 2008).
- From Figure 7.3, the distributions of average products do not look bimodal.

Next, we present an empirical test to give detailed analyses of the firm size distribution. First, We used the Dip test to investigate the null hypothesis H_0 that the sample distribution has a unimodal density, against the alternative hypothesis H_1 that the sample distribution has more than one mode.(for details, see Hartigan and Hartigan, 1985). According to the results of the Dip test[13], the null hypothesis is rejected at the 5% significance level for all years. The Dip statistics are all smaller than 0.05, which indicates the presence of bimodality or multimodality.[14]

Then, we conducted the bimodality coefficient test[15]. The test is designed to examine the relationship between bimodality, skewness and kurtosis of the distribution. The bimodality coefficient (BC) statistics ranged from 0 to 1, with those exceeding 0.55 suggesting bimodality[16]. The BC's of all years are shown in Table 1. As shown, all BC statistics were higher than 0.55, which indicates the existence of bimodality.

Table 1: The Bimodality Coefficients

Year	'00	'01	'02	'03	'04	'05	'06	'07	'08
BC*	0.69	0.72	0.71	0.71	0.63	0.69	0.71	0.62	0.56

* - BC stands for the Bimodality Coefficients.

[13] The detailed result is presented in Table 11 of Appendix.

[14] A more detailed interpretation of this statistic is found in Freeman and Dale (2013). Finding out how many modes were present in the distributions was not our aim in this paper, and so we did not consider this issue further.

[15] The logic behind the bimodality coefficient is that a bimodal distribution would have very low kurtosis, an asymmetric character, or both, all of which increase this coefficient. The formula itself does not assume a particular distribution. On the other hand, the value for the uniform distribution or the exponential distribution is $5/9$. Values greater than $5/9$ may indicate a bimodal or multimodal distribution in the data.

[16] For more details, see page 1258 in SAS Institute Inc. (2008). The bimodality coefficient is $\frac{m_0^2+1}{m_1+\frac{3(n-1)^2}{(n-2)(n-3)}}$ where m_0 is skewness and m_1 is the excess kurtosis. The BC test, however, has been criticized because of its sensitivity to the skewness of the distribution. Thus, we also included the results of the Dip test, which has been considered to be more robust (see Freeman and Dale, 2013).

Further, we computed the BC's for each sector, as shown in Table 2. Interestingly, the BC's in Sector 22, Sector 26, and Sector 29 indicate that these sectors' firm size distributions are not bimodal, unlike in other sectors. Sector 22 (Publishing, printing, etc.) includes many small firms (i.e., many small printing stores) and fewer large firms. In contrast, the other two sectors, Sector 26 (Nonmetallic mineral products) and Sector 29 (Machinery and equipment) contain large firms with a need for large capital. In either case, firm size distribution tends to be non-bimodal for these sectors.

Table 2: The Bimodality Coefficients in each Sector

Sector	14	15	17	18	19	20	21	22	24	25	26	28	29	36
BC*	0.56	0.68	0.74	0.64	0.68	0.69	0.62	0.49	0.72	0.71	0.36	0.65	0.54	0.69

* - BC stands for the Bimodality Coefficients.

Overall, our result for the whole sample does not reject the hypothesis H_1: The result shows a bimodal distribution; thus, we reject the null hypothesis that the distribution is uni-modality. [17] Moreover, the result implies that the features of the firm size distributions vary substantially across sectors. As our sectoral analysis of the BC indicates, there are some sectors for which the bimodality is rejected[18]. In the next section, we discuss productivity and efficiency across different sizes. Finally, Table 2 shows that the firm size distribution varies substantially across sectors; hence, we conducted analysis of productivity and efficiency for each sector. Our results in the next section indicate that the sectors rejecting bimodality in Table 2, particularly Sector 26 and Sector 29 show different features in productivity and efficiency compared to other sectors.

[17] We also checked the distributions against Zipf's law and the hypothesis that they follow Zipf's law was rejected in our data set. The details of analysis are available upon request.

[18] We also computed the BCs for each sector by year. Because the number of observations in each year for each sector decreased, the results fluctuated. In the yearly results, the BCs tended to increase, and even for sectors 14, 26, and 29, the BCs were higher than 0.56 (except for sector 14 in 2008). Perhaps the interesting case was sector 26. Although it showed the lowest score for any year, sector 26's score for each year was higher than 0.56. We further analyzed sector 26's histograms of firm size distribution for each year and found that small firms grow noticeably in this sector year by year. According to Anh et al. (2014), the industry of non-metallic mineral products was one of the "sunrise" industries in Vietnam. This industry's ranking in the share of total output has increased from the eighth (in 2001) to the second in 2007, and continued to be the second highest after Food products.

4.2.3 Productivity and efficiency analysis

This section consists of two parts. In the first part, we discuss the measurement of the efficiency of production, using the efficiency score[19]. In the second part, we describe our analysis of the inefficiency associated with scale at the firm level using nonparametric methods, which imposed few if any assumptions on the data. In contrast to Hsieh and Olken (2014)'s investigation of the average product of inputs of the economy, we examined the production efficiency of firms at a sectoral level.

We considered the situation in which a firm uses capital, labor, and intermediate materials to produce goods. All firm information came from the census data of Vietnamese manufacturing industries. Real revenue was used as a proxy for output.[20]. Three inputs were included in our estimation: intermediate inputs, labor, and capital. Labor was measured by the total income of employees in a firm. This measure included total wages and other employee labor-related costs such as social security, insurance and other benefits[21]. The intermediate material variable includes costs such as fuel and the value of other materials. Capital was measured as assets to be used in production. All variable values were adjusted to account for inflation to obtain a real value.

In the following analysis, we construct a production function for each sector without imposing any restrictions on the parametric relationship between inputs and outputs.

Our analysis relies on the following theoretical framework. Consider one sector in the manufacturing industry. Firm i's input and output combination is called a *production unit*, denoted by (x_i, y_i) with $x_i \in R^3_+$ and $y_i \in R_+$. Then, we define firm i's production set by

[19]To compute the efficiency scores, we use the R-package *FEAR* (see Wilson, 2008). We are grateful to Wilson, who provided us with a license to use the software.

[20]At a firm level, prices and quantities may not be well measured, and revenue, instead of gross output or cost, is normally used. In the literature, there have been some arguments that the elasticities of labor and capital, in a revenue estimate, may be downward biased (Basu and Fernald, 1997). Klette and Griliches (1996) reported that changes in sector prices are substantially diversified and correlated with changes in labor and capital. However, according to Jacques and Jordi (2005), "the introduction of individual output prices into the production function does not markedly affect the estimate...[and]... the estimation of a production function in terms of "physical quantities" is, in fact, meaningless, unless we confine the analysis to a very precisely defined industry where goods are so homogeneous that firm outputs can be well measured and compared across firms". Accordingly, we used this measure in the analysis.

[21]Alternatively, labor input can be captured by quantity measure such as number of employees or total hours worked. This quantity measure, however, implicitly assumes uniform skill distribution across firms. In this analysis, we used total income of a employee to account for differences in skill distribution across firms in the economy.

$\Psi_i = \{(x, y) : x \geq x_i, y \leq y_i\}$. The union of all the firm's production sets in the sector, $\bigcup_i \Psi_i^l$, is the production set of this sector. The surface of this set is the production function. Using these data, we construct a production function for each sector.[22] Next, we compute each firm's efficiency score, which is the minimal proportional reduction of all inputs while maintaining the same output level within $\bigcup_i \Psi_i^l$. Figure (4.1b) graphically illustrate in the case of a single input and a single output (an explanation for the case of multiple inputs is provided later). Each dot represents a productive unit. A solid line is the constructed production function. The multiple θ in Figure (4.1b) shows the efficiency score of this white circle whose inputs are given by x. This is our first method to compute the efficiency scores.

Next, by using the algorithm developed by Soleimani-Damaneh and Reshadi (2007), we analyzed returns to scale or inefficiencies associated with production scale.[23]

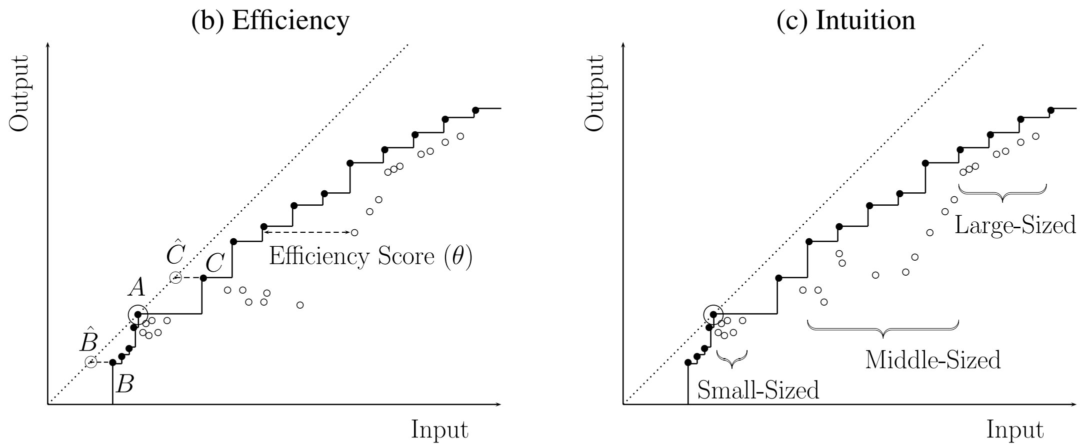
Regarding our method to estimate inefficiencies associated with scale, we note that the circled dot A has the highest average product. The other black dots on the production function, such as B and C, are efficient in terms of efficiency scores, unlike the white circles, but do not have the highest average product. If the same technology and management skills of Firm A are available for Firm C, and the underlying technology is constant returns to scale, Firm C could achieve Firm A's average product by adopting the production practices of Firm A, which is represented by point $\hat{C}$. Otherwise, this indicates that there is some inefficiency associated with production scales or decreasing returns to scale technology between A and C. Similar logic applies to points A and B, although in this case, the technology may be increasing returns to scale, but not decreasing returns to scale.

Conceptually, the above method is intended to assess the existence of some management inefficiency related to scales or decreasing returns to scale technology (or both), although the method does not identify what percentage of the difference between Firm B and C's and Firm

[22]To clarify, we did not impose any functional form on the estimation, and a production function was estimated as a frontier of a production set.

[23]Diewert et al. (2011) used the following definition of *returns to scale*. If all inputs are multiplied by a positive scaler, t, and the consequent output can be represented as $t^\gamma y$, the value of γ may be said to indicate the magnitude of returns to scale. If $\gamma = 1$, there are constant returns to scale; if $\gamma > 1$, there are increasing returns to scale and if $\gamma < 1$, there are decreasing returns to scale. In the literature on productivity analysis, a production function is constructed from the data, and we applied this definition to each firm's production by comparing it to the production of the firm with the highest average product among the set of firms with similar production scales. Applying this definition to each firm's production might be confusing, because in the classical production theory of economics, returns to scale is a property of the production technology. Thus, we only used the term returns to scale when we discussed production technology itself.

(a) Illustrative Example



A's production levels is attributable to management inefficiency and the returns to scale technology. For instance, suppose that there are two firms, where Firm 1 uses 10 units of labor to produce 5 units of goods and Firm 2 uses 15 units of labor to produce 7 units of goods. If the technology used is constant returns to scale, then by simply extrapolating Firm 1's production, Firm 2 should be able to produce 7.5 units by using 15 units of labor. Then, comparing Firm 2's production, the technology may be causing a decrease in returns to scale, or there may be some inefficiency in managing the resources in Firm 2. [24] Merging data from multiple sectors might mix these two features even further, and thus, we may not be able to identify a relationship between management inefficiency and firm size. It is particularly important to study this measure at the sectoral level, because the underlying technology in each sector would not be much different in comparison with different sectors.[25]

Figure (4.1c) shows the intuition of our empirical findings in the case of one input and one output, although our analysis with multiple inputs is more complicated. Large-sized firms produce closer to the production function than do middle-sized firms, although compared with the highest average product firms, they tend to have lower average products (except for the sectors

[24] As pointed out by Feng and Zhang (2014), not considering the technological heterogeneity of an individual firm even within a single sector can cause bias in measuring returns to scale. Attempting to separate the two factors by measuring returns to scale considering technological heterogeneity for each firm might be an interesting and promising research path to further investigate to what extent inefficiencies only associated with scale affect the missing middle.

[25] Another way to approach this problem is by assuming that the basic natures of the technologies are not so different between developed and developing nations, one may try to study this measure at the sectoral level in developed nations and compare the outcomes with ours.

with lower bimodality coefficients). In the case of multiple inputs, these lines can be thought of as a cone. In contrast, the middle-sized firms tend to be further away from efficient production and have lower average products.

The following summarizes our findings.

- In all the sectors, many firms have low efficiency scores, smaller than 0.5, across different firm sizes.
- Except in some sectors whose bimodality coefficients are low, the relationship between firm size and efficiency score tends to be U-shaped, which indicates that smaller-sized firms and larger-sized firms produce efficiently compared to middle-sized firms.
- Firm size at the bottom of the U-shape differs across sectors.
- The large-sized firms are likely to exhibit inefficiencies associated with scale.

In the following sections, we describe in detail our methodologies and findings.

4.2.4 Efficiency Score and Results

Suppose that there are L industrial sectors in the economy. Take a sector $l \in \{1, \cdots, L\}$ (we repeat the same procedure for each sector) where inputs $x \in R^3_+$ are used to produce an output $y \in R_+$. Let N^l denote the number of firms in sector l of the data set. Firm i's input and output combination is called a *productive unit*, denoted by (x_i, y_i) with $x_i \in R^3_+$ and $y_i \in R_+$.[26] As stated above, for firm i, we formally define

$$\Psi_i^l = \{(x, y) : x \geq x_i, y \leq y_i\}. \tag{1}$$

Further, we define $\Psi^l_{FDH} = \bigcup_i \Psi^l_i$, and the convex hull of Ψ^l_{FDH} is defined as Ψ^l_{DEA}. The efficiency score for a productive unit (x_i, y_i) is defined by

$$E(x_i, y_i) = \inf_{\theta}\{\theta : (\theta x_i, y_i) \in \Psi^l\}, \tag{2}$$

[26]To keep the notation simple, we did not explicitly denote a period and an industry for each productive unit.

for each $\Psi^l = \Psi^l_{FDH}$ under *free disposal hull* (hereafter, FDH), or Ψ^l_{DEA} under *data envelopment analysis* (hereafter DEA), respectively. Further, the PS Ψ^l is assumed to satisfy the regularity conditions; namely, boundedness, closedness, no free-lunch[27], and free disposability[28]. We say that Ψ^l exhibits *convexity* if for every (x_1, y_1), $(x_2, y_2) \in \Psi^l$, and any $\alpha \in [0, 1]$, $\alpha(x_1, y_1) + (1 - \alpha)(x_2, y_2) \in \Psi^l$ holds.

The efficiency score[29] lies between 0 and 1, and represents the minimal proportional reduction of all inputs while maintaining the same output level within the production set (hereafter, PS). Solving Problem 2 requires constructing the PS, using linear programming. In this analysis, we used two different methods, the FDH and DEA methods as defined above. The PS for the FDH method is different from the PS for the DEA method, and the difference between the two scores indicates the existence of nonconvexity in the production technologies.

Figure 7.4 is a scatter-plot of efficiency scores for the FDH method in our sample industries for the entire nine-year period, and the curve represents the smooth trend line of the efficiency scores for every 0.1 interval of $log(size)$.[30] The x-axis represents the logarithm of firm size. In our analysis, we also present the ratios of the two efficiency scores obtained from the DEA and the FDH methods in Figure 7.5. The ratio of the two scores indicates the existence of nonconvexity in the production technology.

Insert Figure 7.4, and Figure 7.5

According to Figure 7.4, except in sector 14 (Other mining and quarrying), sector26 (Nonmetallic mineral products), and sector 29 (Machinery and equipment, etc.), the relationship between firm size and efficiency score is clearly U-shaped, which indicates that smaller firms and larger firms produce efficiently compared to middle-sized firms. In contrast, Figure 7.4 for sectors 14, 26, and 29, displays the low BCs in Table 2, showing that the smooth trend lines of the efficiency scores are mostly decreasing as the size increases. In sector 29,a spike is shown in the trend line for very large-sized firms, although this seems to be caused by a few efficient large-sized firms. Moreover, firm size at the bottom of the U-shape—that is, the lowest efficiency score—differs

[27] A positive amount of production cannot occur without a positive amount of inputs.

[28] The increase in inputs must lead to increased or constant outputs, and a smaller output vector than a feasible vector is also feasible.

[29] This is called the “input-oriented” efficiency score. There is another definition called the “output-oriented” efficiency score. In this analysis, we used the input-oriented efficiency score because we have computed output-oriented scores for some sectors, which did not change substantially.

[30] We used Matlab’s command `smooth`. This inserts a smooth trend line by using the local regression method.

across sectors. The heterogeneity of firm size at the bottom of the U-shape across industries indicates the importance of sector-level analysis in investigating the missing middle.

Finally, the ratios between the DEA scores and the FDH scores are less than one for most of the firms, which indicates nonconvexity in the production technologies, and the shapes of the trend lines are quite different across sectors. This shows that the nonconvexity of the technologies is also significantly different across sectors.

4.2.5 Inefficiency Associated with Scale and Results

In this section, we discuss the relationship between production scale and average product at the firm level in each industry. Figure 7.5 in the previous section indicates the existence of some nonconvexities in production technologies.[31] Thus, we used an FDH estimate, which does not require the assumption of convex technology.

Similar to the process described in the previous section, we consider manufacturing sector l. Let D^l denote a set of all the observations in sector l. Then, we consider the following problem for each productive unit $(x_i, y_i) \in D^l$,

$$\begin{aligned} &\text{Minimize } \theta \\ &\text{subject to } \textstyle\sum_{j\in N^l} \lambda_j x_j \leq \theta x_i, \sum_{j\in N^l} \lambda_j y_j \geq y_i, \lambda_j = \delta w_j, \\ &\qquad w_j \in \{0,1\} \text{ for every } j \in N^l, \delta \geq 0, \textstyle\sum_{j\in N^l} w_j = 1. \end{aligned} \tag{3}$$

The aim was first to compute each firm's hypothetically best productive unit by projecting it onto the line of the highest average products. To illustrate, let us consider the inefficiency associated with scale C in Figure 4.1b. In Figure 4.1b, $\hat{B}$ and $\hat{C}$ correspond to the projected points of B and C, respectively. Note that here, only one λ_j is nonzero, but it does not equal 1. The algorithm finds some nonzero values of $(\lambda_j x_j, \lambda_j y_j)$, which are associated with $(\theta_i x_i, y_i)$ for each i. In Figure 4.1b, A has the highest average product. To obtain $(\lambda_C x_A, \lambda_C y_A) = (\theta_C x_C, y_C)$, which is represented by $\hat{C}$, it must be the case that $\lambda_C > 1$. This implies that if the same technology and management skills of Firm A are available for Firm C, and the underlying technology is

[31] Farrell (1957) emphasized indivisibilities and economies of scale as important sources of nonconvexity and emphasized that convexity can only be justified in terms of time divisibility (ignoring not only any setup times but also indivisibilities, increasing returns to scale, positive or negative production externalities, etc., that can lead to nonconvexity). For some recent theoretical or empirical studies, see Brown (1991) or (Ramey, 1991).

constant returns to scale, Firm C could achieve Firm A's production efficiency by adopting the production processes of Firm A. Otherwise, it indicates that there is some inefficiency associated with production scales or decreasing returns to scale technology between firms A and C. A similar argument applies to points A and B.

Now, let θ_i denote the solution θ associated with observation i, and let λ_i^+ and λ_i^- denote the maximal and the minimal of the solutions $\sum_{j \in N^l} \lambda_j$ to Problem 3 for each i, respectively.[32] When λ_i^+ is smaller than 1, it indicates that there is a firm with higher average products whose scale is larger than firm i.

Figure 7.6 presents the bar graphs indicating how many percentages of firms have $\lambda_i^+ < 1$ (dark color), $\lambda_i^- > 1$ (light color), and else (white) in each group of firm sizes. To facilitate understanding, we group the samples by number of employees into five groups, where groups 1 through 5 include firms with $1 - -10$, $11 - -50$, $51 - -100$, $101 - -200$, and 201 or more employees, respectively.[33] The white portions are the groups of firms with the highest average products; in Figure 4.1a they are shown as the circled dots. According to the categorization defined above, we calculated how many firms belonged to each group and then divided these numbers by the total number of firms in each group, so that we could obtain the percentages of firms in each category of each group. We repeated this procedure for all sectors.

Insert Figure 7.6

Figure 7.6 shows that in most of the sectors, the proportion of firms with the property $\lambda_i^+ < 1$ is highest in Group 1. This is understandable, because their size is small and thus their production perhaps tends to be small compared with the optimal scale. Interestingly, we observed that in many sectors, the large-sized firm category included the most firms with the property of $\lambda_i^- > 1$, particularly in sectors 22 (Publishing, printing, etc.) through sector 25 (Rubber and plastic products). This finding contrasted with the results seen in sector 14 (Other mining and quarrying), sector 26 (Nonmetallic mineral products), and sector29 (Machinery and equipment, etc.), which were the sectors with the low BCs in Table 2. The profiles of the production functions revealed by the analysis are typical of firms using large quantities of inputs because they show decreasing returns to scale, and many large-sized firms are located in this region. As

[32] A more detailed description of the numerical algorithm to calculate λ^+ and λ^- is found in the Appendix. In the productivity analysis literature, a firm with $\lambda_i^+ < 1$ is said to be *operating under Increasing Returns to Scale* and a firm with $\lambda_i^- > 1$ is said to be *operating under Decreasing Returns to Scale*.

[33] This set of cutoff sizes ensured that we had enough firms in each group.

Figure 7.5 shows, at larger scales, the production sets in most of the sectors are more convex. This implies that the production functions in the regions of larger quantities of inputs show concavity. From the analysis shown in Figure 7.4, we can see that many large-sized firms operate close to the function. This may indicate that even though the large-sized firms can produce in a relatively efficient manner, they may not be able to take full advantage of scale.

Finally, we should note that we have also computed the annual efficiency scores, λ_i^+ and λ_i^-, for each sector. In this analysis of annualized data, because the number of observations for each sector decreases dramatically, it becomes more difficult to obtain a general trend. As described in Daraio and Simar (2007), under both the FDH and DEA method, larger quantities of data provide more sensible estimates.[34] Still, similar to the analysis including all the years for each sector, the analysis of the annualized data shows that the middle-sized firms tend to have low efficiency scores. On the other hand, as also mentioned in footnote 18, the BCs of sectors 14, 26, and 29 tend to increase and the annual bar graphs of these three sectors are similar to those of other sectors. Finally, we also observe that there are still many low-efficiency firms, as shown in Figure 7.4.

4.3 Bribes and Firm Size

4.3.1 Corruption in Vietnam

Since the initiation of the "Doimoi" process in 1986, Vietnam has shifted from a centrally planned economy to a market-based economy and thus can be categorized as a transition economy, although the government still maintains a central role in the economy (see Gainsborough, 2006). The level of corruption in Vietnam is still quite high. In Transparancy International (2007), Vietnam was ranked 118 out of 163 countries and corruption accounted for 3% - 4% of lost GDP annually. Governmental efforts against corruption did not achieve significant improvement. The 2011 Corruption Perceptions Index of Vietnam was around 2.9 on a scale of 0 (highly corrupt) to 10 (highly clean). A 2010 World bank report (see Anderson et al., 2010) also stated that the control of corruption indicator of Vietnam only increased from 22.9 in 2004 to 33 in 2010 on a 0 to 100 scale.

[34]The detailed results are available from the authors.

According to CIEM (2006), the major causes of corruption in Vietnam include: “(a) abuse of power by public officials; (b) arbitrary decisions related to policies and administration; (c) weak accountability of officials and government agencies; and (d) weak state implementation and monitoring”. Also, Maitland (2012) specified two major forms of corruption in Vietnam. The first type is petty and bureaucratic corruption which refers to low-level, small scale corruption practices. This includes the “greasing the wheel” money that firms pay to officials to obtain licenses, permits, or to escape from other government regulatory requirements, taxes and fees. The primary victims of this kind of corruption are mostly small and middle-sized private enterprises. Another target of bribery is the foreign sector. Maitland (2012) claimed that foreign enterprises in Vietnam could be targeted by rent-seeking activities of civil servants who handle taxes, customs, and export/import licensing.

The last common type of corruption is grand payment, in which bribed officials use their authority to secure contracts for bribers. This type of corruption is largely observed among large state-owned enterprises, especially in public procurements for infrastructure projects and land transactions. In 2009, the annual value of public procurement was around USD S20.47 billion, consisting of more than 22% of the national GDP (see Anti-Corruption Center, 2012). Moreover, according to Thang et al. (2011), the number of competitive tenders in the total value of government procurement contracts decreased dramatically in 2009. Only 53% of procurement contracts followed a competitive bidding process in 2009, compared to more than 72% in 2008.

In our study, we focus mainly on petty corruption, because our analysis focuses on small- and medium-sized enterprises. Although we do not have an estimates of the proportion of this type of corruption relative to the nation’s entire economy, many have argued that petty corruption is not so “petty” when taken as a whole. Recently Centre for Community Support and Development Studies, Centre for Research and Training of the Viet Nam Fatherland Front, and United Nations Development Programme (2015) co-authored a report based on the survey of experiences and assessments of the state’s governance and public administration performance using data collected from 13,552 randomly selected citizens across various locations in Vietnam. The authors of the report claimed that 42% of respondents said they had to pay unofficial fees for services at district-level hospitals, and 30% agreed that corruption existed when people applied for land-use licenses.

4.3.2 Data Description

We used the Small and Medium Manufacturing Enterprise (SME) to analyze the relationship between bribes and firm size. The SME survey was conducted by the Central Institute for Economic Management (CIEM) in collaboration with the Vietnamese Institute of Labor Science and Social Affairs (ILSSA), the Department of Economics (DoE) of the University of Copenhagen, and UNU-WIDER in conjunction with the Royal Embassy of Denmark in Vietnam. We excluded joint ventures from the sample because of the high degree of governmental and foreign involvement in ownership structures. Unlike the first data set, this survey only deals with nonstate manufacturing enterprises. It excludes all state-owned firms as well as firms in other sectors such as the service sector. Third, firms in the survey include both formally registered ones (enterprises with a business registration license and/or a tax code) and informal households. All informal firms included in the survey operate alongside the officially registered enterprises. These informal household establishments (firms without a business registration license or a tax code and not registered with District authorities) are included in the surveys based on the on-site identification process. The inclusion of unregistered firms is another important contribution of the survey.

Our initial sample contains 9567 observations, collected from all firms in the SME survey covering four years: 2005, 2007, 2009 and 2011. We excluded all firms for which we did not have information on firm size, wages and total assets. We also discarded firms for which we did not have information on firm age. The final sample consisted of $8,421$ non-state manufacturing firms in 10 provinces of Vietnam. Employees in our analysis included both regular full-time and casual workers.

4.3.3 Snapshots

We used the SME survey data to create Table 3, which is a matrix of employment transitions showing what percentage of firms that shift from one group to another.[35] A micro group was defined as firms with one to nine employees. A small group consists of firms with 10 to 49 employees. A medium firm-size group consists of firms with 50 to 300 employees.[36] Few firms

[35]For a data description, refer to Section 4.3.2.

[36]This is the definition of small and medium enterprises given in Vietnamese law (Decree no. 90/2001/CP-ND on "Supporting for Development of Small and Medium Enterprises").

shifted from the micro category to the small or medium category, or from the Small category to the Medium category. Most firms remain in the same category. Another feature is that there are not many firms who shift down from Medium to Small. This matrix indicates that, in these years, very few firms shifted upward from Micro to another category, and on the other hand, very few firms shift downward from medium to micro group.

Table 3: Employment Transition Matrix (2005–2011)

'05 \ '07	Micro	Small	Medium
Micro	1,436 (90.0%)	144 (10.0%)	1 (0.0%)
Small	134 (19.8%)	490 (72.5%)	52 (7.3%)
Medium	3 (1.6%)	47 (25.5%)	134 (72.8%)

'07 \ '09	Micro	Small	Medium
Micro	1053 (90.2%)	109 9.3%)	5 (0.4%)
Small	121 (22.2%)	390 (71.4%)	35 (6.4%)
Medium	3 (2.2%)	35 (25.2%)	101 (72.7%)

'09 \ '11	Micro	Small	Medium
Micro	1,193 (92.2%)	96 (7.4%)	5 (0.4%)
Small	152 (25.3%)	416 (69.3%)	32 (5.3%)
Medium	5 (2.7%)	48 (26.1%)	131 (71.2%)

Further, in the survey, firm owners were asked whether they had paid bribes in the surveyed year. Table 4 shows the number of firms that bribed or did not bribe in each group. It is clear that the proportion of firms paying bribes was much higher in the last three groups, compared to the proportion in the first or second groups.

Table 4: The Number of Firms Bribing/Not Bribing in Each Group

	group 1	group 2	group 3	group 4	group 5
Not Bribe	3184	1449	171	91	54
Bribe	1383	1605	304	186	60

In this section, we presented Figure 7.7 and Figure 7.8, which are based on a classification by a relative ranking.[37] We ranked firms in ascending orders of size and partitioned them.[38] In Figure 7.7, each group contains 20 firms, and in Figure 7.8, each group represents 10 firms. Each dot shows how many firms paid bribes in each group. For example, the first dot shows the proportion of the smallest 20 firms that paid bribes, out of the 20 firms in this group.

Moreover, in Figure 7.8, we separate the firms into two groups, depending on whether each firm had a tax code. We called those with tax codes *formal* firms and the rest *informal* firms.

[37]Figures based on a classification by the absolute number of employees are also available upon request. The figures show that the likelihood of paying bribes increases with size based on a relative ranking.

[38]If there were some firms of the same size, we used a computer program to permutate them randomly. We repeated this procedure several times to see whether the feature of increasing likelihood of bribes as size increased still helds.

In Figure 7.8, we selected firms having fewer than 200 employees. Informal firms tended to be smaller, and for a good comparison, we selected 200 firms instead of 500, unlike Figure 7.7.

Insert Figure 7.7 and Figure 7.8

From Figure 7.7 and Figure 7.8, it appears that larger-sized firms were more likely to pay bribes, and formal firms were more likely to pay bribes than informal firms in each size group, particularly in 2005 and 2007.

4.3.4 Logit Regression on Bribes

As a preliminary investigation, we conducted an analysis of variance (ANOVA) on the categorized samples, as shown in Table 4. Let Y_{ij} denote the portion of firms bribing out of the total number of firms in each group $i \in \{1, \cdots, 5\}$ for the year $j \in \{05, 07, 09, 11\}$. Then, the ANOVA model is represented by:

$$Y_{ij} = \mu + \alpha_j + \epsilon_{ij},$$

where μ is the overall average, α_j is the treatment effect for group j, and the error terms, ϵ_{ij}, are assumed to be independent and normally distributed with mean 0 and standard deviation σ.

To expand our study of the relationship between the probability of paying bribes and the firm sizes, we conducted a logit regression on bribes and assessed whether paying bribes was related to firm sizes. Assets included both short-term and long-term assets of the firm. Tax code (denoted by TAX) is a dummy that represents the formality of firms. This variable was given the value of 1 if a firm acquired both a business registration certificate (BRC) and a tax code and otherwise the value was 0.[39] The 10 dummies for firms' location are represented by PRO_j for each district j. Finally, TYP_k represents the five firm-type dummies for each k. The firms'

[39] There are two ways to define a formal firm in the survey: (i) a firm that holds a BRC, and (ii) a firm that holds both a BRC and a tax code. A firm can have a BRC without a tax code, but it is not possible to have a tax code without a BRC, because the tax authorities require a BRC before issuing a tax code. The two definitions differ in the level of commitment to formalization and thus the extent to which firms are visible to civil servants. Further investigation using the transition matrix of formal incidence using the first definition provides some contradictions. For example, about 10.7% of firms obtained BRCs in 2007, while about 13.1% of registered firms lost their BRCs in that year. This may indicate some misreporting. Therefore, we used the second definition in the analysis.

dummies include five types: a household business, a limited liability company (hereafter Ltd), a private enterprise, a joint stock, and a cooperative.[40]

Unregistered firms usually have restricted access to credit, infrastructure, public services and markets. Such firms, therefore, may try to become formal to overcome such limitations. On the other hand, remaining informal enables firms to escape the heavy burdens of regulations such as taxes or labor-related requirements. It is, therefore, of interest to see which force is dominant in the Vietnamese case.

Let Y_{ij}^k denote whether firm i of type k in district j paid bribes. If the firm paid bribes, Y_{ij}^k equals 1, and otherwise 0. Then, we considered the following formulation

$$Y_{ij}^k = \beta_0+\beta_1 \cdot \log(size_i)+\beta_2 \cdot \log(age_i)+\beta_3 \cdot \log(asset_i)+\beta_4 \cdot TAX_i+\beta_5^j \cdot PRO_j+\beta_6^k \cdot TYP_k+\epsilon_{ij}^k.$$

To estimate the logit regression, we used the dummy of household business from the set of firm type dummies and the dummy of Province 10 from the set of province dummies for normalization.

Finally, we estimated two sets of regressions in order to check the robustness of our regression results. The first one was a fixed-effects regression that used each firm's ID recorded in the survey, because there might have been some reverse causality between bribes and firm size. The second one took differences in formality into account using a mixed-effect regression, and we assume a fixed effect over the formality, and random effects over age and location.[41]

[40]A household business is a business that is not registered as an enterprise under Vietnam's Enterprise Law. Many businesses operate as a household business, both informally (i.e.,without a license) and formally (i.e., with a license). An Ltd is established by member capital contribution to the company. An Ltd is "a legal entity separate from the owner(s), the owner's liability for the firm's debts and obligations is limited to his capital contribution" (Socialist Republic of Vietnam, 2005). A private enterprise is "a firm owned by an individual, who is its legal representative. The owner has total discretion in making business decisions and is liable for its operations because he owns all firm assets. Each individual can only establish one private enterprise" (Socialist Republic of Vietnam, 2005). A joint stock company is "a company whose capital is divided into shares, and the liability of each shareholder is limited to the par value of the shares held by him/her." (Socialist Republic of Vietnam, 2005). A cooperative is a collective economic organization that is founded by individuals, households or legal entities who have common demands and benefits, and who volunteer to contribute capital and labor (Socialist Republic of Vietnam, 2003).

[41]We estimated various combinations of fixed and random effects over these variables and used Akaike Information Criteria (AIC) to choose this particular combination.

4.3.5 Empirical Findings on Bribes

Our regression results from the logit regression are summarized in Table 12. The results from the fixed-effect and mixed-effect regressions are presented in Table 13. Table 14 shows the odds ratio and the 95 % confidence interval. The odds ratios computed from the results of these regressions are also presented in Table 14. Finally, Table 15 shows that the p-value from the ANOVA, conducted for a preliminary study of the data, was 0.0039; thus, we rejected the null hypothesis that the proportion of firms paying bribes came from the same distribution across different groups at the 5 % significance level.

Insert Table 12, Table 13, Table 14 and Table 15.

Table 14 indicates that firm Size has some positive association with the likelihood of paying bribes, and further, that the likelihood of bribing compared to not bribing increased by 1.68 if a firm had a tax code, compared to a firm not having a tax code. Further, observing a negative effect of firm age may appear to be counterintuitive, because the longer a firm has operated, the greater the firm's visibility is likely to be, which may lead to an increased likelihood of paying bribes. Our interpretation of this result is that a firm is perhaps more likely to need to pay bribes at the start of business.[42] Comparing the odds ratio from the ordinal OLS and the one from the fixed-effect regression, we see that age had a negative effect in the ordinal OLS, whereas age had a positive effect in the fixed-effect regression. We can conclude that if a firm pays bribes at the start of business, then the likelihood of paying bribes increases over time.

Table 14 also indicates that firms' location is important to the likelihood of paying bribes. Compared to being located in Province 10, being located in Province 9 resulted in the highest likelihood of paying bribes. Province 7 had the second highest likelihood of paying bribes. Firm type also had an interesting effect on the likelihood of paying bribes. Ltd and Private firms had higher likelihoods of paying bribes than did cooperative or joint stock firms.

Finally, Figure 7.8 shows that the formality of firms was an important determinant of the likelihood of paying bribes. It is well known that the informal sector plays an important role in many developing economies. According to Schneider and Enste (2002), the informal sector generates from 10% to 20% of the aggregate output in developed countries and more than 30% for

[42] Another interpretation could be that firms paying bribes might be less competitive and thus exit rates for these firms might be higher, which lead to lower survival rates for those firms.

developing countries, with some reaching more than 50%. Our analysis indicates that corruption could be one possible explanation for the dominance of the informal sector in a developing country.

It is also known that corruption is one of the most significant barriers to economic growth. Many studies have provided evidence that corruption reduces human capital, discourages investment, leads to a mis-allocation of resources, and slows down economic development. However, few studies have analyzed possible causes of corruption, and most of them have used cross-country data on corruption perception(Méndez and Sepúlveda, 2006, Mo, 2001b), rather than firm-level data from one nation. Our analysis provides insights on the dynamics of bribery at the firm level.

4.4 Conclusion

In this chapter, we used Vietnamese data to examine the relationship between firm size and production efficiency, as well as the relationship between firm size and the likelihood of paying bribes. We presented results indicating that middle-sized firms tend to produce less efficiently relative to small-sized or large-sized firms; and that the likelihood of paying bribes is correlated with firm size. Thus, our study fills in the gap in the literature by analyzing firms in a developing country in which small firms dominate the economy and exhibit higher efficiency levels, compared to their middle-sized counterparts.[43] Moreover, the heterogeneity of firm size at the bottom of the U-shape across industries indicates the importance of sector level analysis in investigating the missing middle. Our analysis also contributes in this point.

One might question whether our findings from a transition economy—Vietnam—provide a sufficient basis for generalization, particularly because much of the missing middle literature has concerned Africa. The business environment and its features vary substantially from country to country. It would be interesting to see if we observe a similar feature in a different country.

Several interesting future research paths follow from our analysis. A next step is to conduct a

[43]In the literature, there has been an increasing interest in the relationship between firm sizes and their characteristics such as innovation and market structure (see Acs and Audretsch, 1987), growth and productivity (see Bentzen et al., 2011) or job creation (see Dalton et al., 2011). There are, however, few empirical studies on the relationship between firm size and efficiency level, and studies of a developing country are particularly rare. Also, in preceding works, large firms have been found most efficient (see Angelini and Generate, 2008, Leung et al., 2008).

similar analysis on production efficiency with data from developed countries and to compare the results for Vietnam. If we observe lower productivity for middle-sized firms, we could then consider the cause. If we only observe this U-shaped pattern in developing countries, we could examine the reasons that this pattern exists in developing countries but not in developed countries.

Finally, it would be interesting and important to study the causes of these differences in productivity across different firm sizes in developing countries. OECD Publishing (2013) reported two main reasons among OECD countries for the differences in productivity across different firm sizes: (a) firm size matters for productivity, and (b) structural differences at the industrial level of an economy affect the relative performance of firms with different scales across countries. First, in most countries, it has been shown that there is possibility for improving production efficiency by increasing firm size. Larger firms are on average more productive than are smaller ones. In large-scale industrial sectors in relatively low income countries, large firms are, on average, 2–3 time more productive than smaller firms. However, large services sectors in relatively high income countries, small firms are usually more productive than are large firms. It would be interesting to investigate whether these factors are also important in developing nations.

CHAPTER 5

INDUSTRY INEFFICIENCIES AND FIRM SIZE

5.1 Introduction

In the previous chapter, we investigated empirically the relationship between firm size and production efficiency at the firm level. We showed that middle-sized firms' production efficiencies tend to be lower than the efficiencies of small-sized or large-sized firms in most manufacturing industries in Vietnam. The result, however, was based on a firm-level efficiency analysis. Because the missing middle in firm size distribution represents a macro phenomenon, summarized these scores into a few industry-level efficiency score and investigated those in relation to firm size.

The problem of aggregation has been studied in various fields of economics, such as consumption, production, and investment. Klein (1946) opened the discussion on the consistent aggregation of individual economic relations. Later, Debreu (1951) introduced the coefficient of resource utilization (CRU) to measure the overall efficiency of an economy.

In efficiency analysis, the aggregation problem is more complicated, compared to firm-level analysis because the analysis can be carried out at many levels of aggregation, for example, at the plant, firm, industry or economy levels. Moreover, in a multi-input, multi-output analysis, the level of input and/or output aggregation significantly affects the accuracy of efficiency measurements.

The inefficiency of an industry (or a group of firms or production plants) is defined as the difference between the observed production of the industry and the potential production that would be possible if all resources were allocated and used efficiently. To the best of our knowledge, the first paper to address industry inefficiency explicitly in a linear programming framework was presented by Ray and Hu (1997). Ray and Hu introduced a basic model in a primal context, in which only input and output quantities were observed. Later, Lozano et al. (2004) introduced the so called centralized resource allocation model which used the same idea (and it is in fact a special case of Ray and Hu (1997)'s model in which the number of firms is fixed to the observed one). Ray and Hu interpreted the same optimization model from the point of view of a central planner who must allocate resources efficiently. After these attempts, other researchers developed models to accommodate alternative empirical settings (see Aparicio et al., 2013, Aparicio and Pastor, 2012, Asmild et al., 2009, 2012, Fang, 2013, Fang and Zhang, 2008, Giménez-García et al., 2007, Lotfi et al., 2010, Lozano and Villa, 2004, 2005, Lozano et al., 2004, 2009, 2011, Mar-Molinero et al., 2012, Ray, 2007, Ray and Mukherjee, 1998). Ray et al. (2008) extended the industry efficiency model using a cost function approach with input prices varying across locations. Following this idea, Peyrache (2015) introduced an indirect characterization of the problem, which involved solving a mixed-integer non-linear program.

Similar to the definition for the inefficiency at the firm level, the definition of the inefficiency of an industry normally depends on the underlying assumptions of the production possibility set. Convexity, thus, is normally considered a critical assumption. In this chapter, we provide an approximation theorem to report all the different specifications introduced in the literature applied to the same basic optimization program. This program is linear and can be solved using standard simplex methods. This process introduces advantages in terms of both computation and interpretation. The computational advantage is obvious. The interpretational advantage comes from the fact that we now need to discuss a very simple basic model to which we can refer for any type of aggregate analysis. Thus, this model is a unifying tool for a body of literature that has been fragmented and dispersed on this topic. Essentially, in our analysis, we prove that the industry inefficiency is approximately the same if measured under different technology, and the convexity and scale assumptions made regarding the firm-level technology become irrelevant when we consider the overall inefficiency of the industry. The theorem is of particular importance in our case. As indicated in the previous chapter, evidence shows the existence of non-convexity in firm production technology, and efficiency score obtained from a convex frontier such as DEA may be biased. In contrast, with the theorem, the discussion on industry inefficiency is independent of the convexity of the production technology.

Moreover, to investigate the relationship between aggregate efficiencies and the missing middle further, we extended the proposed framework to the meta-technology approach pioneered by Battese and Rao (2002), Battese et al. (2004) and to the group-wise efficiency proposed in Nesterenko (2007). In particular, we segmented the input and output matrices of Vietnamese manufacturing industries into groups of different size and investigated the differences in efficiencies between size groups and the meta-estimates. We found differences between the group-frontiers and the meta-frontiers. In addition, the meta-technology efficiencies (MTE) were highest for the small-sized group, and in particular for groups of firms with fewer than 50 employees. This finding confirms the differences in the characteristics of the firms' operating environments across size groups.

Thus, the chapter contributes to the literature on meta-frontier analysis. Hayami (1969) first proposed a meta-production function to measure the difference in productivity in the agricultural sector across differences among countries. Hayami and Ruttan (1970), subsequently, developed a meta-frontier concept based on the meta-production function to compare efficiencies across groups of firms/units. The meta-production function is, thus, defined as the boundary of an unrestricted technology set. Battese and Rao (2002), Battese et al. (2004) were first to introduce the concept of the meta-frontier for measuring the technical efficiency and technology gap effects separately, using stochastic frontier analysis (SFA). Later, O'Donnell et al. (2008) introduced the analytical framework necessary for a metafrontier. They also show that a metafrontier can be estimated either by non-parametric or parametric methods.

The rest of the paper is organized as follows. In section 2, we define the firm (or plant) and the industry (or aggregate) technologies and present our approximation results. Section 3 provides a decomposition of this inefficiency into sources components. In section 4, we discuss, for the sake of completeness, the extension to the dynamic case. Section 5 provides an empirical application based on a number of manufacturing industries of the Vietnamese economy to shed light on the missing middle phenomenon. Section 6 concludes the chapter.

5.2 Technology definition

5.2.1 Firm level technology and inefficiency

In the previous chapter, the production set of a firm was defined in the one-output, three-input space. In the following analysis, we extend the terminology is to a multiple-input and -output case. In particular, we consider an industry that produces $\mathbf{y} = (y^1, ..., y^M) \in R_+^M$ outputs using $\mathbf{x} = (x^1, ..., x^N) \in R_+^N$ inputs. Data are observed for a group of K firms. We consider (possibly unbalanced) panel data settings by assuming that there are T periods (each period is denoted by $t = 1, ..., T$). We can collect all the observations at time t into input and output matrices which represent our dataset:

$$\left(\mathbf{X}^t, \mathbf{Y}^t\right) \ , \ \forall t = 1, \dots, T$$

These matrices show the observations on the rows. The data-generated production possibilities set is defined as

$$\Psi^t = \left\{(\mathbf{x}, \mathbf{y}) : \ \phi\mathbf{x} \geq \boldsymbol{\lambda}\mathbf{X}^t\,, \ \phi\mathbf{y} \leq \boldsymbol{\lambda}\mathbf{Y}^t\,, \ \sum \boldsymbol{\lambda} = 1\right\}. \tag{1}$$

The production possibilities set Ψ^t is similar to Ψ_i defined in the previous chapter, and satisfies all regularity conditions as Ψ_i. The inclusion of the intensity vector λ and the scale variable ϕ, however, enable us to represent different technology specifications in one formula.

Assumptions about the intensity variables $(\boldsymbol{\lambda}, \phi)$ are crucial for the definition of alternative production sets because they define the convexity and scale properties of the technology [1]. The scale properties can be easily characterized using the parameter ϕ:

- constant returns to scale (CRS): if $\phi \geq 0$;

[1] We note that in principle it is possible to nest the meta-technology approach pioneered by Battese and Rao (2002), Battese et al. (2004). To do so we segmented the input and output matrices into groups and imposed additional restrictions on the intensity vectors. Because this does not change the main result of this paper, we omitted the consideration of group specific technologies to economize on notation. The group technologies are inherited from our definition by imposing constraints on the intensity variable λ; in particular:

- $\boldsymbol{\lambda_l}\boldsymbol{\lambda_l}^T = 1\,,\ \boldsymbol{\lambda_p} = 0\,\forall p \neq l$ defines group specific technologies and a meta-technology which is the union of these group specific sets ($\Psi^t = \cup_l \Psi_l^t$);
- $\boldsymbol{\lambda}\boldsymbol{\lambda}^T = 1$ defines a meta-technology which is the convex hull of these group technologies;

- variable returns to scale (VRS): if $\phi = 1$;
- non-decreasing returns to scale (NDRS): if $\phi \geq 1$;
- non-increasing returns to scale (NIRS): if $\phi \leq 1$;

The intensity variables, $\boldsymbol{\lambda}$ characterize the convexity properties of the technology:

- $\boldsymbol{\lambda} \geq 0$ implies convex technologies;
- $\lambda_j \in \{0, 1\}$, $\forall j = 1, \ldots, K$ implies non-convex technologies;

All the previous assumptions obout the intensity variables can be used in any mix to produce alternative technology specifications. Though this definition incorporates most of the specifications discussed in the literature (see Banker et al., 1984, Charnes and Rhodes, 1978), we focus on the following:

(i) DEA-BCC (Ψ_{BCC}): obtained assuming convexity and VRS;

(ii) DEA-CCR (Ψ_{CCR}): obtained assuming convexity and CRS;

(iii) Free Disposal Hull (FDH) (Ψ_{FDH}): obtained assuming non-convexity and VRS;

The advantage of this formulation is that it provides a unified definition of all basic technologies under different returns to scale and convexity assumptions[2]. In the following discussion, we consider the previous three technologies because of their widespread use and because they pose bounds on all the other alternative specifications. It is easily be seen that

$$\Psi_{CCR} \supseteq \Psi_{BCC} \supseteq \Psi_{FDH}.$$

The definition of technology as presented in Equation 1 corresponds to the firm-level benchmark technology, which we will use to assess the technical efficiency of firms.

[2]The definition of firm technology allows also to nest the definition of a non-convex CRS hull. This is obtained assuming $\phi \geq 0$ and $\boldsymbol{\lambda} \in \{0, 1\}$. We decided to focus our discussion only on the BCC, CCR and FDH models because of their popularity. This is without loss of generality because as shown in the later section, the bounds that we derive for the inefficiency measures, are based on the CCR and FDH models: these two models produce the most and the least conservative sets. All the other sets are included between these two.

In this analysis, in contrast to the work in the previous chapter, we used the directional distance function (DDF) (Chambers et al., 1996, 1998) to measure technical inefficiency because we are interested in a measure of aggregate inefficiency, and the additive nature of DDF allows for the summation of individual firm inefficiencies[3]. The DDF is defined as:

$$D(\mathbf{x}, \mathbf{y}; \mathbf{g}) = \max\{\beta : (\boldsymbol{y} + \beta \boldsymbol{g_y}, \boldsymbol{x} - \beta \boldsymbol{g_x}) \in \Psi^t\}. \tag{2}$$

where $\boldsymbol{g} = (\boldsymbol{g_x}, \boldsymbol{g_y})$ specifies the direction in which inputs are to be contracted and outputs are to be expanded to reach the efficient boundary[4]. $D(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g})$ satisfies some important properties including:

(i) homogeneous of degree -1 in the directional vector $(\boldsymbol{g})$: $D(\boldsymbol{x}, \boldsymbol{y}; \lambda \boldsymbol{g}) = \frac{1}{\lambda} D(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g})$;

(ii) translation: $D(\boldsymbol{x} - \alpha \boldsymbol{g_x}, \boldsymbol{y} + \alpha \boldsymbol{g_y}; \boldsymbol{g}) = D(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}) - \alpha, \ \forall \alpha \in R$;

(iii) if the technology exhibits CRS, then, $D(\lambda \boldsymbol{x}, \lambda \boldsymbol{y}; \boldsymbol{g}) = \lambda D(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}), \lambda > 0$.

The DDF can be computed with respect to any of the technologies we defined earlier. In order to obtain a measure of inefficiency at the firm-level, one can compute the DDF with respect to the firm level technologies. Thus, we obtained a measure of firm technical inefficiency for each of the firm technology definitions (focusing on the CCR, BCC, and FDH models for the reasons explained above):

- BCC (Convex, VRS): $D_{BCC}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}) = D(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g} \mid \Psi_{BCC})$;
- CCR (Convex, CRS): $D_{CCR}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}) = D(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g} \mid \Psi_{CCR})$;
- FDH (Non-convex, VRS): $D_{FDH}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}) = D(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g} \mid \Psi_{FDH})$;

For the way the firm technologies are constructed, it holds that:

$$D_{CCR}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}) \geq D_{BCC}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}) \geq D_{FDH}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}).$$

[3]Readers should also note that the DDF basically generalizes the traditional Shephard distance function defined in the previous chapter.

[4]We can retrieve the Shephard distance function with a radial direction of DDF out of the origin.

In other words, the CCR model provides the least conservative estimate of technical efficiency, and the FDH model provides the most conservative estimate of technical efficiency within the class of technologies considered by definition 1.

5.2.2 Industry technology

In the previous section, we introduced the representation of the firm technology. The definition of technology explicitly allows for alternative specifications of the scale and convexity properties of the technology. In this section, we extend these notions to the industry level. We assumed that any number of firms can operate in the industry (entry and exit of firms is allowed) and all the firms in the industry face the same technology set Ψ^t[5]. The peculiar characteristic of an industry technology definition (as opposed to the firm technology definition) is that we allow for free replicability of production plans. A technology set is said to satisfy free replicability if

$$\forall\,(\boldsymbol{x},\boldsymbol{y})\,,\,(\boldsymbol{x'},\boldsymbol{y'})\in\Psi^t\,,\text{ then }(\boldsymbol{x}+\boldsymbol{x'},\boldsymbol{y}+\boldsymbol{y'})\in\Psi^t.$$

In other words, free replicability allows for any replication or any addition of existing production plans. It should be noted that free replicability does not, in general, imply additivity of the industry or firm technology, although it excludes the possibility of sub-additivity of the industry technology (see Färe, 1986, for a definition of additivity). Sub-additivity of the technology can be excluded by assuming that the production set is convex. This means that free replicability is the same as additivity in a DEA model, but it is weaker than additivity in a non-convex (FDH) model. We can think of a sub-additive industry technology in cases where the addition of new production plants or firms produces negative externalities; but this topic is outside the scope of this chapter.

An alternative interpretation of the industry technology comes from a management perspective, where we rename firms as “production plants” and the industry as a “company”. Our definition of the company technology, then, implies that the company can activate as many production

[5] If one insists in considering the meta-technology approach, then the industry technology will be either the union or the convex hull of the group specific technologies. Our assumption is then that every firm can choose any of the available group technologies (i.e. firms face the same meta-technology). We should also emphasize that once allowance for free replicability is made, the distinction between the meta-technology and the group technologies becomes much less relevant. In fact, the free replicability assumption implies that the industry can pick any production plan from any group technology and in any combination.

plants as desired.

At the industry level, it is possible for a specific production plan to be replicated many times within the industry. In other words, free replicability allows for the replication of any individual firm as well as the combination of proportional production plans in the industry. With this assumption in place, we can define the industry technology as an enlargement of the firm-level technology. The industry production possibilities set will be the union of all the possible sums of production plans of the firm technology (see Peyrache, 2013):

$$\Psi_I^t = \bigcup_{S=1}^{\infty} \sum_{s=1}^{S} \Psi^t. \tag{3}$$

In this definition, S is the number of replicates and these replicates can come from the sum of any allowable production plans at the firm level. It is easy to see that the firm technology is a subset of the industry technology. It should be noted that in general $\sum_{s=1}^{S} \Psi^t \neq S\Psi^t$, unless the technology is convex (see Färe et al., 2008, Li and Ng, 1995, Zelenyuk, 2006, for a proof of this result). Because we allowed for non-convex technologies, we maintained the general definition of free replicability contained in equation 3. Applying this definition to our data-generated technology model (model 1) returns the following explicit definition of industry technology:

$$\Psi_I^t = \left\{ (\boldsymbol{x}, \boldsymbol{y}) : \phi \boldsymbol{x} \geq \boldsymbol{\lambda X^t} \,,\ \phi \boldsymbol{y} \leq \boldsymbol{\lambda Y^t} \,,\ \sum_k \boldsymbol{\lambda_k} = S \,,\ S \in \{0, 1, 2, \ldots\} \right\}. \tag{4}$$

The industry technology represents an enlargement of the firm-level production technology. We note that in the CRS convex case the industry technology definition is trivially satisfied by the firm-level production set (i.e., a CRS convex firm-level technology is equivalent to the associated industry technology enlargement). In all other cases, the free replicability assumption enlarges the firm production possibilities set. Alternative assumptions on the intensity vector return different industry-level technologies. We shall consider the following ones (in which the parameter S is free)[6]:

- the convex, VRS, free replicable set (Ψ_I^{BCC}): $\phi = 1$;

[6] We are once again only focusing on the free replicable enlargement of the BCC, CCR and FDH technologies. This is without loss of generality.

- the convex, CRS free replicable set ($\Psi_I^{CCR} = \Psi_{CCR}$): as mentioned before, this is a trivial case in which the firm technology definition actually satisfies free replicability;
- the non-convex, VRS, free replicable set ($\Psi_I^{FDH} = \Psi_{FRH}$): $\phi = 1$ and $\lambda \in \{0, 1\}$. This industry technology is equivalent to the free replicability hull (FRH) proposed in Tulkens (1993), which allows for integer replications of all observations and sum of observations.

From the previous definitions, we can make the following conclusion about the relationship among different industry technology sets:

$$\Psi_I^{CCR} \supseteq \Psi_I^{BCC} \supseteq \Psi_I^{FRH}.$$

In other words, all the industry technologies we consider were contained in the convex CRS industry technology and contained the non-convex VRS industry technology (the free replicable set). To measure inefficiency at the industry level, we benchmarked any input-output combination against an industry technology set. The industry-level DDF was then calculated by solving either a linear or a mixed-integer linear program. The result of such a program returned the optimal intensity vector ($\boldsymbol{\lambda}$), the optimal value of the DDF and the optimal value for the parameter S. The parameter S is interpreted as the optimal number of firms that should populate the industry in order to reach a fully efficient configuration. At this point, it is relevant to present the optimization program explicitly. Thus, we write the program for the FRH because this is the set for which we have the highest number of restrictions on the intensity variables (the other sets can be obtained by relaxing these restrictions):

$$\begin{aligned} D_I^{FRH}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}) = \max_{\boldsymbol{\lambda}, \phi, S} \ & \beta \\ st \ & \boldsymbol{x} - \beta \boldsymbol{g_x} \geq \boldsymbol{\lambda X^t} \\ & \boldsymbol{y} + \beta \boldsymbol{g_y} \leq \boldsymbol{\lambda Y^t} \\ & \sum_k \lambda_k = S \\ & \lambda_k \in \{0, 1\} \quad \forall k = 1, \ldots, K \\ & S \in \{0, 1, 2, \ldots\}. \end{aligned}$$

In accordance with the three types of industry level technologies, three different types of inefficiency scores can be obtained:

- Convex, VRS: $D_I^{BCC}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g})$;
- Convex, CRS: $D_I^{CCR}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g})$;
- Non-convex, VRS: $D_I^{FRH}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g})$;

All these measures of industry inefficiency are based on alternative convexity and scale assumptions of the underlying firm-level production technology. From the definition of industry-level technology, we can conclude that:

$$D_I^{CCR}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}) \geq D_I^{BCC}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}) \geq D_I^{FRH}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}).$$

Therefore, all the industry inefficiency measures are bounded from the convex CRS technology and the free replicability hull. In the next proposition, we clarify the connection between these two fundamental industry inefficiency measures.

Proposition 5.1. *If the industry is large enough, then:*

$$D_I^{FRH}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}) \simeq D_I^{CCR}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}). \tag{5}$$

Proof. From our definitions, we have

$$D_I^{FRH}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}) + \delta = D_I^{CCR}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}).$$

where the residual difference $\delta \geq 0$ is clearly bounded. Without loss of generality, we can write any input-output combination as $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}) = \frac{1}{K}(\boldsymbol{x}, \boldsymbol{y})$ where $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$ is the average input-output bundle and K is the number of replicates (number of firms). Applying property (iii) of the DDF we can write:

$$D_I^{FRH}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}) + \delta = D_I^{FRH}(K\bar{\boldsymbol{x}}, K\bar{\boldsymbol{y}}; \boldsymbol{g}) + \delta = K D_I^{CCR}(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}; \boldsymbol{g})$$

By dividing both sides by K, we have

$$\frac{1}{K}D_I^{FRH}\left(K\bar{\boldsymbol{x}}, K\bar{\boldsymbol{y}};\boldsymbol{g}\right)+\frac{\delta}{K}=D_I^{CCR}\left(\bar{\boldsymbol{x}},\bar{\boldsymbol{y}};\boldsymbol{g}\right)$$

The right-hand side $D_I^{CCR}\left(\bar{\boldsymbol{x}},\bar{\boldsymbol{y}};\boldsymbol{g}\right)$ is independent of the number of firms K. We have, δ is bounded. Moreover, since replicability implies super-additivity (Bogetoft and Wang, 2005)[7], we have $\frac{\delta}{K}$ converges to zero and the left hand side of the above equation converges to $D_I^{CCR}\left(\bar{\boldsymbol{x}},\bar{\boldsymbol{y}};\boldsymbol{g}\right)$ as $K\rightarrow\infty$. In other words, when industry production expands because the number of firms producing the average input-output vector increases, we have

$$\frac{1}{K}D_I^{FRH}(K\bar{\boldsymbol{x}}, K\bar{\boldsymbol{y}};\boldsymbol{g})\rightarrow D_I^{CCR}(\bar{\boldsymbol{x}},\bar{\boldsymbol{y}};\boldsymbol{g})$$

A more direct interpretation of this result can be achieved when we consider a large K:

$$\frac{1}{K}D_I^{FRH}(K.\bar{\boldsymbol{x}}, K.\bar{\boldsymbol{y}};\boldsymbol{g})\simeq D_I^{CCR}(\bar{\boldsymbol{x}},\bar{\boldsymbol{y}};\boldsymbol{g})$$

in which case we obtain the statement of the proposition (after re-arranging terms). □

This proposition shows that the industry inefficiency is approximately the same if measured under a convex CRS technology or a free replicable hull. In other words, the convexity and scale assumptions made on the firm-level technology become irrelevant when we consider the overall inefficiency of the industry! This is further clarified in the next corollary.

Corollary 5.2. *If the industry is large enough, then*

$$D_I^{FRH}\left(\boldsymbol{x},\boldsymbol{y};\boldsymbol{g}\right)\simeq D_I^{BBC}\left(\boldsymbol{x},\boldsymbol{y};\boldsymbol{g}\right)\simeq D_I^{CCR}\left(\boldsymbol{x},\boldsymbol{y};\boldsymbol{g}\right).$$

This corollary follows immediately from the fact that the FRH inefficiency is converging to the convex CRS inefficiency, and the BCC distance is bounded by these two. To restate the result, the industry inefficiencies are approximately the same regardless of the underlying convexity and scale assumptions imposed on the firm-level technology. As a result, the industry inefficiency of any technology in the class defined by equation 3 can be approximated by the

[7] $(x_A, y_A), (x_B, y_B)$ are called super-additive if we have: $f(x_A + x_B, y_A + y_B; g) \geq f(x_A, y_A; g) + f(x_B, y_B; g)$.

efficiency of a convex CRS technology. This bears some important computational and analytical advantages because we no longer have to solve complex optimization programs (integer, mixed-integer or non-linear mixed integer) for measuring industry inefficiency. In fact, all the industry inefficiency notions proposed in the literature can be re-constructed to a basic case that can be solved via a simple linear program. In addition, the discussion regarding whether to impose convexity and different scale assumptions onto the technology becomes irrelevant at the industry level. To simplify notation, in the next two sections, we omitted references to the convexity and scale assumptions of the industry model, and just use $IE^t = D_I(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g})$ to refer to the industry inefficiency (as approximated by the CRS convex hull).

5.3 Industry Inefficiency Decomposition

In the previous section, we showed that the industry inefficiency is approximately the same, regardless of the assumptions on returns to scale and convexity of the underlying firm technology. In this section, we discuss the components of such an industry inefficiency measure. An aggregate indicator of firm technical inefficiency was obtained by summing the DDF scores for all the firms in the sample (see Briec et al., 2003, Färe and Zelenyuk, 2003, Färe et al., 2008, Färe and Primont, 2003). Because we have three alternative definitions of firm-level inefficiency (CCR, BCC, FDH) we will obtain three alternative measures of aggregate performance:

$$ITE^t_{BCC} = \sum_k D_{BCC}(\boldsymbol{x_k}, \boldsymbol{y_k}; \boldsymbol{g}).$$

$$ITE^t_{CCR} = \sum_k D_{CCR}(\boldsymbol{x_k}, \boldsymbol{y_k}; \boldsymbol{g}).$$

$$ITE^t_{FDH} = \sum_k D_{FDH}(\boldsymbol{x_k}, \boldsymbol{y_k}; \boldsymbol{g}).$$

where it holds that:

$$ITE^t_{CCR} \geq ITE^t_{BCC} \geq ITE^t_{FDH}.$$

It should be noted from the above equation that the free disposal hull and the CRS hull pose bounds on the movements of the aggregate firm technical inefficiency. This is not surprising, considering the fact that the FDH results in the most conservative estimate of technical inefficiency, and the CRS results in the least conservative estimate of technical inefficiency. Because we know from the theorem in the previous section that the industry inefficiency for each time period can be approximated by the convex CRS technology (irrespective of the underlying assumptions on scale and convexity), we obtain the following decomposition of the industry inefficiency:

$$IE^t = ITE^t + IRE^t. \tag{6}$$

where the ITE^t component is one of the previous technical efficiency measures (and clearly $IE^t \geq ITE^t$). IE^t is the overall industry inefficiency, which is a measure of aggregate loss in outputs or waste in inputs at the industry level; IRE^t is a reallocation effect, that measures the contribution to the overall industry inefficiency, arising because of a mis-allocation of resources across production units. As it stands, the left-hand side of this equation remains approximately unchanged if one changes the convexity and scale assumptions on the firm-level technology. In contrast, the decomposition of the equation on the right-hand side depends on the assumptions one is willing to make regarding the technology. The convex CRS firm technology returns the most conservative estimate of the reallocation effect; in contrast, the FDH firm technology returns the most conservative estimate of the technical efficiency effect. The FDH and CRS inefficiencies set bounds for the values that this decomposition can take. Therefore, the difference between the two gives insight into on the stability of the decomposition relative to alternative convexity and scale specifications of the firm technology. An alternative way of looking at the result contained in our theorem is in terms of this decomposition. Because the left-hand side is approximately invariant to the specification of the firm-level technology, choosing the most conservative estimate of firm-level inefficiency (the FDH) implicitly means that one is assigning the least conservative weight to the reallocation component in the decomposition (and vice versa). In other words, trying to build a benefit-of-the-doubt indicator for the firm-level inefficiency returns the highest possible value of the reallocation component!

A simple interpretation of the previous decomposition can be given in terms of the percentage contribution of each component:

$$\%ITE^t + \%IRE^t = 1.$$

where $\%ITE = ITE^t/IE^t$ and $\%RE^t = RE^t/IE^t$. Showing the percentage contribution of each component makes the decomposition easier to understand.

5.4 Industry productivity change

In this section, for the sake of completeness, we briefly consider a comparison of the industry between two time periods. The industry inefficiency change (IEC) is defined as the difference of the industry inefficiencies at the two time periods:

$$IEC = IE^t - IE^{t+1}. \tag{7}$$

IEC has a positive (negative) value if the industry performance deteriorates (improves) over time. Because the measure of industry inefficiency is approximately equal for alternative convexity and scale assumptions, so is its change in time. The change in inefficiency is only one component of productivity change, the other one being the shift of the production frontier itself. A measure of technical change, which captures the shift of the production frontier can be calculated using the Malmquist-Luenberger method as the average of industry technical change at the base (ITC^t) and comparison (ITC^{t+1}) periods. More explicitly,

$$ITC = \frac{1}{2}\left[D_I^{t+1}(\boldsymbol{x^{t+1}},\boldsymbol{y^{t+1}};\boldsymbol{g}) - D_I^{t}(\boldsymbol{x^{t+1}},\boldsymbol{y^{t+1}};\boldsymbol{g}) + D_I^{t+1}(\boldsymbol{x^{t}},\boldsymbol{y^{t}};\boldsymbol{g}) - D_I^{t}(\boldsymbol{x^{t}},\boldsymbol{y^{t}};\boldsymbol{g})\right]. \tag{8}$$

ITC has a positive (negative) value if the industry production frontier shifts inwards (outwards). Putting the previous two definitions together, we obtain a measure of industry productivity change (IP):

$$IP = IEC + ITC.$$

We note, once again, that these measures of technical change and efficiency change at the industry level are approximately invariant to the choice of the scale and convexity properties of the underlying firm technology set. Therefore, when it comes to measuring productivity change at the industry level, the underlying convexity and scale assumptions on the firm level technology become essentially irrelevant. This very strong result stems from our approximation theorem. This result also justifies the common practice of using CRS technologies to measure productivity change at the industry level.

We can decompose industry productivity change further by noting that the industry inefficiency change component itself can be decomposed. Taking into account the efficiency decomposition introduced in the previous section, productivity change can thus be further decomposed into

$$IP = ITC + ITEC + IREC, \tag{9}$$

where $ITEC = ITE^t - ITE^{t+1}$ and $IREC = IRE^t - IRE^{t+1}$. The first component on the right-hand side (ITC) is invariant to the choice of the convexity and scale properties of the technology. The other two components depend on the bounds we identified earlier.

5.5 An empirical illustration

5.5.1 Data

In this chapter, to be consistent with the analysis presented in the third chapter, we continued to analyze the Census Enterprise Survey collected by the General Statistics Office of Vietnam. We focused on the nine-year period from 2000 to 2008 and included 14 sectors of the manufacturing industry that had at least 200 observations for each year. The number of observations ranged from approximately 3,350 to 32,000 per industry (Details of the dataset can be found in Chapter 3).

Next, we considered the situation in which a firm uses capital, labor, and intermediate materials to produce goods, which are measured in monetary terms labor is measured by the total income of employees in a firm. Real revenue was used as a proxy for output. Three inputs were included in our estimation: intermediate inputs, labor, and capital. The intermediate material included costs such as fuel and the value of other materials. Labor was measured by the total income of employees in a firm. This included total wages and other employee labor-related costs such as social security, insurance and other benefits. Capital was measured as assets to be used in production. All input values were adjusted to account for inflation to obtain a real value.

Empirical Results

We applied our previous models to the dataset of Vietnam. For simplicity and to save space, we report only results derived from the static decomposition which focused on industry inefficiency, rather than on productivity change[8]. This choice was dictated by the fact that our main result was based on the static decomposition rather than on the productivity decomposition (the approximate invariance of the productivity decomposition hinges on its static version). Table 16 shows the results for each sector of the Vietnamese economy. For each sector, we report two tables: the first table displays the industry inefficiency (IE) and the aggregate firm inefficiency (ITE) for each year. These quantities are reported for each of the technology specifications considered in the previous sections (CCR, BCC, and FDH). The second table shows the percentage contribution of the ITE and IRE components to the overall inefficiency of the industry. The numbers were calculated for each year and each model. All the inefficiency values reported in these tables are computed using the following directional vector: $\mathbf{g_x} = \mathbf{1}$ and $\mathbf{g_y} = \mathbf{0}$.

First, it should be noted that for all sectors, the industry inefficiency (IE) value (as computed using the directional distance function) was approximately the same under alternative convexity and scale properties of the underlying firm level technology. For example, if we look at sector 14 in table 16, the industry inefficiency shows large variation across different years, but it is largely the same under different model specifications. In year 2000, it was 450,200 for the CCR and BCC models and 450,194 for the FRH model. At the end of the period in 2008, these figures weremore than three times higher but still very similar to each other: 1,466,380 for CCR and BCC and 1,466,340 for FRH. If we look at the aggregate firm inefficiency (ITE) numbers in the same table, it is possible to appreciate the impact of the scale and convexity assumptions on the firm technical efficiency scores. For example, in year 2000, the scores were 250,567, 189,358 and 147,479 for the CCR, BCC and FDH models respectively. These are quite large variations, considering that the ITE-CCR inefficiency for 2000 was almost double the one from the ITE-FDH model. Another interesting feature of these numbers is that the ITE-CCR did not grow as much as the industry inefficiency, being less than double the initial value in year 2008 (474,055). In contrast, the ITE-FDH inefficiency measure grew almost proportionally to the inefficiency of the industry (414,499 in year 2008). Moreover, the difference between alternative models in the first time period is much larger than it was in the last time period in the sector.

The differences between alternative models are more evident if we look at the percentage results for the sectors shown in Table 16. In particular, in year 2000, the CCR model shows that 55.7% of the industry inefficiency can be attributed to aggregate firm inefficiency and 44.3% to

[8]We used Matlab functions to solve the linear and integer-mix programs associated with our efficiency measures under CCR, BCC and FDH technologies. The code is available upon request.

reallocation of resources. These numbers are quite different for the FDH model, with aggregate firm inefficiency accounting for 32.8% of industry inefficiency and the reallocation component accounting for about 67.2% of the industry inefficiency. It should be noted that although the overall inefficiency of the industry remains unchanged thanks to our theorem, the underlying components changed quite a lot.

If we look at the other sectors, our main story is similar. As noted, the industry inefficiency is approximately invariant to the choice of the underlying convexity and scale assumptions on the firm-level technology. The components of this industry inefficiency, however, depend quite crucially on these assumptions. Sector 17 in table 16 is a good example. The industry inefficiency in year 2000 was 1,232,733 and about double at the end of the period at 2,432,830. If we now look at the decomposition of this overall inefficiency, it is possible to appreciate the differences in the decomposition. In most years, with the exception of year 2008, the ITE-CCR accounted for about 90% of this inefficiency. On the contrary, according to the FDH model, aggregate firm-level inefficiency was only about a third of the total inefficiency. Interestingly, the results from these two models indicate quite opposite policy implications. In contrast, according to the CCR model, the central planner should focus on improving technical inefficiency, while according to the FDH model the main problem is the reallocation of resources across production plants.

As noted in chapter 3, the DEA and FDH ratios of firm technical efficiency scores in each sector have prevailed over the existence of non-convexity in production technology of manufacturing firms. The above discussion justifies the use of non-convex technology to investigate efficiency scores in the dataset. Moreover, we emphasize the importance of the reallocative effect or inefficiencies arising because of a mis-allocation of resources across production units in each sector.

In the next step, to analyze further the missing middle in relation to industry efficiencies, we extended the proposed framework to the meta-technology approach pioneered by Battese and Rao (2002), Battese et al. (2004). The meta-frontier framework, first, evaluates each observation relative to its own group's best-practice frontier (where the units of the group are assumed to face homogeneous environmental factors and share a common technology) and, second, to the overall meta-frontier constituted from the best practices of all groups. The meta technology, therefore, is the union of group-specific technologies, and firms in the sector can choose any of the available group technologies. The comparison of these two efficiency scores, then, reveals the technology gap or the meta-technology efficiency (MTE) (see Battese and Rao, 2002, Battese et al., 2004)

In this analysis, we segmented firms in each sector according to their size differences in size[9]. In manufacturing industries, firms of different sizes tend to be exposed to different production environments and technologies. The choice of manufacturing plant scale may prevent manufacturers from taking advantage of productive techniques associated with different plant configurations. Small firms, especially informal ones, for example, usually have restricted access to credit, infrastructure, public services and markets and are subjected to different technology compared to their larger peers. It is also known that corruption is one of the most significant barriers to economic growth. Many studies have provided evidence that corruption reduces human capital, discourages investment, leads to a mis-allocation of resources, and slows down economic development. Moreover, as shown from our result in Chapter 3, as firm size increases, the likelihood of paying bribes also increases. Further, formality of firms matters in the sense that the likelihood of paying bribes is higher for formal firms, which tend to be larger in size.

Table 17 shows the meta frontier estimates for each of the sectors of Vietnamese economy in each year . For each sector, we grouped firms into five different categories according to size (1 to 9 employees, 10 to 49 employees, 50 to 99 employees, 100 to 199 employees and more than 200 employees)[10]. For each group, we reported the industry efficiency scores (IE) under both group technology and meta-technology and the meta technology ratio (MTR) which is the ratio between the two industry efficiency scores[11]l. All the inefficiency values reported in these tables were input-oriented to be consistent with the firm level analysis in Chapter 3. In particular, the scores are computed using the following directional vector: $\mathbf{g_x} = \mathbf{x}$ and $\mathbf{g_y} = \mathbf{0}$. The key findings from the table are discussed in the following paragraphs

First, for the years examined, the industry efficiencies were low under both group and meta frontiers. This indicates that at the aggregate level, there was much room for efficiency improvement in all analyzed manufacturing sectors.

[9]In the meta frontier literature, most studies divided firms into groups of different industries, regions and/or countries. According to Simar and Zelenyuk (2007), firms can be grouped as they are “operating under different regulatory regimes, or different ownership structures (private vs. public, domestic vs. foreign, etc.), or operating in different regions, or efficiencies of countries at different stages of economic development or transition”.

[10]To ensure the robustness of the analysis, we also tried other grouping which returned similar results. The details of the analysis are available upon request.

[11]We calculated the industry inefficiency (IE) under all three different model (CCR, BCC, and FRH). The all three models returned similar results as predicted by our theorem. We, therefore, only report the result of CCR mode.

Second, a difference exists between the group-frontiers and the meta-frontiers. Except for some particular years in the small-sized firm group, the majority of efficiencies were lower under meta technology (i.e. MTRs were smaller than 1). This indicates that the results may be distorted with the group-frontier.

Third, among different firm-size groups, small - sized firms (from 1 to 9 employees and from 10 to 49 employees) tended to have much higher values for MTR. Since MTR captures the distance, and hence, the relative ability of a group to catch up with rest of the firms in the sample, it is evident that small firms tended to operate closer to the industrial potential. Meanwhile, large sized groups (firms with 100 to 199 employees and firms with more than 200 employees) tended to have lower MTRs (the average level was lower than 50%). These firms operated at much lower levels, compared to the industry potential. Another interpretation for the obtained result is that because MTR represents the restrictive nature of the production environment, large-sized firms tended to have a more restrictive production environment. Interestingly, the empirical results in Chapter 3 also showed the larger burden of corruption was borne by larger-sized firms in Vietnam.

5.6 Conclusion

In this chapter, we provided an approximation theorem that simplifies the computation of industry inefficiency scores. The industry inefficiency (IE) value is approximately the same under alternative convexity and scale properties of the underlying firm-level technology. Thus, the theorem bears some important computational and analytical advantages. First, we no longer must solve complex optimization programs for measuring industry inefficiency. In fact, all the industry inefficiency notions proposed in the literature were de-constructed to a basic case that can be solved with a simple linear program. In addition, the discussion regarding whether to impose convexity and CRS onto the technology becomes irrelevant at the industry level. We provided a discussion on the decomposition of the industry inefficiency, which depends on the assumptions one is willing to make about the technology. In particular, the convex CRS firm technology will return the most conservative estimate of the reallocation effect, and the FDH firm technology will return the most conservative estimate of the technical efficiency effect.

The proposition was, then, illustrated using data from a number of different industries in the Vietnamese economy. We showed that the industry inefficiency measure is approximately invariant to alternative convexity and scale properties of the underlying technology. Moreover, the proposed framework was extended to the group and meta-frontier approach. The presence of significant levels of MTR confirmed differences existed across size groups in all analyzed sectors. This finding implies that there is, indeed, a potential to improve the efficiencies of Vietnamese firms, even using the current technology. Moreover, the results also imply that it would be challenging to eliminate the existing inefficiencies in Vietnamese manufacturing sectors because policies and programs for efficiency improvement must target both firms and the firm's operating environment.

CHAPTER 6

CONCLUSION

6.1 Introduction

Cross-country differences between firms' output levels and growth are large and persistent (Hall and Jones, 1999; Klenow and Rodriguez-Clare,1997). In the literature, many approaches have been explored in an attempt to explain such differences. In this thesis, we focused on firm size distribution (FSD) as the manifestation of firm dynamics. In particular, we analyzed the missing middle phenomenon in a developing country, Vietnam. The missing middle refers to the empirical fact that most employment in developing countries is located in either small-sized or large-sized firms. In view of the fact that the missing middle is more evident in developing nations, we surmised that there were some reasons that particularly applied to developing nations, but not so much to developed nations. Thus, understanding the mechanism behind the phenomenon could shed light on cross-country differences in outputs.

In this thesis, we focused on two factors: productivity and corruption. These two factors may seem unrelated. However, productivity and corruption may represent the distortions driving the mis-allocation of resources of firms in the economy. We argue that small firms may be reluctant to become formal to avoid a rise in expected corruption costs. Small firms can then substitute capital for labor to scale-up production and wait for a productivity shock that will offset the cost of growth. Another explanation could involve differences in technology or productivity between firm-size groups.

In this conclusion chapter, we summarize the main results and contributions presented in each chapter. In general, the results we obtained adequately fulfill the objectives of the study, which were to investigate the missing middle phenomenon and to provide some initial evidence of the underlying mechanism behind the phenomenon.

In this chapter, we first present a brief summary of all the results described in preceding chapters. The third section highlights the main contributions of the thesis, followed by discussion of their implications. In the fourth section, we embrace a slightly broader perspective, highlight some short-comings of the thesis and note analyses we did not try (but might have if not for time constraints). We also propose lines of research that we believe are of great interest for future projects.

6.2 Chapter summaries

6.2.1 Chapter 2 - Literature review

The aim of the second chapter was to review the literature. The main content of the chapter was divided into two parts. The first part presented an overview of the major concepts in the thesis. In the second part, we examined the FSD-related literature, both empirically and theoretically.

In addition, we analyzed both the definition of a firm and different measures of firm size.

The definition of a firm has been mentioned in many recent theoretical approaches, including transaction cost theory, property rights theory, knowledge-based view theory, and evolutionary theory, among others. In this thesis, we adopted the definition of a firm as a collection of assets that includes both physical and human capital. This definition of a firm was an extension of the definition used in property rights theory because our definition included labor as a part of the firm operation process.

The choice of measure for firm size can have a significant impact on the analysis. The size of a firm can be represented by many proxies, and variables capture the internal process of a firm differently. Among the possible measures used so far, the number of employees of a firm has been the most widely used indicator of organizational complexity and seemed the most suitable measure for distribution analysis. Thus, in this thesis, we used the number of employees of a firm as the measure of firm size and described the distribution of firms according to this size measure.

The second section of Chapter 2 was devoted to providing an overview of the FSD literature. FSD is a complex issue that requires an integrated analysis of various factors, and this section showed the difficulty of determining the major factors that affect the distribution of firms in the market. First, we presented some features of the FSD, mainly of developed countries. In particular, firm density has been found to be leptokurtic, stable and highly skewed to the right. Moreover, many studies have shown that FSD tends to follow the Gibrat's law whereby firm size is dependent of random shocks that are independent of each other and of the initial size of the firm. Related to the shape of the FSD, the literature provided evidence supporting both the log-normal and Pareto distributions in developed countries.

Factors that have important effects on firm growth and FSD formulation were also presented. These factors were (a) productivity, (b) economies of scale, (c) mergers and acquisition, (d) policy-related factors and (e) the stochastic factor. The next part focused on analyzing these factors under different theories. These included stochastic models, learning and selection models, and behaviorism models. In general, these models differed in their implications for the distribution of firms. Thus, researchers have tended to adopt one approach or another and justify their preference.

These models, however, are based mainly on studies of developed economies. Moreover, it is also a well-known fact that profound differences exist between developed and developing economies in the business environment . As a result, a sharp empirical contrast has been identified between FSDs in developed and developing economies. Although the size distribution in developed countries can be approximated by either the log-normal distribution or the Pareto law, or a combination of the two, the FSDs in developing countries are characterized by the missing middle where middle-sized firms account for the smallest proportion of the total number of firms. The third part of the chapter was, therefore, devoted to summarizing the literature on the missing middle. We acknowledged various explanations that mainly focused on the differences in the business environments of developing countries. These included capital constraints, high corruption levels, differences in regulatory structure, and firms' access to infrastructure, among others. Nevertheless, there is no unified agreement on the factors that can satisfactorily explain such an empirical finding. Given this lack of agreement, the thesis focused on examining the missing middle phenomenon in a developing country - Vietnam, focusing on two factors: corruption and productivity.

Vietnam represented an interesting case. Vietnam is a transitional, developing country. Over the last two decades, with the emergence of the “Doi moi” (Renovation) process, Vietnam has experienced a period of continuously high economic growth. However, Vietnam has failed to become a tiger in Asia as expected. One of the possible explanation for this failure might be the under-representation of medium-sized firms which are commonly the most dynamic and growth-oriented firms. Thus, serious considerations of the missing middle phenomenon was needed to understand the mechanism behind such a bias in FSD because it underpins the misallocation of resources and hinders firm growth.

6.2.2 Chapter 3 - Theoretical model

In Chapter 3, we extended the model presented in Luttmer (2007), which captured the increasing costs of large-sized firm groups, compared to the smaller ones. This phenomenon has lead to the concentration of a continuum of infinitesimal firms in the market. Luttmer noted that the model was more applicable to a developed economy such as the U.S. economy. With our model, in contrast, we attempted to consider the specific characteristics of a developing economy. The model describes firm dynamics - in terms of entry, exit, and growth consistent with empirical evidence in developing countries, while still allowing for heterogeneity in the characteristics of firms.

In the model, the economy we analyzed was a monopolistic competitive market. Time was assumed to be continuous. On the consumption side, there was a continuum of infinitely-lived consumers and each one inelastically supplied one unit of labor at every point of time. The consumer faced a standard prevent-value budget constraint in which the income consisted of claims from firms and labor wages. On the production side, firms were monopolistic competitors. There were many producers, each with a certain market power, but the free-entry condition limited profit opportunities to a certain extent. Firms produced differentiated goods using linear, constant return to scale technologies. Firms had heterogeneous productivity which evolved over time stochastically. Potential entrepreneurs had to pay certain costs to enter the market whereas incumbents needed to pay a fixed cost per unit of time to continue. All costs were measured by labor.

Active firms in the market were subject to firm-specific and permanent shocks to both productivity and demand for their differentiated commodities. In particular, the states of consumers' tastes and firm productivity were assumed to follow a geometric Brownian motion which had a

mean and variance growth rate independent of size. In the model, firms face with the non-linear cost of corruption and thus tend to stay small. Meanwhile, active firms in the market are subject to firm-specific, permanent shocks to both productivity and demand for their differentiated commodities. In particular, the states of consumers' tastes and firms' productivities are assumed to follow a geometric Brownian motion which has a mean and variance growth rate independent of size. Therefore, firms can achieve a bigger scale to medium or large-scale groups if they experience a large enough positive productivity shocks.
Relating to the case of Vietnam, different from other countries, Vietnam experienced a particular development path. Large-scale firms in Vietnam are mainly State-owned Enterprises which were established during the period that the economy followed the center planning mechanism[1]. Another source of large-scale firms is foreign invested enterprises (FIEs). Since the establishment of the Foreign Investment Law in 1987, there has been a surge in the number of foreign registered firms[2]. Among those, the most popular form of investment was joint-venture with the principle partners being SOEs (98% of joint ventures). These enterprises mainly operated in heavy industries or strategic sectors such as the extraction of oil and natural gas, chemicals and automotive with large scale production.

Later, these processes translated into the selection mechanism of firm productivity over time. This mechanism was crucial for the development of the stationary FSD in the economy. Technologies available to entrants were assumed to develop at an exogenous rate, which later determined the economic growth rate. A stationary FSD arose as productivity and demand shocks improved the productivity of existing firms at an average rate that was not too high, compared to those that was available among potential entrants.

In the model, we assumed that fixed cost represented the level of corruption and regulatory burdens that firms of different groups faced in the market. In contrast to Luttmer (2007), fixed cost in our model was a function of size. As fixed cost increased with the size of the firm, the value of the firm decreased and the exit barrier increased. This process decreased the size density for larger firms. As such, changes in fixed cost caused a change in the shape of the FSD given the variation was not uniform across different size groups in the FSD. In other words, the increase in fixed cost implied a smaller number of firms in the large-sized firm group, which

[1] After the war against French in 1954, the state sector of Vietnam was quickly formed by nationalizing exiting private enterprises and establishing new ones. By 1960, 100% of industrial establishments, 99.4% commercial firms and 99% of transport facilities were transformed into SOEs (Vu, 2002)

[2] By 1995, the registered foreign capital was over 2 billion USD annually (a higher level compared to those of other Asian neighbors like Thailand, South Korea or Taiwan in the same period).

was in accordance with empirical fact in many developing countries.

This study addressed a limited area of research within the domains of corruption, firm size, and distribution of firms. The importance of the findings lies in the asymmetry of the growth opportunities for small firms and large firms in a developing market. This work is one of the few to offer evidence for the inverse relationship between growth and initial firm size (i.e., violation of Gibrats law). Evidence for this violation is cited as justification for differential treatment for small businesses, an issue that has gained attention with both academics and policymakers.

6.2.3 Chapter 4- Empirical Analysis - Corruption and Productivity

The objective of Chapter 4 was to examine the FSD of a developing country, Vietnam and to explore the driving forces behind the skew distribution of firm size in Vietnam. In accordance with the theoretical framework developed in the previous chapter, we focused on two factors: corruption and productivity.

First, we sought to determine whether any evidence of the missing middle phenomenon existed in Vietnamese data. We presented a comprehensive analysis of the FSD in Vietnam in which the hypothesis of bimodality was not rejected. The bi-modality statistics for the overall sample of the manufacturing industry rejected the hypothesis of unimodality although the statistics for each sector indicated that firm size distribution in some sectors rejected bimodality. This finding confirmed the importance of examining the FSD and the underlying mechanism at the sectoral level.

Next, using firm-level data, we examined different explanations for the existence of the missing middle in Vietnam. We focused on the relationship between firm size and production efficiency, as well as on the relationship between firm size and the likelihood of paying bribes in Vietnam.

More particularly, we analyzed production efficiency in each sector of the manufacturing industry by using the frontier approach within a nonparametric model. Our analysis showed that in most of the manufacturing industries studied, middle-sized firms' production efficiencies tended to be lower than those of small-sized or large-sized firms. In addition, we found the scale at which minimum efficiency occurred as well as the level of this efficiency were different across industries. This finding showed the importance of studying the FSD and productivity at the sectoral level. Together with our result related to the corruption issue, this finding provided evidence that the middle-sized firms were disadvantaged in Vietnam.

Further, we analyzed the inefficiencies associated with scale for each firm in each sector. Our analysis indicated that large-sized firms and small-sized firms tended to produce closer to the production function, although, compared to the highest average products' firms, they tended to have lower average products. In contrast, the middle-sized firms tend to be further from the efficient production level, and had lower average products. Firms in manufacturing sectors of Vietnam experienced size inefficiencies in the sense that they can achieve higher average product if they increase their size. In other words, most firms choose to operate in the smaller than optimal scales in the analyzed period. One possible explanation is that there are certain barriers to development, for example, corruption or financial constraints and firms choose not to produce at the optimal scale or cannot achieve the optimal scale because of those burdens.
In addition, firms in Vietnam enter the market with different sizes. State-owned Enterprises or Foreign join-ventures tend to be big while private firms are small. Once entering the market, firms are facing with different burdens depending on their initial sizes. Small or micro firms can avoid attention of officials and thus pay a smaller corruption cost, compared to their medium and large peers. The large firms, on the other hand, can utilize effectively corrupted officials to avoid taxes and other legal requirements. As a result, medium-sized firms are those who are mostly affected by corruption.

To the best of our knowledge, this analysis was the very first work to connect the missing middle and production efficiencies across different firm sizes at the sectoral level in a developing country. Moreover, the results of the analysis indicated the necessity of sector-by-sector analyses to study the underlying mechanisms of the missing middle phenomena.

Next, we conducted a logit regression to study the relationship between firm size and the likelihood of paying bribes. In this analysis, we used the SME survey, which provided comprehensive information on all forms of small-sized and medium-sized firms. We analyzed the important factors involved in firms' bribing activities, in particular the relationship between firm size and the likelihood of paying bribes. Our results about corruption indicated that as firm size increased, the likelihood of paying bribes also increased. The formality of firms also mattered in the sense that the likelihood of paying bribes was higher for formal firms, which tended to be larger. We also analyzed the relationship between the likelihood of paying bribes and firm age. Our analysis showed that firms' ownership types and locations were also important determinants of bribing activities. Interestingly, our analysis also showed that firms perhaps needed to pay bribes at the beginning of operation, and if a firm paid bribes at the beginning, then the likelihood of paying bribes increased over time.

6.2.4 Chapter 5 - Productivity at the industry level

In the previous chapter, we investigated empirically the relationship between firm size and production efficiency at the firm level. The result, however, was based on the firm-level efficiency analysis. Because the missing middle in firm size distribution represented a macro phenomenon, we summarized the scores into a few efficiency scores at the industry level and investigated those in relation to firm size.

In Chapter 5, we provided an approximation theorem that encompassed all the different specifications of industry efficiency introduced in the literature and applied them to a same basic optimization program. This program was linear and could be solved using standard simplex methods. This introduced advantages in terms of both computation and interpretation. The computational advantage was obvious. The interpretational advantage came from the development of a simple, basic model to which we could refer for any type of aggregate analysis. This model served as a unifying tool for a body of literature that has become fragmented and dispersed on this topic. Essentially, we proved that the industry inefficiency was approximately the same if measured under different technology and in fact, the convexity and scale assumptions made on the firm-level technology became irrelevant when we considered the overall inefficiency of the industry. We also discussed the components of such an industry inefficiency measure as the sum of firm technical efficiency and the reallocation efficiency. Although the industry efficiency remained approximately unchanged if one changed the convexity and scale assumptions on the firm-level technology, its decomposition depended on the assumptions one was willing to make about the technology. In addition, for the sake of completeness, we also briefly considered the industry productivity change.

Finally, to investigate the relationship between aggregate efficiencies and the missing middle further, we extended the proposed framework to the meta-technology approach pioneered by Battese and Rao (2002), Battese et al. (2004)) and to the group-wise efficiency proposed in Nesterenko (2007). In particular, we segmented the input and output matrices into different sized groups and then, investigated the differences in efficiencies between size groups and the meta estimates. We found differences between the group-frontiers and the meta-frontiers. In addition,the meta-technology efficiencies (MTE) were highest for the small-sized firm group, in particular, for groups of firms with fewer than 50 employees. This confirmed the existence of differences in the characteristics of the firm's operating environments across size groups.

6.3 Limitations and future research questions

Before closing this thesis we note that this work presented some limitations. These limitations should not be regarded as failures but as opportunities to further knowledge about this topic. These limitation are summarized in the following paragraphs:

First, among the many possible factors that could affect the missing middle, the thesis focused on only two factors: productivity and corruption. Other plausible explanations for the missing middle phenomenon, such as the financial constraint, should be considered in future studies.

Second, because the missing middle is a common phenomenon discussed in the context of many developing countries, a similar analysis with developed economies is needed to highlight possible differences between the two types of economies. The analysis could shed light on the mechanism behind the missing middle. Moreover, similar analysis using other developing countries' data is crucial to generalize the conclusions of our study to settings in other developing countries. Such an exercise is still open because of the limited data currently available.

Third, related to the empirical study, particularly to the productivity analysis, as pointed out by Feng and Zhang (2014), choosing to exclude technological heterogeneity of an individual firm even within a single sector can cause bias in measuring efficiencies and returns to scale. Measuring returns to scale considering technological heterogeneity for each firm represents an interesting and promising research path to investigate further the impacts on the missing middle of inefficiencies associated with scale.

Fourth, relating the theoretical model, we note that the evolution of FSD in different industries is presumably more complex than the FSD evolution discussed in our study. In particular, many studies on the dynamics of the FSD have provided evidence on the complexity of forces that can affect the FSD (see Geroski, 1999, for a survey of growth of forces influencing in certain stages of the industry life cycle). In addition, the main force operating in the formulation of the FSD in our model - the decision of bribery was assumed to be exogenous to firms.

Taking into account all the above-mentioned limitations of the study, one outcome of this thesis is that we can propose the basis for an ambitious and diversified research agenda. Future research should include the following:

- Similar analyses to be conducted using the bribery data to directly investigate the linkage if any between bribery and the inefficiencies of firms.

- Further investigation on the source of inefficiencies at the firm level and the relationship between the inefficiencies and both the level and the growth rate of output.

- Similar analyses should be conducted on production efficiency using data from other developed and developing countries and compare the findings to the results for Vietnam. If lower productivity for middle-sized firms in developed economies is observed, researchers could then consider the causes. If we observe this U-shaped pattern only in developing countries, we could examine the reasons that this pattern exists in developing countries but not in developed countries.

- The causes of these differences in productivity should be examined across different firm sizes in developing countries. OECD Publishing (2013) reported two main reasons for differences in productivity across different firm sizes among OECD countries: (a) firm size matters for productivity, and (b) structural differences in the industrial composition of economies affect the relative performance of large and small firms across countries. It would be interesting to investigate whether these factors are also important in the productivity of developing nations.

- A more comprehensive theoretical framework should be developed that could generate a missing middle in FSD. In particular, a potential future study could involve endogenizing the bribery decision of firms in the interaction with governmental officials and political elites.

In general, the main contributions of this study are relevant both empirically and theoretically. Many questions raised in this thesis could help to improve knowledge about the Vietnamese case and could be extended to other developing countries. The analysis of firm size distribution in general and of the missing middle phenomenon in particular was a first step in an interesting field of research that deserves further investigation. The crucial implications of this study for policy, market structure and the economic evolution of regions are sufficient justifications for why this process should be taken into account in economic research.

Bibliography

A., Cooper W. W. Charnes and E. Rhodes, "Measuring the efficiency of decision making units," *European Journal of Operational Research*, 1978, pp. 429–444.

Acs, Zoltan J. and David B. Audretsch, "Innovation, marketstructures, and firm size," *Review of Economics and Statistics*, 1987, *69*, 567–74.

Aidt, T., J. Dutta, and V. Sena, "Governance regimes, corruption and growth: theory and evidence," *Journal of Comparative Economics*, 2008, *36*, 195–220.

Alfaro, L., S. Kalemli-Ozcan, and V. Volosovych, "Why doesn't capital flow from rich to poor countries? An empirical investigation," *Review of Economics and Statistics*, 2008, *90*, 347–368.

Alvaro, C., "Who cares about corruption?," *Journal of International Business Studies*, 2006, *37*, 803–822.

Anderson, J. H., G. Alcaide, D. Maria, and P. T. Tuyet, "Vietnam Development Report 2010: Modern Institutions," Technical Report, World Bank, Washington DC. 2010.

Angelini, P. and A. Generate, "On the evolution of firm size distribution," *American Economic Review*, 2008, *98*, 426–438.

Anh, Nguyen Thi Tue, Luu Minh Duc, and Trinh Duc Chieu, "The Evolution of Vietnamese Industry," *WIDER Working Paper*, 2014, *2014/76.*

Anti-Corruption Center, "Overview of corruption and anti-corruption in Vietnam," 2012.

Aparicio, J. and J. T. Pastor, "Directional distance function and Rate-of-Return Regulation," *Advances in Decision Sciences*, 2012, *doi:10.1155/2012/731497.*

__ , __ , **and S.C. Ray**, “An overall measure of technical ineffiiency at the firm and at the industry level: the “lost profit on outlay”,” *European Journal of Operational Research*, 2013, *226*, 154–162.

Arthur, W. B., *The self-reinforcing mechanism in economic systems, in: Anderson P. W., Arrow K. J., Pines D., The economy: as an evolving system*, Springer, Berlin, 1988.

Asmild, M., J. C. Paradi, and J. T. Pastor, “Centralized Resource Allocation BCC Models,” *Omega. The International Journal of Management Science*, 2009, *37* (1), 40–49.

__ , __ , **and** __ , “DEA based models for reallocations of police personnel,” *OR Spectrum*, 2012, *34* (4), 921–941.

Athanasouli, Daphne., Antoine Goujard, and Pantelis. Sklias, “Corruption and firm performance: evidence from Greek firms,” *International Journal of Economic Sciences and Applied Research*, 2012, *5*, 43–67.

Axtell, L., “Zipf Distribution of US firm sizes,” *Science*, 2001, *293*, 1818–1820.

Ayaydin, Hasan and Pinar Hayaloglu, “The effect of corruption on firm growth: Evidence from firms in Turkey,” *Asian Economic and Financial Review*, 2014, *4* (5), 607–624.

Bai, Jie, Seema Jayachandran, Edmund J. Malesky, and Benjamin A. Olken, “Does Economic Growth Reduce Corruption? Theory and Evidence from Vietnam,” NBER Working Papers 19483, National Bureau of Economic Research, Inc September 2013.

Bakhtiari, Sasan, “Firm Size Evolution and Outsourcing,” *The University of New South Wales Discussion Papers*, 2013, *07*.

Banerjee, Abhijit V. and Esther Duflo, “Growth Theory through the Lens of Development Economics,” in Philippe Aghion and Steven Durlauf, eds., *Handbook of Economic Growth*, Vol. 1a, Amsterdam: Elsevier, 2005, chapter 7, pp. 473–552.

Banker, R. D., A. Charnes, and W. W. Coope, “Models for the Estimation of Technical and Scale Inefficiencies in Data Envelopment Analysis,” *Management Science*, 1984, *30*, 1078–1092.

Bartelsman, E.J. and M. Doms, “Understanding Productivity: Lessons from Longitudinal Microdata,” *Journal of Economic Literature*, 2000, *38*, 569–594.

Basu, S. and J. G. Fernald, "Returns to Scale in U.S. Production: Estimates and Implications," *Journal of Political Economy*, 1997, *105* (2), 249–283.

Battese, G.E. and D.S.P. Rao, "Technology potential, efficiency and a stochastic metafrontier function," *International Journal of Business Economics*, 2002, *1*, 17.

_ , _ , and C.J. O'Donnell, "A metafrontier production function for estimation of technical efficiencies and technology potentials for firms operating under different technologies," *Journal of Productivity Analysis*, 2004, *21*, 91103.

Baumol, W. J., J. C. Panzar, and R. D. Willig, "Contestable markets and the theory of industrial structure, rev. ed.," *American Economics Review*, 1983, *73* (3).

_ , _ , and _ , *Contestable markets: an uprising in the theory of industry structure: repply*, San Diego: Harcourt Brace Jovanovich, Academic Press, 1988.

Beck, T., A. Demirguc-Kunt, and V. Maksimovic, "Financial and Legal Constraints to Firm Growth: Does Size Matter," *Journal of Finance*, 2005, *60* (1).

Bentzen, J., E. S. Madsen, and V. Smith, "Do firms' growth rates depend on firm size?," *Small Business Economics*, 2011, *39* (4), 937–47.

Boerner, S. and T. Macher, "Transaction cost economics: An assessment of empirical research in the social sciences," *Working paper*, 2002.

Bogetoft, P. and D. Wang, "Estimating the Potential Gains from Mergers," *Journal of Productivity Analysis*, 2005, *23*, 45–171.

Brandt, Urs and Gert Svendsen, "Why does bureaucratic corruption occur in the EU?," *Public Choice*, 2013, *157* (3), 585–599.

Briec, W., B. Dervaux, and H. Leleu, "Aggregation of directional distance functions and industrial efficiency," *Journal of Economics*, 2003, *79*, 237–261.

Brown, D. J., "Equilibrium analysis with non-convex technologies," *Handbook of Mathematical Economics*, 1991, *4.*

Brunetti, A., K. Gregory, and W. Beatrice, "Institutional obstacles to doing business: Region - by - region results from a worldwide survey of the private sector," *World Bank working paper*, 1997, *1759.*

Cabral, L. M. B. and J. Mata, "On the evolution of the firm size distribution: Facts and theory," *American Economic Review*, 2003, *93*, 1075–1090.

Cai, Wenbiao and Pandey Manish, "Size-Dependent Regulations and Structural Transformation in India," *Economics Letters*, 2013, *119*, 272–275.

Calvo, J. L., "Testing Gibrat's law for small, young and innovating firms," *Small Business Economics*, 2006, *26*, 117–123.

Carrizosa, M. and S. B. Agusti, "Productivity growth and competition in Spanish manufacturing firms: What has happened in recent years?," *Working paper CREAP 2006-09*, 2006.

Carrizosa, M. T., "Firm growth, persistence and multiplicity of equilibra: An analysis of Spanish manufacturing and service industries," *Rovira I Virgili University*, 2007.

Cefis, E., O. Marsili, and H. Schenk, "The effects of mergers and acquisitions on firm size distribution," *Journal of Evolutionary Economics*, 2008, *19*, 1–20.

Centre for Community Support and Development Studies, Centre for Research and Training of the Viet Nam Fatherland Front, and United Nations Development Programme, "The Viet Nam Governance and Public Administration Performance Index (PAPI) 2014: Measuring Citizens' Experiences," 2015.

Chambers, R. G., Y. Chung, and R. Färe, "Benefit and distance functions," *Journal of Economic Theory*, 1996, *70*, 407–419.

_ , _ , and _ , "Profit, directional distance functions and Nerlovian efficiency," *Journal of Optimization Theory and Applications*, 1998, *98*, 351–364.

Chowdhury, M., M. Azam, and S. Islam, "Problems and Prospects of SME Financing in Bangladesh," *Asian Business Review*, 2013, *2*, 1–8.

CIEM, *Vietnam's economy in 2005*, Hanoi: Political Publishing House of Political Theory, 2006.

Cirillo, P. and J. Husler, "On the upper tail of Italian firms' size distribution," *Physica A: Statistical Mechanics and its Applications*, 2009, *399*, 1546–1554.

Coad, A., *The growth of firms: A survey of theories and empirical evidence*, Cheltenham, UK: Edward Elgar, 2009.

_ , "The Exponential Age Distribution and the Pareto Firm size Distribution," *Journal of Industry, Competition and Trade*, 2010, *10*, 389–395.

Crosato, L., S. Destefanis, and P. Ganugi, "Firms size distribution and returns to scale. Non-parametric frontier estimates from Italian manufacturing," in A.M. Ferragina, E. Taymaz, and K. Yilmaz, eds., *Innovation, globalization and firm dynamics lessons for enterprise policy*, Rouledge, London, 2009, pp. 71–94.

Dabla-Norris, Era, Mark Gradstein, and Gabriela Inchauste, "What causes firms to hide output? The determinants of informality," *Journal of Development Economics*, 2008, *85*, 1–27.

Dalton, Sherry, Erik Friesenhahn, James Spletzer, and David Talan, "Employment Growth by Size Class: Firm and Establishment Data," *Monthly Labor Review*, 2011, *134*, 3–24.

Daraio, Cinzia and Léopold Simar, *Advanced Robust and Nonparametric Methods in Efficiency Analysis*, New York: Springer US, 2007.

Dasgupta, K., "The Missing Middle in Developing Countries Revisited," *University of Toronto*, 2010.

Dhammika, D., J. Slemrod, and J.D. Wilson, "Tax Policy and the Missing middle: Optimal Tax Remittance with Firm level Administrative Costs," *Working paper*, 2010.

Diaz, M. A. and R. Sanchez, "Firm size and productivity in Spain: a stochastic frontier analysis," *Small Bus Econ*, 2008, *30*, 315–323.

Diewert, W. Erwin, Takanobu Nakajima, Alice Nakamura, Emi Nakamura, and Masao Nakamura, "Returns to scale: concept, estimation and analysis of Japan's turbulent 1964-88 economy," *Canadian Journal of Economics*, 2011, *44* (2), 451–485.

Doraszelski, Ulrich, "Measuring returns to scale in nineteenth-century French industry," *Explorations in Economic History*, 2004, *41*, 256–281.

Dosi, G., "Statistical regularities in the evolution of industries, a guidance through some evidences and challenges for the theory," *The Cambridge University Press*, 2007.

Dutta, Nabamita, Saibal Kar, and Sanjukta Roy, "Corruption and persistent informality: An empirical investigation for India," *International Review of Economics and Finance*, 2013, *27*, 357–373.

Eifert, Benjamin, “Infrastructure and Market Structure in Least-Developed Countries,” *University of California Working paper*, 2007.

Emerson, M. Patrick, “Corruption and Industrial dualism in less developed countries,” *The Journal of International Trade and Economic Development: An international and Comparative Review*, 2001, *11*, 63–76.

Ernesto, Dal B and Martin A. Rossi, “Corruption and Inefficiency: Theory and Evidence from Electrical Utilities,” *Journal of Public Economics*, 2007, *91* (5-6), 939–962.

Evan, D. S., “Tests of Alternative theories of firm growth,” *The Journal of Political Economics*, 1987, *95*, 657–674.

Fang, L., “A generalized DEA model for centralized resource allocation,” *European Journal of Operational Research*, 2013, *228* (2), 405–412.

_ **and C. Q. Zhang**, “Resource Allocation Based on the DEA Model,” *Journal of the Operational Research Society*, 2008, *59*, 1136–1141.

Färe, Rolf, “Addition and Efficiency,” *The Quarterly Journal of Economics*, 1986, *101* (4), 861–866.

_ **and D. Primont**, “Luenberger productivity indicators: aggregation across firms,” *Journal of Productivity Analysis*, 2003, *20*, 425–435.

_ **and V. Zelenyuk**, “On aggregate Farrell efficiencies,” *European Journal of Operational Research,*, 2003, *146*, 615–620.

_ **, S. Grosskopf, and V. Zelenyuk**, “Aggregation of Nerlovian profit indicator,” *Applied Economics Letters*, 2008, *15*, 845–847.

Farrell, M. J., “The measurement of productive efficiency,” *Journal of the Royal Statistic Society*, 1957, *120*, 253–282.

Feng, Guohua and Xiaohui Zhang, “Returns to scale at large banks in the US: A random coefficient stochastic frontier approach,” *Journal of Banking & Finance*, 2014, *39* (C), 135–145.

Freeman, Jonathan B. and Rick Dale, “Assessing bimodality to detect the presence of a dual cognitive process,” *Behavioral Research Methods*, 2013, *45* (1), 83–97.

Fujiwara, Y., "Do Pareto - Zipf and Gibrat law hold true? An Analysis with European firms," *Statistical Mechanics and its Applications*, 2004, *335*, 197–216.

Gainsborough, Martin, "Corruption and the Politics of Decentralization in Vietnam," *Journal of Contemporary Asia*, 2006, *33*, 69–84.

Garicano, L. and E. R. Hansberg, "Inequality and the organization of knowledge," *American Economic Review*, 2004, *94*, 197–202.

Geroski, P. A, "The growth of firms in theory and in practice," *CEPR Discussion Papers*, 1999, *1986.*

Geroski, P. A. and K. Gugler, "Corporate growth convergence in Europe," *Oxford Economic Papers*, 2004, *56*, 597–620.

_ , S. Lazarova, G. Urga, and C. F. Walter, "Are differences in firm size transitory or permanent?," *Journal of Applied Econometrics*, 2003, *18*, 47–59.

Gerrit, D.W., "Firm Size Distribution, an Overview of Steady State Distribution Resulting from Firm Dynamic Models," *International Journal of Industrial Organization*, 2005, pp. 423–450.

Gibbons, R., "For formal(izable) theories of the firm?," *journal of Economic Behavior and Organization*, 2005, *58*, 200–245.

_ , "For formal(izable) theories of the firm?," *EWU Institutional Repository Working papers*, 2011.

Gibrat, R., "Les Inegalities Ecomiques," *Sirey, Paris*, 1931.

Giménez-García, V. M., J. L. Martínez-Parra, and F. P. Buffa, "Improving Resource Utilization in Multi-Unit Networked Organizations: The Case of a Spanish Restaurant Chain," *Tourism Management*, 2007, *28* (1), 262–270.

Giovanni, D. J. and Andrei A. Levchenko, "Putting the Parts Together: Trade, Vertical Linkages, and Business Cycle Co-movement," Working Papers, Research Seminar in International Economics, University of Michigan 2008.

Giovanni, G. and J. Goyette, "Corruption and the Size Distribution of Firms," *University of British Columbia, Working paper*, 2009.

Groenendijk, G. M., "A Principal-Agent Model of Corruption," *Crime, Law and Social Change*, 1997, *27*, 217–218.

Grossman, S. and O. Hart, "The costs and benefits of ownership: A theory of vertical and lateral integration," *Journal of Political Economy*, 1986, *94*, 681–719.

Guner, Nezih, Gustavo Ventura, and Yi Xud, "Macroeconomic implications of size-dependent policies," *Review of Economic Dynamics*, 2008, *11*, 724–744.

Gustavasson, P., "The dynamics of European industrial structure," *FIEF working pper series*, 2002, *176.*

Hannan, M. T. and J. Freeman, "The population ecology of organizations," *The American Journal of Sociology*, 1977.

Hansberg, R. and M. L. J. Wright, "Establishment size dynamics in the aggregate economy," *American Economic Review*, 2007, *97*, 1639–1666.

Harrell, Frank E. Jr, *Regression Modeling Strategies Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*, New York: Springer, 2001.

_ , "Regression Modeling Strategies," Technical Report 2014.

Harrison, J., *Brownian Motion and Stochastic Flow Systems*, New York: Wiley, 1985.

Hart, O. and J. Moore, "Property rights and the nature of firm," *Journal of Political economy*, 1990, *98*, 1119–1158.

Hart, P. and S. Prais, "The Analysis of Business Concentration: A Statistical Approach," *Journal of the Royal Statistical Society*, 1956, *119*, 50–191.

Hart, P. E. and N. Oulton, "The growth and size of firms," *Economic Journal*, 1996, *106*, 1242–1252.

Hartigan, J.A. and P. M. Hartigan, "The Dip Test of Unimodality," *The Annals of Statistics*, 1985, *13* (1), 70–84.

Hayami, Y., "On the Use of the Cobb-Douglas Production Function on the Cross-Section Analysis of Agricultural Production," *American Journal of Agricultural Economic*, 1969, *52* (2), 327–329.

_ **and V. W. Ruttan**, “Agricultural Productivity Differences among Countries,” *American Economic Review*, 1970, *60* (5), 895–911.

He, Zhenke, “Corruption and Anti-Corruption in Reform China,” *Communist and Post Communist Studies*, 2000, *33*, 250–272.

Hill, B. M., “A simple general approach to inference about the tail of a distribution,” *The Annals of Statistics*, 1975, *5*, 1163–1174.

Holmstrom, M., “The firm as a sub-economy,” *Journal of Law, Economics and Organization*, 1999, *15*, 74–102.

_ **and P. Milgrom**, “Multi-task principle - agent analysis: Linear contracts, asset ownership and job design,” *Journal of Law, Economics and Organization*, 1991, *7*, 24–52.

Ijiri, Y. and H. Simon, “Interpretations of Departures from the Pareto Curve Firm-Size Distributions,” *Journal of Political Economy*, 1974, *82*, 315–331.

Ingram, M., V. Ramachandran, and V. Desai, “Why do firms choose to be informal? Evidence from enterprise surveys in Africa,” *Mimeo, World Bank, Washington, DC*, 2007.

Ishikawa, A., “The uniqueness of firm size distribution function from ten-shaped growth rate distribution,” *Physica*, 2007, pp. 79–84.

Jacques, M. and J. Jordi, “Panel-data estimates of the production function and the revenue function: What differences does it make?,” *Scandinavian Journal of Economics*, 2005, *107*, 651–672.

Jain, A. K., “Corruption: a review,” *Journal of Economic Surveys*, 2005, *15* (1), 71–121.

Johannes, V. B., “Firm size matters: growth and productivity growth in African manufacturing,” *University of Chicago Press*, 2005, *53* (3), 651–672.

Johnson, S., D. Kaufmann, J. McMillan, and C. Woodruff, “Why do firms hide? Bribes and unofficial activity after communism,” *Journal of Public Economics*, 2000, *76*, 495–520.

Jovanovic, B., “Selection and the evolutionary of Industry,” *Econometric Society*, 1982, pp. 649–470.

Kalecki, M., “On the Gibrat Distribution,” *Econometrica*, 1945, *13*, 161–170.

Kalirajan, K.P. and R.T. Shand, "Modelling and measuring economic efficiency under risk," *Indian Journal of Agricultural Economics*, 1994, *49*, 579–90.

Kendall, M. G., "A New Measure of Rank Correlation," *Biometrica*, 1938, *30* (1/2), 81 – 93.

Kirchhoff and E. Norton, "Schumpeterian Creative Destruction versus Gibrat's law: An Examination of the Relationship between Firm Size and Growth," *ICSB World Conference*, 1995.

Klette, K. and S. Kortum, "Innovation Firms and Aggregate Innovation," *Journal of Political Economy*, 2002, *112*, 586–1018.

Klette, T. and Z. Griliches, "The inconsistency of common scale estimators when output prices are unobserved and endogenous," *Journal of Applied Econometric*, 1996, *11*, 343–361.

Knack, S. and P. Keefer, "Institutions and economic performance: cross-country tests using alternative institutional measures.," *Economics and Politics*, 1995, *7*, 207–227.

Krugman, P. R., "The Self-Organizing Economy," *Cambridge, MA*, 1996.

Kruskal, William H., "Ordinal Measures of Association," *Journal of of the American Statistical Association*, 1958, *53* (284), 814–8 61.

Kwok, C. and S. Tadesse, "The MNC as an agent of change for host-country institutions: FDI and corruption," *Journal of International Business Studies*, 2006, *37*, 767–785.

Leung, D., C. Meh, and Y. Terajima, "Productivity in Canada: does firm size matter?," *Bank of Canada Review*, 2008, *1*.

Li, S. K. and Y. C. Ng, "Measuring the productive efficiency of a group of firms," *International Advances in Economic Research*, 1995, *1*, 377–390.

Liedholm, Carl and Donald C. Mead, "Small Scale Industries in Developing Countries: Empirical Evidence and Policy Implications," Technical Report, Michigan State University, Department of Agricultural, Food, and Resource Economics 1987.

Little, I., M. Dipak, and P. John, "Small manufacturing enterprises: a comperative analysis of India and other economies," Technical Report 10118, A World Bank research publication 1987.

Lotfi, F. H., A. A. Noora, G. R. Johanshahloo, J. Gerami, and M. R. Mozaffari, “Centralized resource allocation for enhanced Russell models,” *Journal of Computational and Applied Mathematics*, 2010, *235*, 1–10.

Lozano, S. and G. Villa, “Centralized Resource Allocation Using Data Envelopment Analysis,” *Journal of Productivity Analysis*, 2004, *22* (1-2), 143–161.

_ and _ , “Centralized DEA Models with the Possibility of Downsizing,” *Journal of the Operational Research Society*, 2005, *56*, 357–364.

_ , _ , and B. Adenso-Díaz, “Centralized Target Setting for Regional Recycling Operations Using DEA,” *Omega, The International Journal of Management Science*, 2004, *32* (2), 101–110.

_ , _ , and D. Canca, “Application of Centralized DEA Approach to Capital Budgeting in Spanish Ports,” *Computers & Industrial Engineering*, 2011, *60* (3), 455–465.

_ , _ , and R. Braennlund, “Centralized Reallocation of Emission Permits using DEA,” *European Journal of Operational Research*, 2009, *193* (3), 752–760.

Luttmer, E. G. J, “Selection, Growth and the Size Distribution of Firms,” *Quarterly Journal of Economics*, 2007, *122*, 1103–1068.

_ , “On the Mechanism of Firm Growth,” *Review of Economic Studies*, 2010, *78*, 1042–1068.

Maitland, E., “Corruption and the Outsiders: MNEs in Vietnam,” *Federation Press*, 2012, pp. 147–166.

Mansfield, E., “Entry, Gibrats Law, Innovation, and the Growth of Firms,” *American Economic Review*, 1962, *52*, 1023–1051.

Mar-Molinero, C., D. Prior, M. M. Segovia, and F. Portillo, “On Centralized Resource Utilization and its Reallocation by using DEA,” *Annals of Operations Research*, 2012, *221*, 273–283, DOI: 10.1007/s10479–012–1083–8.

Marchiavello, R., “Vertical Integration and Investor Protection in Developing Countries,” *Journal of Development Economic*, 2010, *93*, 1162–172.

Marsili, O., “Technology and the size distribution of firms: Evidence from Dutch manufacturing,” *Review of Industrial Organization*, 2005, *27*, 303–328.

Masten, E., “Econometrics of contracts: An assessment of recent developments in the empirical literature on contracting,” *Revue d’Economie Industrielle*, 2000, *92*, 215–236.

Méndez, Fabio and Facundo Sepúlveda, “Corruption, growth and political regimes: Cross country evidence,” *European Journal of Political Economy*, 2006, *22*, 82–98.

M‘eon, P. and K. Sekkat, “Does corruption grease or sand the wheels of growth?,” *Public Choice*, 2005, *122*, 69–97.

Miller, Stephen M. and Athanasios G. Noulas, “The technical efficiency of large bank production,” *Journal of Banking & Finance*, 1996, *20*, 495–509.

Mishra, Ajit and Ranjan Ray, “Informality and Corruption,” 2013. mimeo.

Mo, P., “Corruption and economic growth,” *Journal of Comparative Economics*, 2001, *29*, 66–97.

Mo, Pak Hung, “Corruption and Economic Growth,” *Journal of Comparative Economics*, 2001, *29*, 66–79.

Neeman, Z., D. Paserman, and A. Simhon, “Corruption and openness,” *The B.E. Journal of Economic Analysis & Policy*, 2008, *8* (1), 1–38.

Nelson, R. and G. Winter, “An evolutionary theory of economic change,” *Havard University Press*, 1982.

Newman, M. E. J., “Power laws, Pareto distribution and Zipf’s law,” *Contemporary Physics*, 2004, *46*, 323–351.

Nowak, Robert, “Corruption and Transitional Economies,” 2001. Presented at the Ninth OSCE Economic Forum.

O’Donnell, C. J., D. S. P. Rao, and G. E. Battese, “Metafrontier Frameworks for the Study of Firm-level Efficiencies and Technology Ratios,” *Empirical Economics*, 2008, *34* (2), 231–255.

OECD Publishing, “Enterpreneurship at a glance 2013,” 2013. OECD.

Okuyama, K., “Zipf’s law in Income Distribution of Companies,” *Physics A*, 1999, *269*, 125–131.

Oliveire, B. and A. Fortunato, "Does firm size matter for investment and R&D? Evidence from a panel data of manufacturing firms in the Portuguese economy," *Economics working paper*, 2004, *11.*

_ **and** _ , "Testing Gibrat's law: Empirical evidence from a panel of Portuguese manufacturing firms in the Portuguese economy," *International Journal of the Economics of Business*, 2006, *13*, 65–81.

Penrose, Edith. T., *The Theory of the Growth of the Firm*, Oxford University Press, 1959.

Peyrache, A., "Structural inefficiency and potential gains from mergers and break-ups: a Comprehensive approach," *European Journal of Operational Research*, 2013, *230*, 422–430.

_ , "Cost Constrained Industry Inefficiency," *European Journal of Operational Research, doi:10.1016/j.ejor.2015.06.067*, 2015.

Quandt, R. E., "On the size distribution of firms," *American Economic Review*, 1966, *3*, 416–432.

Ramey, V. A., "Non-convex costs and the behavior of inventories," *The Journal of Political Economy*, 1991, *99*, 45–61.

Rand, J. and F. Tarp, "Characteristics of the Vietnamese business environment: evidence from a SME survey in 2005," *Business sector program Support*, 2012.

Rausch, J. E., "Modeling the informal sector formally," *Journal of Development Economics*, 1991, *35*, 33–47.

Ray, S. C., *Data Envelopment Analysis*, Cambridge, United Kingdom: Cambridge University Press, 2004.

_ , "Are some Indian banks too large? An examination of size efficiency in Indian banking," *Journal of Productivity Analysis*, 2007, *27*, 41–56.

_ **and K. Mukherjee**, "Quantity, quality and efficiency for a partially superadditive cost function: Connecticut public schools revisited," *Journal of Productivity Analysis*, 1998, *10*, 47–62.

_ **and X. Hu**, "On the technically efficient organization of an industry: a study of U.S. airlines," *Journal of Productivity Analysis*, 1997, *8*, 5–18.

— **, L. Chen, and K. Mukherjee**, "Input price variation across locations and generalized measure of cost effciency," *International Journal of Production Economics*, 2008, *116*, 208–218.

Richardson, G. B., "The organization of industry," *The Economic Journal*, 1972, *82*, 883–896.

Robinson, E., "The structure of competitive industry," *University of Chicago Press*, 1958.

Romer, Paul, "New Goods, Old Theory and the Welfare Costs of Trade Restrictions," *Development Economic Journal*, 1994, *43*, 5–38.

Rosa, Donato De, Nishaal Gooroochurn, and Holger Gorg, "Corruption and Productivity: Firm-level Evidence," *Policy Research Working Paper Series, The World Bank*, 2010, *5348.*

Santarelli, E., L. Klomp, and A. R. Thurik, *Gibrat's law: An overview of the Empirical literature*, in E. Santarelli (Ed.), Entrepreneurship, Growth and Innovation: the dynamics of firms and industries, Springer, New York, 2006.

SAS Institute Inc., "SAS/STAT user's guide 9.2," 2008. Cary, NC: SAS Institute.

Saussier, S., "Transaction costs and contractual incompleteness: the case of Electricite de France," *Journal of Economic Behavior and Organization*, 2000, *42.*

Schiffer, M. and B. Weder, "Firm Size and the Business Environment: Worldwide survey Results," *World Bank Publication*, 2001, *85.*

Schneider, F. and D. Enste, "The Shadow Economy: Theoretical Approaches, Empirical Studies, and Political Implications," *Cambridge University Press*, 2002.

Sequeira, . and S. Djankov, "An Empirical Study of Corruption in Ports," *MPRA Paper, University Library of Munich, Germany*, 2010, *21791.*

Shapiro, S. S. and M. B. Wilk, "An Analysis of Variance Test for Normality (Complete Samples)," *Biometrika*, 1965, *52* (3/4), 591–611.

Simar, L. and V. Zelenyuk, "Statistical inference for aggregates of Farrell-type efficiencies," *Journal of Applied Econometrics*, 2007, *22*, 1367–1394.

Simon, A. H. and C. P. Bonini, "The Size Distribution of Business Firms," *American Economic Review*, 1958, *4*, 607–617.

Simon, H. A., "On a Class of Skew Distribution Functions," *Biometrika*, 1955, *52*, 425–440.

_ , “Some Further Notes on a Class of Skew Distribution Functions,” *Information and Control*, 1960, *3*, 80–88.

Singh, A., “Take-overs, economic natural selection and the theory of the firm: evidence from the post-war United Kingdom experience,” *Economic Journal*, 1975, *85.*

Sleuwaegen, L. and M. Goedhuys, “Growth of Firms in Developing Countries, Evidence from Cte dIvoire.,” *Journal of Development Economics,*, 2002, *68*, 117–135.

Socialist Republic of Vietnam, “Law on Cooperatives,” National Assembly 2003.

_ , “Law on Enterprise,” National Assembly 2005.

Soleimani-Damaneh, M. and A. Mostafaee, “Stability of the classification of returns to scale in FDH models,” *European Journal of Operational Research*, 2009, *196*, 1223–1228.

_ **and M. Reshadi**, “A polynomial-time algorithm to estimate returns to scale in FDH models,” *Computer and Operations Research*, 2007, *34*, 2168–2176.

Somers, Robert H., “A New Asymmetric Measure of Association for Ordinal Variables,” *American Sociological Review*, 1962, *27* (6), 799–811.

Steel, W. F. and L. M. Webster, “How Small Enterprises in Ghana Have Responded to Adjustment,” *World Bank Economic Review*, 1992, *6*, 423–438.

Steindl, J., *Random Processes and the Growth of Firms*, London: Hafner, 1965.

Storey, D. J., *Understanding the Small Business Sector*, London: Routledge, 1994.

Suton, J., “Gibrat’s legacy,” *Journal of Economic Literature*, 1977, *35*, 40–59.

Svensson, J., “Who Must Pay Bribes and How Much? Evidence from A Cross Section Of Firms,” *The Quarterly Journal of Economics*, 2003, *118* (1), 207–230.

Tai Hsieh, Chang and Benjamin A. Olken, “The Missing “Missing Middle”,” *Journal of Economic Perspectives*, 2014, *28* (3), 89–108.

Talvitie, A., “Obsrved differences in corruption between Asiaand Africa: the industrial organization of corruption and its cure,” *IPMR*, 2012, *13*, 112–141.

Tanaka, K. and N. Hatsukano, "The size Distribution of all Cambodia Establishments," *IDE Discussion Paper*, 2010, *253.*

Tao, T., "Benford's law, Zipf's law, and the Pareto distribution," *http://terrytao.wordpress.com/2009/07/03/benfords-law-zipfs-law-and-the-pareto-distribution/*, 2009.

Taylor, H. and S. Karlin, *An Introduction to Stochastic Modeling*, California: Academic Press, 1998.

Thang, D.C, MBIS, and MSc, "APEC Procurement transparency standards in Vietnam: the need to move from law to practice," 2011. Transparency International-USA and Center for International Private Enterprise and Towards Transparency.

Transparancy International, *Global Corruption Report 2007: Corruption and judicial systems*, New York: Cambridge University Press, 2007.

Tullock, G., *Rent seeking*, U.K., Hants: Edward Elgar: The Shaftesbury papers, 1993.

Tybout, James R., "Manufacturing firms in Developing Countries: How Well Do They and Why?," *Journal of Economic Literature*, 2000, *38*, 11–44.

_ , "Correspondence," *Journal of Economic Perspectives*, 2014, *28* (4), 235–236.

_ , "The Missing Middle, Revisited," 2014. mimeo.

Wagener, F. O. O., "Incresing returns, structura; analysis of optimal investment for firms with non-concave production," *Working paper, http//www1.fee.uva.nl/cendef/publications/papers/preprint0308.pdf.*, 2003.

Wedeman, A., "Development and corruption: the East Asian Paradox," *Political Business in East Asia*, 2002.

Williamson, O., *Markets and hierarchies: Analysis and antitrust implication*, New York: Collier MacMillan, 1975.

Wilson, Paul W., "FEAR 1.0: A Software Package for Frontier Efficiency Analysis with R," *Socio-Economic Planning Sciences*, 2008, *42*, 247–254.

Witt, Ulrich, "Imagination and leadership - The neglected dimension of an evolutionary theory of the firm," *Journal of Economic Behavior and Organization*, 1998, *35*, 161–177.

Wu, Xun, "Determinants of Bribery in Asian Firms: Evidence from the World Business Environment Survey," *Journal of Business Ethics*, 2009, *87*, 75–88.

Zelenyuk, V., "Aggregation of Malmquist productivity indexes," *European Journal of Operational Research*, 2006, *174*, 1076–1086.

CHAPTER 7

APPENDIX

Gibrat's law

In 1931, a French engineer, Robert Gibrat observed that French manufacturing firms exhibited a highly right-skew distribution. Such an observation leads to a simple dynamic model predicting that all firms grow at a same proportionate rate, irrespective of their initial sizes. Such a law is later known as Gibrat's law or the law of Proportionate Effect. In general, the Law of Proportionate Effect can be derived from a simple dynamic model in the discrete time. Let s_t be the size of a firm at time t, the growth rate of a firm is:

$$\frac{s_t - s_{t-1}}{s_{t-1}} = \varepsilon_t. \quad (1)$$

Solving for s_t by using the recursion formula we have:

$$s_t = (1 + \varepsilon_t)(1 + \varepsilon_{t-1})...(1 + \varepsilon_1)s_0. \quad (2)$$

Taking logarithm of both size:

$$\begin{aligned} \log s_t &= \varepsilon_t + \varepsilon t - 1 + ... + \varepsilon_1 + \log s_0 \\ &= \sum_{i=1}^{t} \varepsilon_i + \log s_0. \end{aligned} \quad (3)$$

We see that the logarithm of firm size at time t is just the accumulation of random shocks ε and the initial condition s_0. These unexpected shocks occur as we do not know what will happen and when it will happen. As a result, the process is unpredictable. In addition, these shocks are assumed to have permanent impacts on firm size in Gibrat's model.

From equation (3) and according to the central limit theorem, the logarithm of firm size, once normalized, can be approximated by the normal distribution since random shocks follow the normal distribution as time goes to infinity. Gibrat's law, thus, implies that firm size follows a random walk. In other words, firms with initial endowments (human resource, technology, financial capacity, etc.) evolve over time through the accumulation of independent draws from a Gaussian distribution.

As summarized in Carrizosa (2007), Gibrat's law implies the identically independent distribution (i.i.d.) of the sequence of random shocks which have finite means and variances. In particular, Gibrat's law also implies:

- the independence of firm growth rate from the initial size or the expected firm growth rate and variance are similar irrespective of the initial size;
- no serial correlation as the previous growth rate has no impacts on current growth;
- the increase with time of firm size diversification or the market concentration will increase if the number of firms is constant;
- the equalization of variances of firm growth rates as small firms grow as fast as large firms;
- no optimum size of firms and thus no stationary distribution and the variance of size approaches infinity as time goes to infinity.

In addition, the continuous version of Gibrats law is closely connected to the *Geometric Brownian motion* (GBM).

To see this, we have the logarithm of firm size, once normalized by the initial value, implies a Wiener process with drift and the size is proportional to the exponential of Wiener process with drift - the GBM, which has the form of:

$$s(t) = s_0 e^{at+bW(t)}, \tag{4}$$

where $s(t)$ is the size of firm at time t, s_0 is the initial size and $e^{at+bW(t)}$ is the GBM which does not depend on the initial value of firms.

Another way to see the relationship between Gibrat's law and the GBM is to consider the stochastic differential equation of the Wiener process with drift. Using the Ito formula, we have:

$$\mathrm{d}s(t) = a.s(t)\mathrm{d}t + b.s(t)\mathrm{d}W(t). \tag{5}$$

Then the growth rate of a firm is:

$$\frac{\mathrm{d}s(t)}{S(t)} = a.\mathrm{d}t + b.\mathrm{d}W(t), \tag{6}$$

which is independent of current firm size.

Power law, the Pareto distribution and Zipf's law

Power law is a profound property of the size distribution[1]. If variable s_i follows Power law, let $p(s_i)ds$ be the fraction between s_i and $s_i + ds$, then the histogram in log-log scale is a straight line and we have:

$$\log p(s_i) = -\alpha \log s_i + c, \tag{7}$$

or

$$p(s_i) = C s_i^{-\alpha}, \tag{8}$$

where α is the exponent of the power law and the constant C is of minor importance because once α is fixed, C can be determined by normalizing $p(s_i)$.

A number of properties of the power law can be listed here.

- In the log-log scale, the cdf of the law can be approximated by a straight line;
- The tail index α of the distribution implies the level of market concentration, which represents the relative difference in size among firms (Simon and Bonini, 1958).
- Power law is scale-independent. The industrial structure, if following the power law, depends only on the interaction among its components rather than any external effect or individual behaviors.

[1]Power law governs the asymptotic distributions of many statistics which (i) take values as positive numbers; (ii) range over many different orders of magnitude; (iii) arise from a complicated combination of largely independent factors; and (iv) have not been artificially rounded, truncated, or otherwise constrained in size (Tao, 2009).

- The slope of power law is relatively stable over time, which in many cases, has the value of unity in the rank - size scale.

In the literature, there are various mechanisms to generate the power law for a distribution. As summarized in Newman (2004), these include, but not exclusively, the combination of exponentials, the inverses of quantities, the random walk, the Yule process, the phase transitions and critical phenomena or the self organized criticality, among others.

Next, we represent two alternatives of looking at the power law: the Pareto distribution and Zipf's law.

First, Pareto distribution is defined for a continuous variable. It is the continuous version of the power law[2]. The Pareto law is normally expressed in terms of a cdf function.

Considering an economy with n firms, each has s employees, we start by asking "What is the size of the i^{th} largest firm? The Pareto distribution says, the number of firms with size greater than or equal to s follows:

$$i = Cs_i^{-(1/\alpha)}, \tag{9}$$

where i is the frequency or the rank of size.

On the other hand, Zipf law - named after George Kingsley Zipf, an American linguist- can be viewed as a discrete Pareto distribution, a "flipped" counterpart of the Pareto distribution. Zipfs Law appears when the Pareto distribution exponent is equal to unity. Zipf's law implies that the probability of a firm size being greater than a given level decays with the growth of size and the size s of the i^th largest firm follows:

$$s = C_i^{-\alpha}. \tag{10}$$

This is exactly the definition of the Pareto distribution, except the two axes are flipped. In a plot, Zipf's law plots the relevant variables on the horizontal axis and the cdf on the vertical one while the Pareto distribution does vice versa.

[2]The distribution is named after Vilfredo Pareto, an Italian economist, who first used this distribution to describe the distribution of income.

The derivation of Eq. (11)

Proof. We have:

$$R_{t,a} - w_{t+a}L_{t,a} = (\frac{\beta Z_{t,a}}{w_{t+a}})^{\beta/(1-\beta)}C_{t+a} - \beta(\frac{\beta Z_{t,a}}{w_{t+a}})^{\beta/(1-\beta)}C_{t+a} \quad (11)$$

$$= (1-\beta)(\frac{\beta Z_{t,a}}{w_{t+a}})^{\beta/(1-\beta)}C_{t+a}. \quad (12)$$

Also, we have:

$$\begin{aligned} w_{t+a}\exp(S_a) &= w_{t+a}\exp(S[Z] + \tfrac{\beta}{1-\beta}[\ln(\tfrac{Z_{t,a}}{Z_{t,0}}) - \theta_E a]) \\ &= w_{t+a}\exp(S[Z]).\exp(\tfrac{\beta}{1-\beta}[\ln(\tfrac{Z_{t,a}}{Z_{t,0}}) - \theta_E a]) \\ &= w_{t+a}(1-\beta)\tfrac{C}{w_{t+a}}(\tfrac{\beta Z_{t,a}}{w_{t+a}})^{\beta/(1-\beta)}\exp(\tfrac{\beta}{1-\beta}[\ln(\tfrac{Z_{t,a}}{Z_{t,0}}) - \theta_E a]) \quad (13) \\ &= (1-\beta)(\tfrac{\beta Z_{t,a}}{w_{t+a}})^{\beta/(1-\beta)}C_{t+a}[\tfrac{(Z_{t,a})^{\beta/(1-\beta)}}{(Z_{t,a}\exp(\theta_E t))^{\beta/(1-\beta)}}][\tfrac{1}{\exp(\theta_E a)}]^{\beta/(1-\beta)} \\ &= (1-\beta)(\tfrac{\beta Z_{t,a}}{w_{t+a}})^{\beta/(1-\beta)}C_{t+a}\tfrac{(Z_{t,a})^{\beta/(1-\beta)}}{(Z_{t,a})^{\beta/(1-\beta)}}. \end{aligned}$$

□

The derivation of Eq. (14) and Eq. (15)

Proof. The size of firm evolves with age according to $\mathrm{d}s_a = \mu\mathrm{d}a + \sigma\mathrm{d}W_{t,a}$. Apply the Ito formula for s_a we have:

$$\begin{aligned} \mathrm{d}s_a &= \tfrac{\partial s_a}{\partial a}\mathrm{d}a + \tfrac{\partial s_a}{\partial W_{t,a}}\mathrm{d}W_{t,a} \\ &= \tfrac{\beta}{1-\beta}\tfrac{\partial s_a}{\partial a}[ln\tfrac{Z\exp(\theta_E t + \theta_I a + \sigma_Z W_{t,a})}{Z\exp(\theta_E t)} + \theta_E a]\mathrm{d}a + \tfrac{\beta}{1-\beta}\sigma_Z\mathrm{d}W_{t,a} \quad (14) \\ &= \tfrac{\beta}{1-\beta}(\theta_I - \theta_E)\mathrm{d}a + \tfrac{\beta}{1-\beta}\sigma_Z\mathrm{d}W_{t,a}. \end{aligned}$$

Moreover, we see that $\exp(s_a)$ is a geometric Brownian motion and can be presented by:

$$\exp(s_a) = \exp(\mu t + \sigma_Z\mathrm{d}W_{t,a}). \quad (15)$$

Let us consider $\exp(\mu t + \sigma_Z \mathrm{d}W_{t,a}) = X_t$. Applying the Ito formula, we can obtain:

$$\mathrm{d}X_t = (\mu + \frac{1}{2}\sigma_Z^2)X_t\mathrm{d}t + \sigma_Z X_t \mathrm{d}W_t. \tag{16}$$

Then, the drift of the Geometric Brownian Motion or the mean of firm size is $\mu + \frac{1}{2}\sigma_Z^2$. The GBM satisfies the Gibrat law which states that the growth rate of firm size is independent of the current state (current size).

According to Proposition 3.3.1 in "The Theory of Zipf's law and Beyond," under all assumption mentioned so far and also if firm size follows the GBM, provided that the intensity of firm birth is constant, (Gibrat's law holds) there exists a steady state mean density of firm size for any $s > 0$. □

The solution of Eq. (19)

Proof. The first equation of (19) is a linear, second order non-homogeneous differential equation of the type:

$$ay''(s) + by'(s) + cy(s) = g(s). \tag{17}$$

Let the associated homogeneous differential equation be:

$$ay''(s) + by'(s) + cy(s) = 0. \tag{18}$$

If $Y_1(s), Y_2(s)$ is the two solutions for the non-homogeneous differential equation (17) and $y_1(s)$ and $y_2(s)$ is the set of fundamental solutions for the associated homogeneous differential equation (18), then $Y_1(s) - Y_2(s)$ is also a solution of equation (18) and can be written as:

$$Y_1(s) - Y_2(s) = n_1 y_1(s) + n_2 y_2(s). \tag{19}$$

Proof. Let us plug $Y_1(s) - Y_2(s)$ into the homogeneous equation (18) we have:

$$\begin{aligned} a(Y_1 - Y_2)'' + b(Y_1 - Y_2)' + c(Y_1 - Y_2) &= 0 \\ aY_1'' + bY_1' + cY_1 - (aY_2'' + bY_2' + cY_2) &= 0 \\ g(s) - g(s) &= 0. \end{aligned} \tag{20}$$

So $Y_1 - Y_2$ is the solution of (18).

□

Also, since $y_1(s)$ and $y_2(s)$ is the fundamental set of solutions for the associated homogeneous differential equation, they form a general solution and any solution of (18) can be written as:

$$y(s) = n_1 y_1(s) + n_2 y_2(s). \tag{21}$$

As $Y_1 - Y_2$ is the solution of equation (18) so we can write:

$$Y_1(s) - Y_2(s) = n_1 y_1(s) + n_2 y_2(s).$$

We use this fact to get the general solution for the non-homogeneous equation (17).

Let $y(s)$ be the general solution for (17) and $Y_N(s)$ a particular solution for (17) that we can get, then we can have:

$$\begin{aligned} y(s) - Y_N(s) &= n_1 y_1(s) + n_2 y_2(s) \\ y(s) &= n_1 y_1(s) + n_2 y_2(s) + Y_N(s). \end{aligned} \tag{22}$$

If we define $Y_H(s) = n_1 y_1(s) + n_2 y_2(s)$ as the complementary solution, the general solution of the differential equation can be written as:

$$y(s) = Y_H(s) + Y_N(s). \tag{23}$$

In other words, the solution of the ordinary differential equation is the sum of the solution of the corresponding homogeneous equation, equation (18) and the particular solution to (17).

Considering our case, we have:

$$-(r-\kappa)V(s) + V'(s)\mu + \frac{1}{2}V''(s)\sigma^2 = -\exp(s) + \lambda_F(s_a).$$

To find the solution, we first need to find the complementary solution for the associated homogeneous equation:

$$-(r-\kappa)V(s) + \frac{1}{2}\sigma^2 V''(s) + \mu V'(s) = 0. \tag{24}$$

The general solution for the homogeneous dynamic equation can be calculated by solving the associated quadratic polynomial $\frac{1}{2}\sigma^2 X^2 + \mu X - (r - \kappa)$. Then, the complementary solution is: $c_1 \exp(\xi s) + c_2 \exp(\xi_* s)$ where ξ, ξ_* is the two roots of the quadratic polynomial.

$$\xi = -\frac{-\mu - \sqrt{\mu^2 + 2(r-\kappa)\sigma^2}}{\sigma^2} = -\frac{\mu}{\sigma^2} - \sqrt{(\frac{\mu}{\sigma^2})^2 + \frac{r-\kappa}{\sigma^2/2}}; \tag{25}$$

and

$$\xi_* = \frac{-\mu + \sqrt{\mu^2 + 2(r-\kappa)\sigma^2}}{\sigma^2} = -\frac{\mu}{\sigma^2} + \sqrt{(\frac{\mu}{\sigma^2})^2 + \frac{r-\kappa}{\sigma^2/2}}. \tag{26}$$

We have $\xi < 0$ and $\xi_* > 0$. The homogeneous equation represents the value of option to exit. Since the probability of the firm with high productivity to exit is small, the value of exit option goes to zero as productivity goes to infinity. And since $\xi_* > 0$, we expect to have $c_2 = 0$ and thus the solution for the homogeneous ODE is:

$$Y_H(s) = c_1 \exp(\xi s). \tag{27}$$

We now need to find the particular solution for the non-homogeneous ODE. We utilize the fact that if $Y_1(s)$ is the particular solution for $ay'' + by' + cy = g_1(s)$ and $Y_2(s)$ is the particular solution for $ay''+by'+cy = g_2(s)$ then $Y_1(s)+Y_2(s)$ is the particular solution for $ay''+by'+cy = g_1(s) + g_2(s)$.

First, we can write:

$$g(s) = -\exp(s) + \lambda_F(s) = g_1(s) + g_2(s). \tag{28}$$

Now, consider:

$$V'(s)\mu + \frac{1}{2}V''(s)\sigma^2 - (r-\kappa)V(s) = -\exp(s) = g_1(s). \tag{29}$$

We use the “Undetermined Coefficient” method to solve the equation. In particular we look into the function form of $g_1(s)$ and guess the form of $Y_1(s)$, leaving the coefficient undetermined. Then, we plug the guess into the differential equation and try to determine the value of the coefficient. If we can determine the value of the coefficients then the guess is correct.

For $g_1(s)$, we can guess the solution since the forcing term has the form of $A\exp(s)$. By plugging the solution into the non-homogeneous ODE, we have:

$$\begin{aligned} \mu V'(s) + \frac{1}{2}\sigma^2 V''(s) - (r-\kappa)V(s) &= -\exp(s) \\ \mu A e^s + \frac{1}{2}\sigma^2 A e^s - (r-\kappa)Ae^s &= -\exp(s) \\ A &= \frac{-1}{1/2\sigma^2+\mu-(r-\kappa)}. \end{aligned} \quad (30)$$

Then, the particular solution for the non-homogeneous differential equation with $g_1(s) = \exp(s)$ is:

$$Y_1(s) = -\frac{1}{\sigma^2 + \mu - (r-\kappa)} \exp(s). \quad (31)$$

Therefore,

$$y(s) = c_1 \exp(\xi s) - \frac{1}{1/2\sigma^2 + \mu - (r-\kappa)} \exp(s) + Y_2(s). \quad (32)$$

Now, utilizing the two boundary conditions: $V(b) = 0$ and $V'(b) = 0$, we can calculate c_1 and b.

We have:

$$V(b) = 0 \quad (33)$$

$$c_1 \exp(\xi b) - \frac{1}{1/2\sigma^2 + \mu - (r-\kappa)} \exp(b) + Y_2(s) = 0. \quad (34)$$

Then,

$$c_1 = \exp(-\xi b)[\frac{1}{1/2\sigma^2 + \mu - (r-\kappa)} \exp(b) - Y_2(s)]. \quad (35)$$

Also, since $V'(b) = 0$, we have

$$c_1 \xi \exp(\xi b) - \frac{1}{1/2\sigma^2 + \mu - (r-\kappa)} \exp(b) = 0. \quad (36)$$

By combing the solution from (35) into (36), we can obtain:

$$\begin{aligned} \exp(-\xi b)[\frac{1}{1/2\sigma^2+\mu-(r-\kappa)} \exp(b) - Y_2(s)]\xi \exp(\xi b) - \frac{1}{1/2\sigma^2+\mu-(r-\kappa)} \exp(b) &= 0 \\ \xi[\frac{1}{1/2\sigma^2+\mu-(r-\kappa)} \exp(b) - Y_2(s)] - \frac{1}{1/2\sigma^2+\mu-(r-\kappa)} \exp(b) &= 0 \\ \frac{\exp(b)(\xi-1)}{1/2\sigma^2+\mu-(r-\kappa)} - \frac{\xi Y_2(s)}{1/2\sigma^2+\mu-(r-\kappa)} &= 0. \end{aligned} \quad (37)$$

The exit barrier is then,

$$\exp(b) = \frac{\xi}{\xi - 1} + Y_2(s). \tag{38}$$

Now if we assume a certain function form for $g_2(s)$ and use the same "Undetermined Coefficient" method we can have exact form of $Y_2(s)$. Then using the boundary conditions ($V(b) = 0$ and $V'(b) = 0$) we can solve for c_1 and value of b to find the value of the firm $V(s)$ and the exit threshold b. In particular, the value of firm size $V(s)$ and the exit threshold are, respectively:

$$\begin{array}{lcl} V(s) & = & \frac{1}{r-\kappa}(\frac{\xi}{1+\xi})(\exp(s-b) - 1 - \frac{1-\exp(-\xi(s-b))}{\xi}) \text{ for } s \geq b \\ \exp(b) & = & (\frac{\xi}{1+\xi})(1 - \frac{\mu+\sigma^2/2}{r-\kappa}). \end{array} \tag{39}$$

□

The derivation of Eq. (26) to (28)

Given the initial density and the lower bound b, the Kolmogorov equation of firms with initial size x provides the transition probability density of the Brownian Motion process $z_{t,a}$ (the productivity of the firm). We can see that this represents the absorbed Brownian Motion with drift where the absorbing barrier is b.

Following Taylor and Karlin (1998), a standard Brownian Motion without drift with the absorbing barrier being zero and the variance being unity has the transition distribution, which can be calculated using the reflection principle as:

$$\Pr(S(a) > s | S(0) = x) = \Phi(\frac{s+x}{\sqrt{a}}) - \Phi(\frac{s-x}{\sqrt{a}}), \tag{40}$$

where $\Phi(.)$ is the cdf of the normal distribution function

If the absorbing barrier equals $b \neq 0$, using the reflection principle again, the transition distribution will be:

$$\Pr(S(a) > s | S(0) = x) = \Phi(\frac{s+x-2b}{\sqrt{a}}) - \Phi(\frac{s-x}{\sqrt{a}}). \tag{41}$$

The next step is to calculate the distribution for the Brownian motion with drift. To construct the

(μ, σ) Brownian Motion with time domain $[0, T]$, we keep the origin process X and change the probability measure. In other words, we replace the probability measure $P(\Omega, F, P)$ with the other probability measure $P^*(\Omega, F, P^*)$ such that X is a Brownian Motion on P^*. Following the Change of Measure Theorem as in Harrison (1985)[3], we define P^* the new probability as:

$$\mathrm{d}P^* = \xi \mathrm{d}P = \exp[\frac{\mu}{\sigma^2}(s-x) - \frac{\mu^2 a}{2\sigma^2}]\mathrm{d}P. \tag{42}$$

We can derive the transition distribution of the Brownian motion with drift with absorbing barrier equals b, which is:

$$\psi(a, s|x) = \frac{1}{\sigma\sqrt{a}}[\phi(\frac{s-x-\mu a}{\sigma\sqrt{a}}) - \exp[-2\mu(x-b)/(\sigma^2)]\phi(\frac{s+x-2b-\mu a}{\sigma\sqrt{a}})], \tag{43}$$

where $\phi(.)$ represents the normal density function

To find the firm marginal, we integrate $m(a, s)$ over all ages:

$$\begin{aligned} m(s) &= \textstyle\int_0^\infty \exp(-\eta a)\psi(a, s|x)\mathrm{d}a \\ &= \textstyle\int_0^\infty [\frac{\exp(-\eta a)}{\sigma\sqrt{a}}\phi(\frac{s-x-\mu a}{\sigma\sqrt{a}})] - \exp[-2\mu(x-b)/(\sigma^2)[\frac{\exp(-\eta a)}{\sigma\sqrt{a}}\phi(\frac{s+x-2b-\mu a}{\sigma\sqrt{a}})]da. \\ &= A_1 - \exp[-2\mu(x-b)/(\sigma^2)]A_2, \end{aligned} \tag{44}$$

where: $A_1 = \frac{\exp(-\eta a)}{\sigma\sqrt{a}}\phi(\frac{s-x-\mu a}{\sigma\sqrt{a}})$ and $A_2 = \frac{\exp(-\eta a)}{\sigma\sqrt{a}}\phi(\frac{s+x-2b-\mu a}{\sigma\sqrt{a}})$.

Now consider A_1, we have:

$$A_1 = \int_0^\infty \frac{\exp(-\eta a)}{\sigma\sqrt{a}}\phi(\frac{s-x-\mu a}{\sigma\sqrt{a}})\mathrm{d}a.$$

Since the standard normal density function of a random variable z, $\phi(z)$, has the form of $\phi(z) = \frac{1}{\sqrt{2\pi}}\exp(-1/2z^2)$, we have:

[3]The Change of Measure Theorem: Given $\theta \in R$, let $\xi, P*$ be defined as: (i) $\xi = V_\gamma(T) = exp[\gamma X_t - (\mu\gamma + 1/2\sigma^2\gamma^2)]$ where $\gamma = \frac{\theta}{\sigma^2}$; (ii)$P^*(A) = \int_A \xi(\omega)Pd(\omega)$ where $A \in \mathcal{F}$ or $\mathrm{d}P^* = \xi\mathrm{d}P$.Then X is a $(\mu+\theta, \sigma)$ Brownian Motion on (Ω, F, P).

$$\begin{aligned} A_1 &= \tfrac{1}{\sigma}\int_0^\infty \tfrac{\exp(-\eta a)}{\sqrt{2\pi a}}\exp(-\tfrac{1}{2}(\tfrac{s-x-\mu a}{\sigma\sqrt{a}})^2)\mathrm{d}a \\ &= \tfrac{1}{\sigma}\int_0^\infty[\tfrac{1}{\sqrt{2\pi a}}\exp[-\tfrac{1}{2\sigma^2 a}(s^2+x^2+\mu^2a^2-2sx-2s\mu a+2x\mu a)-\eta a]]da \\ &= \tfrac{1}{\sigma}\int_0^\infty \tfrac{1}{\sqrt{2\pi a}}\exp(-\tfrac{\mu}{\sigma^2}(s-x)-\tfrac{1}{2\sigma^2 a}[(s-x)^2+(2\eta\sigma^2+\mu^2)a^2])\mathrm{d}a. \end{aligned} \tag{45}$$

Solving the integral, we have:

$$\begin{aligned} A_1 &= \tfrac{1}{\sigma}\exp(\tfrac{(s-x)\mu}{\sigma^2})\tfrac{1}{\sqrt{\frac{\mu^2+2\sigma^2\eta}{\sigma^2}}}\exp(-\tfrac{|s-x|}{\sigma}\sqrt{\tfrac{\mu^2+2\sigma^2\eta}{\sigma^2}}) \\ &= \tfrac{1}{\sqrt{\mu^2+2\sigma^2\eta}}\exp(\tfrac{1}{\sigma^2}[(s-x)\mu-|s-x|\sqrt{\mu^2+2\sigma^2\eta}]) \\ &= \tfrac{\min(\exp(\alpha_*(s-x)),\exp(-\alpha(s-x)))}{\sqrt{\mu^2+2\sigma^2\eta}}. \end{aligned} \tag{46}$$

where $\alpha = \frac{-\mu}{\sigma^2}+\sqrt{(\frac{\mu}{\sigma^2})^2+\frac{\eta}{\sigma^2/2}}$ and $\alpha_* = \frac{\mu}{\sigma^2}+\sqrt{(\frac{\mu}{\sigma^2})^2+\frac{\eta}{\sigma^2/2}}$.

Now consider A_2, we have:

$$A_2 = \int_0^\infty \frac{\exp(-\eta a)}{\sigma\sqrt{a}}\phi(\frac{s+x-2b-\mu a}{\sigma\sqrt{a}})\mathrm{d}a.$$

Since $\phi(.)$ is also the standard normal density function, we have:

$$A_2 = \int_0^\infty \frac{\exp(-\eta a)}{\sigma\sqrt{2\pi a}}\exp(-\frac{1}{2}(\frac{(s-b)+(x-b)-\mu a}{\sigma\sqrt{a}})^2))\mathrm{d}a. \tag{47}$$

Following the similar steps as in equation (45) and solve for the integral, we have:

$$A_2 = \frac{1}{\sigma}\exp(\frac{[(s-b)+(x-b)]\mu}{\sigma^2})\frac{1}{\sqrt{\frac{\mu^2+2\sigma^2\eta}{\sigma^2}}}\exp(-\frac{|(s-b)+(x-b)|}{\sigma}\sqrt{\frac{\mu+2\sigma^2\eta}{\sigma^2}}). \tag{48}$$

Since $s > b$ and $x > b$, we have:

$$\begin{aligned} A_2 &= \tfrac{1}{\sqrt{\mu^2+2\sigma^2\eta}}\exp[(s+x-2b)(\tfrac{\mu}{\sigma^2}-\sqrt{(\tfrac{\mu}{\sigma^2})^2+\tfrac{\eta}{\sigma^2/2}}] \\ &= \tfrac{1}{\sqrt{\mu^2+2\sigma^2\eta}}\exp[(s+x-2b)(-\alpha). \end{aligned} \tag{49}$$

From equation (46) and (49), the density of firm size can be written as:

$$\begin{aligned} m(s) &= \frac{\min(\exp[\alpha_*(s-x)],\exp[-\alpha(s-x)])}{\sqrt{\mu^2+2\sigma^2\eta}} - \exp[\frac{-2\mu(x-b)}{\sigma^2}]\frac{1}{\sqrt{\mu^2+2\sigma^2\eta}}\exp[(s+x-2b)(-\alpha) \\ &= \frac{1}{\sqrt{\mu^2+2\sigma^2\eta}}[M - \exp(\frac{-2\mu(x-b)}{\sigma^2} - (s+x-2b)(-\alpha)], \end{aligned} \tag{50}$$

where $M = \min(\exp[\alpha_*(s-x)], \exp[-\alpha(s-x)])$.

Since $\alpha - \alpha_* = -\frac{2\mu}{\sigma^2}$, we have:

$$\begin{aligned} \exp(\frac{-2\mu(x-b)}{\sigma^2} - [(s+x-2b)(-\alpha)]) &= \exp[(\alpha-\alpha_*)(x-b) + (x-b+s-b)\alpha] \\ &= \exp[(-\alpha_*(x-b)]\exp[-\alpha(s-b)]. \end{aligned} \tag{51}$$

Then,

$$m(s) = \frac{1}{\sqrt{\mu^2+2\sigma^2\eta}}[M - \exp[-\alpha_*(x-b)]\exp[-\alpha(s-b)]]. \tag{52}$$

Also, since $\alpha + \alpha_* = \frac{2}{\sigma^2}\sqrt{\mu^2+2\sigma^2\eta}$, $\frac{1}{\sqrt{\mu^2+2\sigma^2\eta}} = -\frac{1}{\mu}\frac{\alpha-\alpha_*}{\alpha+\alpha_*}$ and then,

$$m(s) = -\frac{1}{\mu}\frac{\alpha-\alpha_*}{\alpha+\alpha_*}[M - \frac{1}{\exp[\alpha_*(x-b)]\exp[\alpha(s-b)]}]. \tag{53}$$

Replace $M = \min(\exp[\alpha_*(s-x)], \exp[-\alpha(s-x)$ and rearrange, we have:

$$\begin{aligned} m(s) &= -\frac{1}{\mu}\frac{\alpha-\alpha_*}{\alpha+\alpha_*}[\frac{\min(\exp[\alpha_*(s-b)]\exp[\alpha(s-b)],\exp[\alpha_*(x-b)]\exp[\alpha(x-b)])-1}{\exp[\alpha_*(x-b)]\exp[\alpha(s-b)]}] \\ &= -\frac{1}{\mu}\frac{\alpha-\alpha_*}{\alpha+\alpha_*}[\frac{\min(\exp[(\alpha+\alpha_*)(s-b)],\exp[(\alpha+\alpha_*)(x-b)])-1}{\exp[\alpha_*(x-b)]\exp[\alpha(s-b)]}] \\ &= -\frac{1}{\mu}\frac{\alpha-\alpha_*}{\alpha+\alpha_*}G, \end{aligned} \tag{54}$$

where $G = [\frac{\min(\exp[(\alpha+\alpha_*)(s-b)],\exp[(\alpha+\alpha_*)(x-b)])-1}{\exp[\alpha_*(x-b)]\exp[\alpha(s-b)]}]$

Then the probability density can be derived from $m(s)$ by integrating $m(s)$ over s to find the corresponding normalizing constant.

$$\begin{aligned}\int_b^{\infty} m(s)ds &= -\frac{1}{\mu}\frac{\alpha-\alpha_*}{\alpha+\alpha_*}\int_b^{\infty}[\frac{\min(\exp[(\alpha+\alpha_*)(s-b)],\exp[(\alpha+\alpha_*)(x-b)])-1}{\exp[\alpha_*(x-b)]\exp[\alpha(s-b)]}]ds \\ &= -\frac{1}{\mu\exp[\alpha_*(x-b)]}\frac{\alpha-\alpha_*}{\alpha+\alpha_*}[\int_b^{\infty}\frac{\min(\exp[(\alpha+\alpha_*)(s-b)],\exp[(\alpha+\alpha_*)(x-b)])-1}{\exp[\alpha(s-b)]}]\mathrm{d}s \\ &= -\frac{1}{\mu\exp[\alpha_*(x-b)]}\frac{\alpha-\alpha_*}{\alpha+\alpha_*}F,\end{aligned} \tag{55}$$

where $F = \int_b^{\infty}\frac{\min(\exp[(\alpha+\alpha_*)(s-b)],\exp[(\alpha+\alpha_*)(x-b)])-1}{\exp[\alpha(s-b)]}\mathrm{d}s$.

Since for $b < s < x$, we have $s - b < x - b$ while for $s > x$ we have $s - b > x - b$, then rearrange the integral we have:

$$F = \int_b^x \frac{\exp[(\alpha+\alpha_*)(s-b)]}{\exp[\alpha(s-b)]}\mathrm{d}s + \int_x^{\infty}\frac{\exp[(\alpha+\alpha_*)(x-b)]}{\exp[\alpha(s-b)]}ds - \int_b^{\infty}\frac{1}{\exp[\alpha(s-b)]}\mathrm{d}s. \tag{56}$$

Solving the integrals, we have:

- $$\int_b^x \frac{\exp[(\alpha+\alpha_*)(s-b)]}{\exp[\alpha(s-b)]}\mathrm{d}s = \int_b^x \exp[\alpha_*(s-b)]\mathrm{d}s = \frac{\exp[\alpha_*(x-b)]}{\alpha_*} - \frac{1}{\alpha_*}.$$

- $$\begin{aligned}\int_x^{\infty}\frac{\exp[(\alpha+\alpha_*)(x-b)]}{\exp[\alpha(s-b)]}ds &= \exp[(\alpha+\alpha_*)(x-b)]\int_x^{\infty}\frac{1}{\exp[\alpha(s-b)]} \\ &= \exp[(\alpha+\alpha_*)(x-b)]\frac{1}{\alpha\exp[\alpha(x-b)]} \\ &= \frac{\exp[\alpha_*(x-b)]}{\alpha}.\end{aligned}$$

- $$\int_b^{\infty}\frac{1}{\exp[\alpha(s-b)]}\mathrm{d}s = \frac{1}{\alpha}.$$

Then,

$$\begin{aligned}F &= \frac{\exp[\alpha_*(x-b)]}{\alpha_*} - \frac{1}{\alpha_*} + \frac{\exp[\alpha_*(x-b)]}{\alpha} - \frac{1}{\alpha} \\ &= \frac{\alpha+\alpha_*}{\alpha\alpha_*}[\exp[\alpha_*(x-b)] - 1].\end{aligned} \tag{57}$$

Moreover, we have $\alpha - \alpha_* = -\frac{2\mu}{\sigma^2}$ and $\alpha\alpha_* = \frac{2\eta}{\sigma^2}$. By replacing into the above equation, we have:

$$\begin{aligned}\int_b^{\infty} m(s)\mathrm{d}s &= -\frac{1}{\mu}\frac{-2\mu}{2\eta}[1-\exp[-\alpha_*(x-b)]] \\ &= \frac{1-\exp[-\alpha_*(x-b)]}{\eta}.\end{aligned} \tag{58}$$

Then, the firm size density is the weighted average of the density $\pi(s|x)$ of size, conditional on the initial size. For any $x > b$, we have:

$$\pi(s|x) = \frac{m(s)}{\int_b^\infty m(s)ds}. \tag{59}$$

Substituting equation (58) and (54), we have:

$$\pi(s|x) = -\frac{1}{\mu} \cdot \frac{\alpha - \alpha_*}{\alpha + \alpha_*} \cdot G] \cdot [-\frac{1}{\mu} \cdot \frac{\alpha - \alpha_*}{\alpha\alpha_*} \cdot [1 - \frac{1}{\exp[\alpha_*(x-b)]}]^{-1}. \tag{60}$$

Because

$$-\frac{1}{\mu} \cdot \frac{\alpha - \alpha_*}{\alpha + \alpha_*} \cdot \left[-\frac{1}{\mu} \cdot \frac{\alpha - \alpha_*}{\alpha\alpha_*} \cdot [1 - \frac{1}{\exp[\alpha_*(x-b)]}]\right]^{-1} = \frac{\alpha\alpha_*}{\alpha + \alpha_*} \cdot \frac{\exp(\alpha_*(x-b)}{\exp(\alpha_*(x-b)) - 1},$$

Rearranging $\pi(s|x)$ we have:

$$\begin{aligned} \pi(s|x) &= [\tfrac{\min(\exp[(\alpha+\alpha_*)(s-b)],\exp[(\alpha+\alpha_*)(x-b)])-1}{\exp[\alpha_*(x-b)]\exp[\alpha(s-b)]}] \times \tfrac{\alpha\alpha_*}{\alpha+\alpha_*} \cdot \tfrac{\exp(\alpha_*(x-b))}{\exp(\alpha_*(x-b))-1}. \\ &= [\tfrac{\min(\exp[(\alpha+\alpha_*)(s-b)],\exp[(\alpha+\alpha_*)(x-b)])-1}{\exp[\alpha(s-b)]}] \times [\tfrac{\alpha+\alpha_*}{\alpha\alpha_*} \cdot (\exp(\alpha_*(x-b)) - 1)]^{-1} \\ &= (\tfrac{\exp[\alpha_*(x-b)-1]}{\alpha_*} \tfrac{\exp[\alpha(s-b)]}{\alpha})^{-1} \min[\tfrac{\exp[[\alpha+\alpha_*](s-b)]-1}{\alpha+\alpha_*}, \tfrac{\exp[(\alpha+\alpha_*)(x-b)]-1}{\alpha+\alpha_*}]. \end{aligned} \tag{61}$$

The derivation of Equation (30)

The total labor hours available in the market are divided into (i) L_E: the number of hours to set up new firms (covering the entry cost λ_E), (ii) L_F: the number of hours to maintain the existing firms (covering the fixed cost λ_F) and (iii) L: the number of hours to produce goods.

In addition, the measure of existing firms in the market, defined on the set of possible ages (a) and sizes (s), grows at the rate of η . This has the density at time t of $m(s)I\exp(\eta t)$ where $I\exp(\eta t)$ is the number of new firms attempting to enter the market per unit of time.

Then the total amount of labor used to cover entry cost λ_E is:

$$L_E = \lambda_E I. \tag{62}$$

The total labor to create new entrants equals entry cost multiplies with the measure of potential entrants.

Since in the balance growth path, the measure of firms in the market is (M), the firm density $(m(s))$, then, has the measure of M/M', where M' capture the measure of potential firms that are trying to enter the market per unit of time. Then, the measure of active firms can be represented as $\lambda_F(s).I\int_b^\infty m(s)ds$. In other words, the total amount of labor to cover fixed cost $\lambda_F(s)$ is:

$$L_F = \lambda_F(s).I\int_b^\infty m(s)ds. \tag{63}$$

The, amount of labor to cover production of a firm is determined through profit maximization condition through Eq. (10) to (13):

From (10) we can write:

$$R_t = \frac{w_t L_t}{\beta}.$$

Substituting R_t into (11) we have:

$$\begin{aligned} \frac{w_t L_t}{\beta} - w_t(L_t + \lambda_F(s)) &= w_t(\exp(s) - \lambda_F(s)) \\ L_t - \beta(L_t + \lambda_F(s)) &= \beta(\exp(s) - \lambda_F(s)) \\ L_t &= \frac{\beta \exp(s) - \lambda_F(s)(\beta - 1)}{1-\beta}. \end{aligned} \tag{64}$$

Thus, the total amount of labor to cover production is:

$$L = \frac{I}{1-\beta} \int_b^\infty [\beta \exp(s) - \lambda_F(s)(\beta - 1)] ds. \tag{65}$$

Calibration

Entry and exit and the heterogeneity of firms have been proved to be important factors that contribute to firm growth in the literature. In chapter 3, we have developed the model in which growth is the result of the increase in variety, technology progress and selection. In this section, using the firm data from Vietnam in 2005, we will measure the magnitude of each factor's impact to firm growth.

Following Luttmer (2007), β parameter in the monopolistic competitive market is set to be close to unity, which is 0.9. This implies that in the economy, goods produced by different firms are close substitutes and hence introducing another variety only moderately increases the utility. The demand curve faced by firms is nearly perfectly elastic. We now try to decompose the growth rate of per capita consumption, κ. From Equation (16), we have:

$$\begin{aligned} \kappa &= \theta_E + (\tfrac{1-\beta}{\beta}\eta) \\ &= \theta_I + (\theta_E - \theta_I) + (\tfrac{1-\beta}{\beta}\eta). \end{aligned} \tag{66}$$

The per capita consumption growth rate can be decomposed into within-firm growth rate (θ_I), the selection component ($\theta_E - \theta_I$) and the increase of variety of goods ($\frac{1-\beta}{\beta}\eta$).

First, the contribution of increase in variety is then:

$$\frac{1-\beta}{\beta}\eta = .00133.$$

The proportion of increase due to changes in variety in per capita consumption growth rate is minor given the fact that goods are close substitute and the population growth rate in Vietnam is about 1.17% in 2005[4].

To calculate the within firm growth rate (θ_I) and the selection component ($\theta_E - \theta_I$), we need to calculate μ, σ^2 and σ_Z.

First, from the definition of α in Equation (15), assuming η is minor, we have:

$$-\mu \approx \frac{\alpha\sigma^2}{2}, \tag{67}$$

where α is the tail index of the firm size distribution and σ represents the variation of firm size.

The variance σ^2 can be calculated from the entry rate (ς) - the rate of new firm successfully enter the market per unit of time over the total number of firms. This, in turn, equals population growth rate plus the exit rate in the equilibrium.

Also, following Luttmer (2007), we have $\varsigma = \eta + \alpha^2\sigma^2/2$, we can now calculate for the variance σ^2 and the drift μ.

In particular, we have

$$\begin{aligned} \mu &= \tfrac{\epsilon-\eta}{\alpha} \\ \sigma^2 &= 2\tfrac{\epsilon-\eta}{\alpha^2}. \end{aligned} \tag{68}$$

In addition, σ_Z can be calculated using:

$$\sigma_Z = \frac{\sigma(1-\beta)}{\beta}. \tag{69}$$

Now, considering Vietnam economy in 2005, we have the entry rate of firms in Vietnam to be

[4]The results for other years are available upon request.

17.8% (GSO, 2006) and the population growth rate to be 1.3%.

The tail indexes α is the slopes of the distribution line. In this chapter, we follow (Axtell, 2001) to estimate the value of the distribution tail index of firm size by regressing the logarithm of frequency of firm appearance in a certain size group against the logarithm of that size, allowing for an error term in the deterministic specification. The results of tail index of Vietnam in different years as well we across size groups in 2005 is presented in the tables below.

Table 1: Values of the tail index - α, 2000-2005

Year	2000	2001	2002	2003	2004	2005
α	-1.07	-1.1	-1.14	-1.14	-1.16	-1.18
P-value	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
R^2	.79	.79	.8	.8	.8	.79

Note: α is the tail index of the FSD. R^2 is the coefficient of determination.

Table 2: Values of α across size groups in 2005

Size group	1-100	100-200	>200
α	-1.31	-1.03	-0.8
P-value	< 0.0001	< 0.0001	< 0.0001

Here, we group the samples into three group which include firms with 1-100, 100-200 and over 200 employees[5].

[5] We use these size cutoffs when grouping the sample to ensure that we have enough firms in each group to ensure the characteristics of size distribution. We have also tried some other different cutoffs, of which results are not substantially different from the presented result.

Table 3: Value of μ and σ^2

size group	α	$-\mu$	σ^2	σ_Z
mean	1.18	0.14	$(0.487)^2$	0.054
1 - 100	1.31	0.12	$(0.439)^2$	0.048
100 - 200	1.03	0.15	$(0.532)^2$	0.06
>200	0.85	0.19	$(0.676)^2$	0.08

Different from developed economies as specified in Luttmer (2007), the tail index of the FSD of Vietnam is significantly different from unity. In the literature, Zipf's law (i.e. the tail index of the firm size distribution equals unity) has been confirmed in many countries including United State, Denmark or many European countries. However, among developing economies, the deviation from Zipf's law is not uncommon. The tail index of FSD can be found to be greater than unity in countries like Cambodia with $\alpha = 1.33$ (Tanaka and Hatsukano, 2010), or Ghana with $\alpha = 1.6$ (Giovanni and Goyette, 2009). This, therefore, also contributes to the scare literature of FSD research by providing relatively significant evidence of the rejection of Zipf's law in the distribution of firm size in developing economies.

The calculations of μ, σ^2 and σ_Z are presented in Table (3).

From the Table 3, we can see that, compared to US economy, Vietnamese economy is characterized with higher variance in productivity growth of small and large firms. This supports the empirical evidence so far that the degree of heterogeneity across firms is greater for developing countries than for the developed ones.

Now we can decompose κ. According to the General Statistic Office of Vietnam, the growth rate of GDP per capita of Vietnam in 2005 is 8.4%, increasing by 80 USD per person, compared to 2004. This number can be used to approximate the per capita consumption growth rate. The detailed decomposition is illustrated as in Table 4.

In such an economy where goods are close substitutes and the population growth rate is small as above, the contribution of variety improvement is minor, only 0.0015 in the total of 8.4%. The majority of growth rate stems from the increase in productivity of incumbent firms. One different point from Luttmer's economy is that, the contribution of selection is relatively small, compared to the productivity growth part of incumbents.

Table 4: κ decomposition

size group	μ	$\theta_I - (\frac{1-\beta}{\beta})\mu$	$\theta_E - \theta_I$	$(\frac{1-\beta}{\beta}\eta)$
mean	0.14	0.067	0.0155	0.0015
1 - 100	0.13	0.070	0.014	0.0015
100 - 200	0.15	0.066	0.017	0.0015
>200	0.19	0.061	0.021	0.0015

Illustration of the Algorithm in Chapter 5

We use the FDH algorithm developed by Soleimani-Damaneh and Reshadi (2007) (see also Soleimani-Damaneh and Mostafaee, 2009). The algorithm is as follows:

Step 1. Denote by D, a set of observations (x_j, y_j) for all $j \in N$. Compute λ^{ij} for all $i, j \in N$ by $\lambda^{ij} = \frac{y_i}{y_j}$ where $y_i, y_j \in D$.

Step 2. Compute θ^{ji} for all $i, j \in N$ by $\theta^{ji} = \max_{s \in S}\{\frac{x_{js} \cdot \lambda^{ij}}{x_{is}} : (x_i, y_i) \in D\}$.

Step 3. Compute θ_i and A_i for all $i \in N$ by $\theta_i = \min_{j \in N} \theta^{ji}$ and $A_i = \{j \in N : \theta^{ji} = \theta_i\}$.

Step 4. Compute λ_i^+ and λ_i^- for all $i \in N$ by $\lambda_i^+ = \max_{j \in A_i} \lambda^{ij}$ and $\lambda_i^- = \min_{j \in A_i} \lambda^{ij}$.

Soleimani-Damaneh and Mostafaee (2009) proves that λ_i^+ and λ_i^- are equivalent to the ones defined in Problem 3. The next simple example demonstrates the algorithm. The original version of the following example is found in Soleimani-Damaneh and Mostafaee (2009).

Table 5: Example for the Algorithm

Firm i	Labor	Capital	Output	$\frac{Output}{Labor}$	$\frac{Output}{Capital}$	θ^{i3}	Average Products (Labor)	Average Products (Capital)	λ^{i1}
Firm 1	1	1	4	4	4	1/2	2/3	1/2	-
Firm 2	2	1	3	3/2	3	3/8	16/9	2/3	3/4
Firm 3	3	4	8	8/3	2	-	-	-	2

Suppose that labor and capital are used to produce one output and our data includes three firms' production. We consider Firm 3's production. By Table 5, we can compute $\theta_3 = \theta^{13} = 2/3$, and

$A_3 = \{1\}$. Firm 1 is the most "efficient". Because $\lambda^{31} = 2$, $\lambda_3^+ = \lambda_3^- = 2$. This implies that if a same technology and management skills of Firm 1 are available for Firm 3, Firm 3 could achieve Firm 1's production efficiency by reducing its production scale. Otherwise, it indicates that there is some problem for Firm 3 to achieve Firm 1's efficiency. The above algorithm extends an idea of this simple example to the case of multiple inputs.

List of Tables

Table 6: ISIC Codes and Industry Description

ISIC	Industry Description	ISIC	Industry Description
14	Other mining and quarrying	15	Food products and beverages
17	Textiles	18	Wearing apparel etc.
19	Tanning and dressing leather	20	Wood and wood products
21	Paper and paper products	22	Publishing, printing etc.
24	Chemical products etc.	25	Rubber and plastic products
26	Nonmetallic mineral products	28	Fabricated metal products
29	Machinery and equipment etc.	36	Furniture

Table 7: Summary of Enterprise Census Data Statistics (Firm Size)

		Size			Capital			Material		
Sector	# Observations	Mean	Median	St. Dev.	Mean	Median	St. Dev.	Mean	Median	St. Dev.
Sector 14	9369	480.00	308.00	669.40	340.63	123.50	702.11	262.50	17.00	702.11
Sector 15	29713	268.63	27.50	699.80	292.63	35.50	683.13	505.75	309.50	683.13
Sector 17	6251	385.00	76.00	697.90	294.13	24.00	2716.69	2303.00	1753.50	2716.69
Sector 18	10177	427.00	224.50	663.52	275.50	19.50	689.05	358.88	73.50	689.05
Sector 19	3347	8918.63	4925.00	12172.47	362.63	31.50	2526.35	2263.63	2015.50	2526.35
Sector 20	11096	283.75	24.50	694.98	712.88	230.00	684.00	316.88	80.00	684.00
Sector 21	6986	309.75	61.00	686.64	337.88	149.50	1087.30	642.75	84.00	1087.30
Sector 22	8435	287.88	13.00	696.97	308.13	57.00	691.28	305.50	33.50	691.28
Sector 24	6927	330.63	94.00	680.31	298.75	57.00	693.97	360.13	72.00	693.97
Sector 25	9457	300.50	42.00	689.49	329.63	51.00	699.46	270.50	16.00	699.46
Sector 26	11813	276.63	22.00	697.41	283.50	40.50	698.78	388.63	73.50	698.78
Sector 28	15853	354.50	110.00	676.98	297.88	53.50	705.04	416.13	106.00	705.04
Sector 29	4203	526.00	39.00	926.41	5114.63	1129.50	690.35	321.75	33.50	690.35
Sector 36	10387	301.25	16.50	694.92	286.88	54.50	1156.72	1087.25	862.00	1156.72

[1] Size is measured by the number of employees.
[2] Capital and Material are measured in home currency (not inflation adjusted).
[3] St. Dev. is standard deviation.

Table 8: Summary of Enterprise Census Data Statistics (Number of Observations)

Year	Total	Sec 14	Sec 15	Sec 17	Sec 18	Sec 19	Sec 20	Sec 21	Sec 22	Sec 24	Sec 25	Sec 26	Sec 28	Sec 29	Sec 36
2000	8020	476	2276	336	474	221	735	355	239	369	420	988	514	201	416
2001	9282	577	2462	381	585	253	662	483	343	442	523	1046	691	267	567
2002	11287	755	2724	500	726	288	845	560	485	556	693	1106	1013	329	707
2003	12989	833	2833	549	872	318	1086	616	620	633	805	1185	1302	395	942
2004	15834	965	3316	662	1168	398	1175	779	894	756	1008	1367	1696	482	1168
2005	17915	976	3632	794	1278	444	1309	914	1068	857	1221	1465	2023	554	1380
2006	19461	926	3859	934	1351	392	1597	920	1476	990	1336	1421	2252	569	1438
2007	24149	1140	4587	1086	1849	492	1907	1158	1729	1139	1726	1705	3066	733	1832
2008	25077	2721	4024	1009	1874	541	1780	1201	1581	1185	1725	1530	3296	673	1937

Note: Each number shows the number of observations in each sector of each industry.

Table 9: Summary Statistics (Number of Employees)

	2005		2007		2009		2011	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Min.	1	2	1	2	2	1	2	1
Max.	1929	90	1300	120	2561	65	496	321
Median	15	5	11	5	11	5	8	8
Mean	37.72	8.13	30.57	7.92	27.01	8.07	20.02	21.62
Standard Deviation	87.89	8.54	70.13	10.40	78.23	9.56	37.11	40.10
# Observations	1083	494	1530	597	1568	446	1364	614
Total	1577		2127		2014		1978	

[1] All entries are measured in terms of the number of employees, except for the number of observations.

Table 10: Shapiro–Wilk Test for Normality

Year	# Observation	W^1	Z	Pr(> $\|Z\|$)
2000	8020	0.28	21.28	$< 0.0001^{***}$
2001	9282	0.26	21.78	$< 0.0001^{***}$
2002	11287	0.21	22.52	$< 0.0001^{***}$
2003	12989	0.18	22.99	$< 0.0001^{***}$
2004	15834	0.18	23.59	$< 0.0001^{***}$
2005	17915	0.14	24.05	$< 0.0001^{***}$
2006	19461	0.14	24.29	$< 0.0001^{***}$
2007	24149	0.15	24.88	$< 0.0001^{***}$
2008	25077	0.11	25.10	$< 0.0001^{***}$

Asterisks correspond to the following p-values:$^{***}p < 0.001$.

[1] The test statistic W is given by $W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$, where $x_{(i)}$ are the ordered sample values, and a_i are constants generated from the means, variances, and covariances of the order statistics of a sample of size n from a normal distribution, and $\bar{x}$ is the mean of the samples x_is.

[2] In general, the statistic is positive and less than or equal to one; being close to one indicates normality.
Note: A more detailed description about the test is found at Shapiro and Wilk (1965).

Table 11: Dip Test for Unimodality

Year	# Observation	Dip^1	Pr(> $\|Z\|$)[2]
2000	8020	0.021	$< 0.0001^{***}$
2001	9282	0.024	$< 0.0001^{***}$
2002	11287	0.022	$< 0.0001^{***}$
2003	12989	0.022	$< 0.0001^{***}$
2004	15834	0.023	$< 0.0001^{***}$
2005	17915	0.023	$< 0.0001^{***}$
2006	19461	0.041	$< 0.0001^{***}$
2007	24149	0.026	$< 0.0001^{***}$
2008	25077	0.027	$< 0.0001^{***}$

Asterisks correspond to the following p-values: $^{***}p < 0.001$.

[1] The Dip statistics measures the departure from unimodality, by computing the maximum distance between the sample distribution and the best fitting unimodal distribution function.

[2] The p-value is calculated by comparing the Dip statistics obtained from the data, and the one from random resamplings.
Note: A more detailed description of the test is found at Hartigan and Hartigan (1985).

Table 12: Logit Regression: Results and Statistics

		Model likelihood ratio test		Discrimination indexes		Rank discrim. indexes	
# Observations	7696	LR χ^2	1296.86	R^2	0.209	[6]C	0.732
# 0	4457	d.f.	17	[2]g	1.054	[7]Dxy	0.465
# 1	3239	Pr($> \chi^2$)	< 0.0001	[3]gr	2.870	[8]gamma	0.466
[1]max |deriv|	3.00E-12			[4]gp	0.226	[9]tau-a	0.226
				[5]Brier	0.205		

	β's	S.E. (β's)	[10]dF/dX	S.E. (dF/dX)	Wald Z	Pr($> \|Z\|$)
(Intercept)	-2.99	0.27			-11.21	$< 0.0001^{***}$
log($size$)	0.36	0.03	0.09	0.01	11.53	$< 0.0001^{***}$
log(age)	-0.19	0.04	-0.05	0.01	-5.38	$< 0.0001^{***}$
log($asset$)	0.11	0.02	0.03	0.01	5.38	$< 0.0001^{***}$
Tax code	0.52	0.07	0.12	0.02	7.65	$< 0.0001^{***}$
Province 1	0.85	0.08	0.21	0.02	10.31	$< 0.0001^{***}$
Province 2	0.70	0.11	0.17	0.03	6.28	$< 0.0001^{***}$
Province 3	-0.27	0.09	-0.06	0.02	-2.83	0.0046^{**}
Province 4	0.69	0.10	0.17	0.02	7.30	$< 0.0001^{***}$
Province 5	0.41	0.10	0.10	0.02	4.23	$< 0.0001^{***}$
Province 6	-0.11	0.12	-0.03	0.03	-0.88	0.3787
Province 7	1.15	0.13	-0.026	0.03	9.00	$< 0.0001^{***}$
Province 8	0.84	0.14	0.21	0.03	5.89	$< 0.0001^{***}$
Province 9	1.23	0.13	0.30	0.03	9.67	$< 0.0001^{***}$
Ltd	0.36	0.10	0.09	0.02	3.66	0.0002^{***}
Cooperative	-0.05	0.10	-0.01	0.02	-0.53	0.5964
Private	0.38	0.10	0.09	0.03	3.70	0.0002^{***}
Joint stock	-0.34	0.06	-0.08	0.01	-5.60	$< 0.0001^{***}$

[1] This is the maximum absolute value of the first derivative of the log likelihood function.
[2] The value g is Gini's mean difference of $X_i\hat{\beta}$.
[3] The value gr is the exponential of g.
[4] In logit regressions, dependent variables (Ys in this case) can be transformed into a probability estimate, and the value gp is Gini's mean difference of these probabilities (those for paying the bribes in this case).
[5] The Brier score is $\frac{1}{N}\sum_{t=1}^{n}(f_t - o_t)^2$, where f_t is the probability that was forecast, o_t is the actual outcome of the event at instance t (0 if it does not happen and 1 if it does happen), and n is the number of forecasts.
[6] C denotes the c-index, and a c-index of 0.5 denotes random splitting, whereas a c-index of 1 denotes perfect prediction.
[7] Dxy represents Somers' Dxy rank correlation between the predicted probabilities and the observed responses. Between Dxy and the c-index, $Dxy = 2(c - 0.5)$ holds. A Dxy of 0 occurs when the model's predictions are random, and when it equals 1, the model is perfectly discriminating. The definition is taken from Somers (1962).
[8] Gamma is Goodman Kruskal's γ, which measures the similarity of the orderings of the data when ranked by each of the quantities. The values range from -1.0 (no association) to 1.0 (perfect association). The definition is taken from Kruskal (1958).
[9] Tau-a is Kendall's τ_a rank correlations between predicted probabilities and observed responses. The definition is taken from Kendall (1938).
[10] The value dF/dX stands for Averaged Marginal Effects.
Note: A detailed description of these indexes is taken from Harrell (2014) and Harrell (2001). Asterisks correspond to the following p-values: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Because our study is about corruption, which may be a delicate issue, we do not disclose each province's name here.

Table 13: Fixed-Effect and Mixed-Effect Regressions

	Fixed-effect	[1]dF/dx(FE)	Mixed-effect	[2]dF/dx(ME)
(Intercept)	0.21		-2.88***	
	(0.25)		(0.29)	
Tax code	0.05*	0.02	0.54***	0.14
	(0.02)	(1.55)	(0.07)	(0.01)
$\log(size)$	0.07***	0.03	0.37***	0.08
	(0.02)	(2.27)	(0.03)	(0.01)
$\log(age)$	0.01	0.01	–	–
	(0.01)	(0.45)	–	–
$\log(asset)$	0.01	0.01	0.10***	0.02
	(0.01)	(0.38)	(0.02)	(0.00)
Ltd	0.14***	0.82	0.34***	0.10
	(0.03)	(5.59)	(0.10)	(0.02)
Cooperative	-0.00	0.0	-0.06	0.02
	(0.03)	(0.09)	(0.10)	(0.02)
Private	0.07	0.03	0.39***	0.13
	(0.03)	(2.49)	(0.10)	(0.02)
Joint stock	-0.07***	-0.03	-0.36***	-0.09
	(0.01)	(2.37)	(0.06)	(0.01)
Residual standard error	0.4466			
d.f.	4294			
Multiple R^2	0.54			
Adjusted R^2	0.18			
F-statistic	1.50			
p-value	<0.0001			
AIC		9265.45		
BIC		9334.93		
Log likelihood		-4622.72		
Num. groups: age		62		
Num. groups: Firm ID		10		
Variance: age.(Intercept)		0.02		
Variance: Firm ID.(Intercept)		0.23		
Variance: Residual		1.00		
# Observations	7696			

Asterisks correspond to the following p-values: $^{***}p < 0.001$.

[1] The variable [1]dF/dx(FE) stands for Averaged Marginal Effects in the Fixed-effect regression.

[2] The variable [2]dF/dx(ME) stands for Averaged Marginal Effects in the Mixed-effect regression. The variables in () are S.E.

Note: In the survey a firm can change its type. For example, among 1516 households in 2005, 35 became private firms, 8 became cooperatives and another 35 became ltd. companies in 2007. We can see that because each firm is assigned with each firm's ID.

Table 14: Odds Ratio (OR) and 95 % Confidence Interval

	OR	2.50%	97.50%	OR (FE)[1]	OR (ME)[2]
(Intercept)	0.05	0.03	0.08	1.24	0.06
$\log(size)$	1.44	1.35	1.53	1.07	1.44
$\log(age)$	0.82	0.77	0.88	1.01	-
$\log(asset)$	1.11	1.07	1.16	1.01	1.11
Tax code	1.68	1.47	1.93	1.05	1.71
Ltd	1.43	1.18	1.74	1.15	1.41
Cooperative	0.95	0.78	1.15	1.00	0.94
Private	1.46	1.20	1.78	1.07	1.48
Joint stock	0.71	0.63	0.80	0.93	0.70
Province 1	2.35	2.00	2.76		
Province 2	2.01	1.62	2.50		
Province 3	0.77	0.64	0.92		
Province 4	2.00	1.66	2.41		
Province 5	1.51	1.25	1.83		
Province 6	0.90	0.70	1.14		
Province 7	3.15	2.46	4.05		
Province 8	2.31	1.75	3.05		
Province 9	3.42	2.67	4.39		

[1] FE stands for fixed effect;
[2] ME stands for mixed effect.

Table 15: ANOVA Table

Source	SS	d.f.	MS	F	Pr($> F$)
Columns	0.34	4	0.086	6.14	0.0039
Error	0.21	15	0.014		
Total	0.55	19			

Table 16: Industry Inefficiency under Different Technologies

Sector 14

	Industry Inefficiency (IE)			Aggregate Firm Inefficiency (ITE)				Industry Reallocation Inefficiency			Aggregate Firm Inefficiency		
Year	CCR	BCC	FRH	CCR	BCC	FDH	Year	CCR	BCC	FRH	CCR	BCC	FDH
2000	450,200	450,200	450,194	250,567	189,358	147,479	2000	44.3%	57.9%	67.2%	55.7%	42.1%	32.8%
2001	568,825	568,825	568,814	312,644	311,653	289,330	2001	45.0%	45.2%	49.1%	55.0%	54.8%	50.9%
2002	909,740	909,740	909,736	546,971	339,535	294,086	2002	39.9%	62.7%	67.7%	60.1%	37.3%	32.3%
2003	886,702	886,702	886,695	478,871	371,054	312,766	2003	46.0%	58.2%	64.7%	54.0%	41.8%	35.3%
2004	919,758	919,758	919,694	485,303	447,643	363,595	2004	47.2%	51.3%	60.5%	52.8%	48.7%	39.5%
2005	1,133,556	1,133,556	1,133,524	592,853	554,010	465,290	2005	47.7%	51.1%	59.0%	52.3%	48.9%	41.0%
2006	1,062,847	1,062,847	1,062,812	536,190	503,729	427,965	2006	49.6%	52.6%	59.7%	50.4%	47.4%	40.3%
2007	1,413,173	1,413,173	1,413,102	746,582	649,698	527,867	2007	47.2%	54.0%	62.6%	52.8%	46.0%	37.4%
2008	1,466,380	1,466,380	1,466,340	474,055	465,444	414,499	2008	67.7%	68.3%	71.7%	32.3%	31.7%	28.3%

Sector 15

	Industry Inefficiency (IE)			Aggregate Firm Inefficiency (ITE)				Industry Reallocation Inefficiency			Aggregate Firm Inefficiency		
Year	CCR	BCC	FRH	CCR	BCC	FDH	Year	CCR	BCC	FRH	CCR	BCC	FDH
2000	2,795,785	2,795,785	2,795,772	2,473,833	1,942,119	1,615,494	2000	11.5%	30.5%	42.2%	88.5%	69.5%	57.8%
2001	2,827,905	2,827,905	2,827,894	2,567,829	1,986,499	1,711,333	2001	9.2%	29.8%	39.5%	90.8%	70.2%	60.5%
2002	4,167,794	4,167,794	4,167,793	3,705,992	2,831,759	2,393,708	2002	11.1%	32.1%	42.6%	88.9%	67.9%	57.4%
2003	4,863,678	4,863,678	4,863,662	4,349,722	3,539,044	2,769,763	2003	10.6%	27.2%	43.1%	89.4%	72.8%	56.9%
2004	5,881,124	5,881,124	5,881,103	5,374,355	4,488,779	3,475,055	2004	8.6%	23.7%	40.9%	91.4%	76.3%	59.1%
2005	9,283,240	9,283,240	9,283,236	8,279,640	7,310,842	6,233,288	2005	10.8%	21.2%	32.9%	89.2%	78.8%	67.1%
2006	7,006,265	7,006,265	7,006,234	5,867,712	5,057,528	4,363,323	2006	16.3%	27.8%	37.7%	83.7%	72.2%	62.3%
2007	9,968,261	9,968,261	9,968,172	8,943,700	7,797,767	7,525,065	2007	10.3%	21.8%	24.5%	89.7%	78.2%	75.5%
2008	5,174,270	5,174,270	5,174,270	1,852,910	1,815,343	1,647,793	2008	64.2%	64.9%	68.2%	35.8%	35.1%	31.8%

Sector 17

	Industry Inefficiency (IE)			Aggregate Firm Inefficiency (ITE)				Industry Reallocation Inefficiency			Aggregate Firm Inefficiency		
Year	CCR	BCC	FRH	CCR	BCC	FDH	Year	CCR	BCC	FRH	CCR	BCC	FDH
2000	1,232,733	1,232,733	1,232,733	1,152,041	798,814	345,466	2000	6.5%	35.2%	72.0%	93.5%	64.8%	28.0%
2001	1,352,347	1,352,347	1,352,345	1,229,991	953,602	510,754	2001	9.0%	29.5%	62.2%	91.0%	70.5%	37.8%
2002	1,553,679	1,553,679	1,553,679	1,379,762	955,453	509,951	2002	11.2%	38.5%	67.2%	88.8%	61.5%	32.8%
2003	1,758,460	1,758,460	1,758,401	1,545,745	933,341	494,006	2003	12.1%	46.9%	71.9%	87.9%	53.1%	28.1%
2004	2,075,899	2,075,899	2,075,730	1,814,785	1,172,900	736,749	2004	12.6%	43.5%	64.5%	87.4%	56.5%	35.5%
2005	2,382,890	2,382,890	2,382,890	2,068,459	1,506,184	1,070,898	2005	13.2%	36.8%	55.1%	86.8%	63.2%	44.9%
2006	2,584,485	2,584,485	2,578,623	2,099,165	2,012,548	1,165,578	2006	18.8%	22.1%	54.9%	81.2%	77.9%	45.1%
2007	3,271,739	3,271,739	3,271,685	2,756,154	2,106,401	1,270,284	2007	15.8%	35.6%	61.2%	84.2%	64.4%	38.8%
2008	2,432,830	2,432,830	2,432,823	1,018,397	901,889	709,514	2008	58.1%	62.9%	70.8%	41.9%	37.1%	29.2%

Sector 18

	Industry Inefficiency (IE)			Aggregate Firm Inefficiency (ITE)				Industry Reallocation Inefficiency			Aggregate Firm Inefficiency		
Year	CCR	BCC	FRH	CCR	BCC	FDH	Year	CCR	BCC	FRH	CCR	BCC	FDH
2000	1,609,537	1,609,537	1,609,509	966,742	896,416	631,053	2000	39.9%	44.3%	60.8%	60.1%	55.7%	39.2%
2001	1,990,064	1,990,064	1,989,782	1,142,087	881,421	596,483	2001	42.6%	55.7%	70.0%	57.4%	44.3%	30.0%
2002	2,601,091	2,601,091	2,601,079	1,596,220	1,244,740	798,273	2002	38.6%	52.1%	69.3%	61.4%	47.9%	30.7%
2003	3,245,427	3,245,427	3,245,411	1,937,086	1,404,561	859,380	2003	40.3%	56.7%	73.5%	59.7%	43.3%	26.5%
2004	3,898,903	3,898,903	3,898,884	2,421,947	1,935,818	1,366,802	2004	37.9%	50.3%	64.9%	62.1%	49.7%	35.1%
2005	4,747,615	4,747,615	4,747,598	2,762,608	2,479,491	2,132,878	2005	41.8%	47.8%	55.1%	58.2%	52.2%	44.9%
2006	4,431,012	4,431,012	4,431,008	2,839,387	2,437,618	1,976,884	2006	35.9%	45.0%	55.4%	64.1%	55.0%	44.6%
2007	6,760,280	6,760,280	6,760,271	4,338,372	4,171,741	2,878,362	2007	35.8%	38.3%	57.4%	64.2%	61.7%	42.6%
2008	2,862,363	2,862,363	2,862,359	1,889,052	1,839,795	1,562,675	2008	34.0%	35.7%	45.4%	66.0%	64.3%	54.6%

Sector 19

	Industry Inefficiency (IE)			Aggregate Firm Inefficiency (ITE)				Industry Reallocation Inefficiency			Aggregate Firm Inefficiency		
Year	CCR	BCC	FRH	CCR	BCC	FDH	Year	CCR	BCC	FRH	CCR	BCC	FDH
2000	1,723,046	1,723,046	1,722,871	1,052,374	657,414	372,040	2000	38.9%	61.8%	78.4%	61.1%	38.2%	21.6%
2001	2,069,534	2,069,534	2,069,380	1,350,336	940,332	496,109	2001	34.8%	54.6%	76.0%	65.2%	45.4%	24.0%
2002	3,233,257	3,233,257	3,233,183	1,765,929	774,927	364,248	2002	45.4%	76.0%	88.7%	54.6%	24.0%	11.3%
2003	3,332,415	3,332,415	3,331,806	2,168,962	1,139,856	653,867	2003	34.9%	65.8%	80.4%	65.1%	34.2%	19.6%
2004	4,583,984	4,583,984	4,583,958	3,154,510	1,645,475	1,243,728	2004	31.2%	64.1%	72.9%	68.8%	35.9%	27.1%
2005	4,909,407	4,909,407	4,909,362	3,124,981	1,528,588	816,719	2005	36.3%	68.9%	83.4%	63.7%	31.1%	16.6%
2006	4,545,504	4,545,504	4,545,497	3,439,618	1,417,372	752,771	2006	24.3%	68.8%	83.4%	75.7%	31.2%	16.6%
2007	7,244,754	7,244,754	7,244,753	4,381,782	2,650,839	1,239,791	2007	39.5%	63.4%	82.9%	60.5%	36.6%	17.1%
2008	896,159	896,159	896,153	658,957	643,016	457,059	2008	26.5%	28.2%	49.0%	73.5%	71.8%	51.0%

Sector 20

	Industry Inefficiency (IE)			Aggregate Firm Inefficiency (ITE)				Industry Reallocation Inefficiency			Aggregate Firm Inefficiency		
Year	CCR	BCC	FRH	CCR	BCC	FDH	Year	CCR	BCC	FRH	CCR	BCC	FDH
2000	620,967	620,967	620,953	474,746	377,710	289,578	2000	23.5%	39.2%	53.4%	76.5%	60.8%	46.6%
2001	454,745	454,745	454,737	353,815	328,981	246,499	2001	22.2%	27.7%	45.8%	77.8%	72.3%	54.2%
2002	665,085	665,085	665,083	662,996	571,019	491,041	2002	0.3%	14.1%	26.2%	99.7%	85.9%	73.8%
2003	963,952	963,952	963,948	662,996	571,019	491,041	2003	31.2%	40.8%	49.1%	68.8%	59.2%	50.9%
2004	998,401	998,401	998,393	740,503	578,739	465,936	2004	25.8%	42.0%	53.3%	74.2%	58.0%	46.7%
2005	1,262,835	1,262,835	1,262,823	980,534	895,411	778,176	2005	22.4%	29.1%	38.4%	77.6%	70.9%	61.6%
2006	1,297,974	1,297,974	1,297,928	916,900	810,267	629,802	2006	29.4%	37.6%	51.5%	70.6%	62.4%	48.5%
2007	1,710,745	1,710,745	1,710,732	1,253,302	1,196,393	1,057,054	2007	26.7%	30.1%	38.2%	73.3%	69.9%	61.8%
2008	1,566,028	1,566,028	1,565,952	1,225,817	631,687	472,771	2008	21.7%	59.7%	69.8%	78.3%	40.3%	30.2%

Sector 21

	Industry Inefficiency (IE)			Aggregate Firm Inefficiency (ITE)				Industry Reallocation Inefficiency			Aggregate Firm Inefficiency		
Year	CCR	BCC	FRH	CCR	BCC	FDH	Year	CCR	BCC	FRH	CCR	BCC	FDH
2000	275,949	275,949	275,906	249,714	163,305	81,016	2000	9.5%	40.8%	70.6%	90.5%	59.2%	29.4%
2001	866,620	866,620	866,582	719,448	503,266	358,461	2001	17.0%	41.9%	58.6%	83.0%	58.1%	41.4%
2002	667,688	667,688	667,650	595,300	465,184	399,245	2002	10.8%	30.3%	40.2%	89.2%	69.7%	59.8%
2003	684,708	684,708	684,699	590,669	452,706	319,644	2003	13.7%	33.9%	53.3%	86.3%	66.1%	46.7%
2004	934,962	934,962	934,961	805,878	622,088	438,397	2004	13.8%	33.5%	53.1%	86.2%	66.5%	46.9%
2005	1,172,592	1,172,592	1,172,585	931,899	755,517	534,074	2005	20.5%	35.6%	54.5%	79.5%	64.4%	45.5%
2006	1,148,496	1,148,496	1,148,483	1,012,439	666,444	472,115	2006	11.8%	42.0%	58.9%	88.2%	58.0%	41.1%
2007	1,792,724	1,792,724	1,792,722	1,584,801	1,196,132	836,709	2007	11.6%	33.3%	53.3%	88.4%	66.7%	46.7%
2008	1,810,194	1,810,194	1,810,044	521,861	506,632	405,337	2008	71.2%	72.0%	77.6%	28.8%	28.0%	22.4%

Sector 22

	Industry Inefficiency (IE)			Aggregate Firm Inefficiency (ITE)				Industry Reallocation Inefficiency					
Year	CCR	BCC	FRH	CCR	BCC	FDH	Year	CCR	BCC	FRH	CCR	BCC	FDH
2000	363,483	363,483	363,179	272,303	226,042	162,045	2000	25.1%	37.8%	55.4%	74.9%	62.2%	44.6%
2001	866,620	866,620	866,582	719,448	503,288	358,461	2001	17.0%	41.9%	58.6%	83.0%	58.1%	41.4%
2002	555,926	555,926	555,924	412,651	314,734	193,726	2002	25.8%	43.4%	65.2%	74.2%	56.6%	34.8%
2003	653,265	653,265	653,213	470,702	384,400	201,186	2003	27.9%	41.2%	69.2%	72.1%	58.8%	30.8%
2004	861,945	861,945	861,940	620,751	400,675	230,405	2004	28.0%	53.5%	73.3%	72.0%	46.5%	26.7%
2005	999,619	999,619	999,585	689,778	498,861	275,950	2005	31.0%	50.1%	72.4%	69.0%	49.9%	27.6%
2006	1,063,161	1,063,161	1,063,123	734,019	505,146	260,905	2006	31.0%	52.5%	75.5%	69.0%	47.5%	24.5%
2007	1,469,579	1,469,579	1,469,557	1,037,098	799,527	491,775	2007	29.4%	45.6%	66.5%	70.6%	54.4%	33.5%
2008	1,418,072	1,418,072	1,417,389	720,503	681,855	484,049	2008	49.2%	51.9%	65.9%	50.8%	48.1%	34.1%

Sector 24

	Industry Inefficiency (IE)			Aggregate Firm Inefficiency (ITE)				Industry Reallocation Inefficiency			Aggregate Firm Inefficiency		
Year	CCR	BCC	FRH	CCR	BCC	FDH	Year	CCR	BCC	FRH	CCR	BCC	FDH
2000	1,043,362	1,043,362	1,043,207	953,475	796,035	373,694	2000	8.6%	23.7%	64.2%	91.4%	76.3%	35.8%
2001	1,199,894	1,199,894	1,199,882	1,107,747	894,181	364,924	2001	7.7%	25.5%	69.6%	92.3%	74.5%	30.4%
2002	1,473,325	1,473,325	1,473,263	1,365,624	1,121,583	605,902	2002	7.3%	23.9%	58.9%	92.7%	76.1%	41.1%
2003	1,805,189	1,805,189	1,805,114	1,689,185	1,397,499	980,284	2003	6.4%	22.6%	45.7%	93.6%	77.4%	54.3%
2004	2,214,746	2,214,746	2,214,564	2,114,002	1,639,794	1,167,968	2004	4.5%	26.0%	47.3%	95.5%	74.0%	52.7%
2005	2,522,780	2,522,780	2,522,663	2,347,241	1,941,652	1,182,457	2005	7.0%	23.0%	53.1%	93.0%	77.0%	46.9%
2006	3,193,722	3,193,722	3,193,663	2,754,149	2,305,881	1,451,798	2006	13.8%	27.8%	54.5%	86.2%	72.2%	45.5%
2007	4,176,795	4,176,795	4,176,767	3,689,089	3,109,457	2,256,897	2007	11.7%	25.6%	46.0%	88.3%	74.4%	54.0%
2008	3,528,423	3,528,423	3,528,395	1,311,296	1,058,691	944,156	2008	62.8%	70.0%	73.2%	37.2%	30.0%	26.8%

Sector 25

	Industry Inefficiency (IE)			Aggregate Firm Inefficiency (ITE)				Industry Reallocation Inefficiency			Aggregate Firm Inefficiency		
Year	CCR	BCC	FRH	CCR	BCC	FDH	Year	CCR	BCC	FRH	CCR	BCC	FDH
2000	702,582	702,582	702,524	634,638	500,896	372,773	2000	9.7%	28.7%	46.9%	90.3%	71.3%	53.1%
2001	794,182	794,182	793,697	656,864	575,234	390,059	2001	17.3%	27.6%	50.9%	82.7%	72.4%	49.1%
2002	1,067,721	1,067,721	1,067,655	932,873	839,961	584,302	2002	12.6%	21.3%	45.3%	87.4%	78.7%	54.7%
2003	1,324,802	1,324,802	1,324,781	1,175,224	1,059,756	825,065	2003	11.3%	20.0%	37.7%	88.7%	80.0%	62.3%
2004	1,799,233	1,799,233	1,799,203	1,474,138	1,381,306	1,180,810	2004	18.1%	23.2%	34.4%	81.9%	76.8%	65.6%
2005	1,938,479	1,938,479	1,938,455	1,729,687	1,557,975	1,236,305	2005	10.8%	19.6%	36.2%	89.2%	80.4%	63.8%
2006	2,088,962	2,088,962	2,088,739	1,725,314	1,587,974	1,230,578	2006	17.4%	24.0%	41.1%	82.6%	76.0%	58.9%
2007	3,396,017	3,396,017	3,395,987	2,847,712	2,567,580	1,965,269	2007	16.1%	24.4%	42.1%	83.9%	75.6%	57.9%
2008	3,213,015	3,213,015	3,212,973	1,105,827	1,066,805	885,205	2008	65.6%	66.8%	72.4%	34.4%	33.2%	27.6%

Sector 26

Year	Industry Inefficiency (IE)			Aggregate Firm Inefficiency (ITE)			Year	Industry Reallocation Inefficiency			Aggregate Firm Inefficiency		
	CCR	BCC	FRH	CCR	BCC	FDH		CCR	BCC	FRH	CCR	BCC	FDH
2000	1,023,969	1,023,969	1,023,960	902,121	537,918	361,425	2000	11.9%	47.5%	64.7%	88.1%	52.5%	35.3%
2001	2,699,898	2,699,898	2,699,747	2,287,799	2,165,294	2,158,565	2001	15.3%	19.8%	20.1%	84.7%	80.2%	79.9%
2002	2,343,688	2,343,688	2,343,665	1,900,576	1,341,054	867,103	2002	18.9%	42.8%	63.0%	81.1%	57.2%	37.0%
2003	2,870,755	2,870,755	2,870,732	2,405,928	1,716,030	1,202,725	2003	16.2%	40.2%	58.1%	83.8%	59.8%	41.9%
2004	3,260,433	3,260,433	3,260,422	2,725,620	2,121,944	1,554,532	2004	16.4%	34.9%	52.3%	83.6%	65.1%	47.7%
2005	3,688,750	3,688,750	3,688,733	3,225,023	2,438,392	1,685,106	2005	12.6%	33.9%	54.3%	87.4%	66.1%	45.7%
2006	3,942,767	3,942,767	3,942,709	3,317,028	2,397,356	1,592,043	2006	15.9%	39.2%	59.6%	84.1%	60.8%	40.4%
2007	5,628,143	5,628,143	5,628,121	4,814,733	3,569,596	2,293,592	2007	14.5%	36.6%	59.2%	85.5%	63.4%	40.8%
2008	4,514,757	4,514,757	4,514,729	1,807,430	1,184,187	950,330	2008	60.0%	73.8%	79.0%	40.0%	26.2%	21.0%

Sector 28

Year	Industry Inefficiency (IE)			Aggregate Firm Inefficiency (ITE)			Year	Industry Reallocation Inefficiency			Aggregate Firm Inefficiency		
	CCR	BCC	FRH	CCR	BCC	FDH		CCR	BCC	FRH	CCR	BCC	FDH
2000	592,360	592,360	592,322	468,590	400,596	286,661	2000	20.9%	32.4%	51.6%	79.1%	67.6%	48.4%
2001	728,255	728,255	728,242	615,008	568,666	480,336	2001	15.6%	21.9%	34.0%	84.4%	78.1%	66.0%
2002	968,831	968,831	968,814	796,283	708,368	569,902	2002	17.8%	26.9%	41.2%	82.2%	73.1%	58.8%
2003	1,274,720	1,274,720	1,274,691	1,048,464	993,766	745,781	2003	17.7%	22.0%	41.5%	82.3%	78.0%	58.5%
2004	1,644,522	1,644,522	1,644,489	1,415,197	1,331,489	1,141,884	2004	13.9%	19.0%	30.6%	86.1%	81.0%	69.4%
2005	2,090,754	2,090,754	2,090,708	1,833,183	1,675,839	1,377,487	2005	12.3%	19.8%	34.1%	87.7%	80.2%	65.9%
2006	2,212,301	2,212,301	2,212,192	1,919,283	1,728,415	1,212,578	2006	13.2%	21.9%	45.2%	86.8%	78.1%	54.8%
2007	3,962,633	3,962,633	3,962,582	3,389,447	3,215,927	2,558,114	2007	14.5%	18.8%	35.4%	85.5%	81.2%	64.6%
2008	4,010,076	4,010,076	4,009,628	1,601,675	1,566,777	1,368,258	2008	60.1%	60.9%	65.9%	39.9%	39.1%	34.1%

Sector 29

	Industry Inefficiency (IE)			Aggregate Firm Inefficiency (ITE)				Industry Reallocation Inefficiency			Aggregate Firm Inefficiency		
Year	CCR	BCC	FRH	CCR	BCC	FDH	Year	CCR	BCC	FRH	CCR	BCC	FDH
2000	340,888	340,888	340,883	322,552	245,344	170,261	2000	5.4%	28.0%	50.1%	94.6%	72.0%	49.9%
2001	516,117	516,117	515,719	453,876	435,310	283,387	2001	12.1%	15.7%	45.1%	87.9%	84.3%	54.9%
2002	549,155	549,155	549,028	484,170	456,443	347,997	2002	11.8%	16.9%	36.6%	88.2%	83.1%	63.4%
2003	710,986	710,986	710,873	636,840	613,963	439,994	2003	10.4%	13.6%	38.1%	89.6%	86.4%	61.9%
2004	796,296	796,296	796,279	753,330	653,735	460,323	2004	5.4%	17.9%	42.2%	94.6%	82.1%	57.8%
2005	858,203	858,203	858,137	795,110	649,884	407,551	2005	7.4%	24.3%	52.5%	92.6%	75.7%	47.5%
2006	771,883	771,883	769,954	677,986	640,279	385,880	2006	12.2%	17.0%	50.0%	87.8%	83.0%	50.0%
2007	1,451,512	1,451,512	1,451,475	1,357,057	1,308,876	1,137,727	2007	6.5%	9.8%	21.6%	93.5%	90.2%	78.4%
2008	1,380,017	1,380,017	1,379,244	493,107	476,435	327,708	2008	64.3%	65.5%	76.3%	35.7%	34.5%	23.7%

Sector 36

	Industry Inefficiency (IE)			Aggregate Firm Inefficiency (ITE)				Industry Reallocation Inefficiency			Aggregate Firm Inefficiency		
Year	CCR	BCC	FRH	CCR	BCC	FDH	Year	CCR	BCC	FRH	CCR	BCC	FDH
2000	599,217	599,217	599,045	403,033	341,076	194,012	2000	32.7%	43.1%	67.6%	67.3%	56.9%	32.4%
2001	849,157	849,157	849,146	699,817	593,546	421,311	2001	17.6%	30.1%	50.4%	82.4%	69.9%	49.6%
2002	1,317,998	1,317,998	1,317,929	1,063,913	921,845	697,726	2002	19.3%	30.1%	47.1%	80.7%	69.9%	52.9%
2003	1,903,470	1,903,470	1,903,301	1,521,766	1,264,369	984,206	2003	20.1%	33.6%	48.3%	79.9%	66.4%	51.7%
2004	2,833,088	2,833,088	2,833,070	2,403,746	2,146,593	1,857,009	2004	15.2%	24.2%	34.5%	84.8%	75.8%	65.5%
2005	3,758,619	3,758,619	3,758,613	3,270,593	2,986,408	2,183,026	2005	13.0%	20.5%	41.9%	87.0%	79.5%	58.1%
2006	3,730,984	3,730,984	3,730,895	3,072,099	2,693,170	1,811,452	2006	17.7%	27.8%	51.4%	82.3%	72.2%	48.6%
2007	6,648,357	6,648,357	6,648,274	5,419,674	4,612,549	3,567,689	2007	18.5%	30.6%	46.3%	81.5%	69.4%	53.7%
2008	3,330,249	3,330,249	3,330,249	1,568,096	1,500,188	1,273,723	2008	52.9%	55.0%	61.8%	47.1%	45.0%	38.2%

Table 17: Industry Inefficiency - Meta Estimates

Sector 14

Group	1-9			10-49			50-99			100-199			>200		
Year	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR
2000	0.471	0.092	0.19	0.058	0.051	0.88	0.140	0.064	0.45	0.291	0.037	0.13	0.451	0.054	0.12
2001	0.299	0.104	0.35	0.025	0.025	1.00	0.232	0.010	0.04	0.230	0.005	0.02	0.073	0.014	0.20
2002	0.191	0.072	0.38	0.097	0.054	0.56	0.179	0.055	0.31	0.176	0.046	0.26	0.286	0.043	0.15
2003	0.262	0.056	0.21	0.054	0.054	1.00	0.161	0.055	0.34	0.182	0.035	0.19	0.179	0.039	0.22
2004	0.154	0.151	0.98	0.132	0.115	0.87	0.174	0.142	0.82	0.143	0.095	0.67	0.186	0.096	0.52
2005	0.146	0.073	0.50	0.065	0.050	0.77	0.063	0.058	0.92	0.127	0.027	0.21	0.345	0.029	0.08
2006	0.188	0.120	0.64	0.081	0.078	0.97	0.130	0.107	0.83	0.180	0.051	0.28	0.295	0.054	0.18
2007	0.181	0.085	0.47	0.120	0.085	0.71	0.169	0.129	0.76	0.135	0.080	0.60	0.327	0.085	0.26
2008	0.109	0.050	0.46	0.061	0.059	0.97	0.213	0.062	0.29	0.145	0.062	0.43	0.217	0.068	0.32

Sector 15

Group	1-9			10-49			50-99			100-199			>200		
Year	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR
2000	0.108	0.107	0.99	0.061	0.061	0.99	0.098	0.044	0.44	0.099	0.033	0.34	0.192	0.031	0.16
2001	0.063	0.054	0.87	0.045	0.044	0.98	0.187	0.035	0.19	0.195	0.024	0.12	0.195	0.023	0.12
2002	0.089	0.068	0.76	0.045	0.043	0.96	0.147	0.036	0.24	0.146	0.027	0.18	0.250	0.025	0.10
2003	0.046	0.046	1.00	0.062	0.047	0.75	0.175	0.033	0.19	0.191	0.031	0.16	0.265	0.030	0.11
2004	0.032	0.027	0.85	0.033	0.028	0.84	0.099	0.021	0.21	0.127	0.016	0.13	0.099	0.017	0.17
2005	0.015	0.015	1.00	0.021	0.012	0.57	0.096	0.010	0.11	0.212	0.008	0.04	0.114	0.008	0.07
2006	0.085	0.074	0.87	0.088	0.080	0.91	0.091	0.061	0.67	0.072	0.055	0.76	0.226	0.054	0.24
2007	0.075	0.013	0.18	0.011	0.011	1.00	0.085	0.009	0.11	0.069	0.007	0.10	0.244	0.009	0.04
2008	0.076	0.061	0.80	0.024	0.024	1.00	0.139	0.020	0.15	0.132	0.019	0.15	0.184	0.020	0.11

Sector 17

Group	1-9			10-49			50-99			100-199			>200		
Year	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR
2000	0.212	0.212	1.00	0.190	0.097	0.51	0.129	0.044	0.34	0.336	0.048	0.14	0.117	0.038	0.32
2001	0.289	0.188	0.65	0.158	0.158	1.00	0.221	0.111	0.50	0.308	0.124	0.40	0.297	0.103	0.35
2002	0.150	0.149	0.99	0.209	0.144	0.69	0.285	0.133	0.47	0.276	0.120	0.44	0.320	0.102	0.32
2003	0.349	0.270	0.77	0.188	0.188	1.00	0.263	0.120	0.46	0.281	0.160	0.57	0.336	0.128	0.38
2004	0.255	0.120	0.47	0.060	0.060	1.00	0.254	0.046	0.18	0.345	0.058	0.17	0.255	0.044	0.17
2005	0.069	0.031	0.44	0.047	0.047	1.00	0.238	0.037	0.16	0.203	0.036	0.18	0.265	0.039	0.15
2006	0.167	0.142	0.85	0.128	0.120	0.94	0.238	0.129	0.54	0.330	0.123	0.37	0.284	0.201	0.71
2007	0.165	0.160	0.97	0.134	0.126	0.94	0.171	0.111	0.65	0.233	0.126	0.54	0.362	0.134	0.37
2008	0.183	0.181	0.99	0.171	0.168	0.98	0.078	0.061	0.79	0.232	0.127	0.55	0.312	0.175	0.56

Sector 18

Group	1-9			10-49			50-99			100-199			>200		
Year	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR
2000	0.422	0.255	0.60	0.113	0.112	0.99	0.211	0.096	0.45	0.320	0.085	0.27	0.180	0.080	0.45
2001	0.414	0.300	0.73	0.181	0.131	0.72	0.199	0.165	0.83	0.304	0.129	0.42	0.166	0.129	0.78
2002	0.111	0.111	1.00	0.157	0.096	0.61	0.193	0.085	0.44	0.161	0.081	0.50	0.228	0.091	0.40
2003	0.229	0.151	0.66	0.124	0.116	0.93	0.371	0.078	0.21	0.101	0.091	0.91	0.223	0.107	0.48
2004	0.125	0.123	0.98	0.132	0.121	0.91	0.104	0.082	0.79	0.223	0.107	0.48	0.225	0.102	0.46
2005	0.105	0.103	0.98	0.139	0.094	0.68	0.251	0.033	0.13	0.192	0.033	0.17	0.044	0.040	0.91
2006	0.146	0.109	0.75	0.205	0.136	0.66	0.289	0.122	0.42	0.216	0.100	0.46	0.127	0.113	0.89
2007	0.184	0.131	0.71	0.132	0.101	0.76	0.206	0.092	0.45	0.248	0.089	0.36	0.179	0.120	0.67
2008	0.333	0.158	0.47	0.174	0.154	0.88	0.197	0.107	0.55	0.181	0.160	0.88	0.238	0.198	0.83

Sector 19

Group	1-9			10-49			50-99		
Year	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR
2000	0.638	0.413	0.65	0.241	0.241	1.00	0.403	0.228	0.57
2001	0.956	0.790	0.83	0.122	0.101	0.83	0.222	0.212	0.95
2002	0.166	0.166	1.00	0.086	0.086	1.00	0.512	0.097	0.19
2003	0.394	0.354	0.90	0.266	0.210	0.79	0.483	0.198	0.41
2004	0.511	0.233	0.45	0.141	0.141	1.00	0.361	0.022	0.06
2005	0.249	0.141	0.57	0.234	0.234	1.00	0.472	0.186	0.39
2006	0.148	0.121	0.81	0.173	0.142	0.82	0.266	0.143	0.54
2007	0.181	0.181	1.00	0.300	0.149	0.50	0.410	0.216	0.53
2008	0.414	0.207	0.50	0.193	0.182	0.94	0.219	0.191	0.87

Group	100-199			>200		
Year	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR
2000	0.496	0.239	0.48	0.317	0.105	0.33
2001	0.544	0.165	0.30	0.346	0.214	0.62
2002	0.494	0.097	0.20	0.388	0.100	0.26
2003	0.514	0.160	0.31	0.386	0.271	0.70
2004	0.608	0.033	0.05	0.225	0.063	0.28
2005	0.619	0.258	0.42	0.372	0.197	0.53
2006	0.586	0.142	0.24	0.316	0.119	0.38
2007	0.532	0.221	0.42	0.313	0.186	0.59
2008	0.350	0.175	0.50	0.292	0.180	0.62

Sector 20

Group	1-9			10-49			50-99		
Year	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR
2000	0.087	0.087	1.00	0.088	0.044	0.50	0.166	0.026	0.16
2001	0.069	0.069	1.00	0.076	0.055	0.72	0.141	0.047	0.34
2002	0.057	0.057	1.00	0.118	0.093	0.79	0.260	0.052	0.20
2003	0.115	0.087	0.76	0.041	0.040	0.99	0.206	0.037	0.18
2004	0.106	0.106	1.00	0.089	0.058	0.65	0.149	0.050	0.34
2005	0.052	0.052	1.00	0.066	0.051	0.78	0.169	0.043	0.26
2006	0.100	0.088	0.89	0.081	0.062	0.76	0.254	0.067	0.27
2007	0.177	0.143	0.81	0.041	0.041	1.00	0.101	0.055	0.54
2008	0.176	0.113	0.64	0.084	0.075	0.89	0.181	0.083	0.46

Group	100-199			>200		
Year	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR
2000	0.302	0.029	0.09	0.205	0.061	0.30
2001	0.457	0.058	0.13	0.140	0.069	0.49
2002	0.194	0.045	0.23	0.278	0.052	0.19
2003	0.309	0.031	0.10	0.223	0.032	0.15
2004	0.179	0.047	0.26	0.332	0.052	0.16
2005	0.205	0.041	0.20	0.265	0.052	0.20
2006	0.201	0.057	0.28	0.423	0.062	0.15
2007	0.202	0.047	0.24	0.187	0.039	0.21
2008	0.261	0.086	0.33	0.136	0.051	0.38

Sector 21

Group	1-9			10-49			50-99			100-199			>200		
Year	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR
2000	0.374	0.287	0.77	0.211	0.207	0.98	0.383	0.270	0.71	0.352	0.207	0.59	0.513	0.223	0.43
2001	0.195	0.120	0.62	0.204	0.158	0.77	0.076	0.074	0.98	0.283	0.070	0.25	0.506	0.056	0.11
2002	0.340	0.221	0.65	0.061	0.051	0.82	0.063	0.047	0.75	0.366	0.037	0.10	0.343	0.028	0.08
2003	0.104	0.104	1.00	0.160	0.094	0.59	0.296	0.053	0.18	0.512	0.054	0.11	0.175	0.042	0.24
2004	0.195	0.132	0.68	0.128	0.128	1.00	0.266	0.119	0.45	0.392	0.107	0.27	0.386	0.113	0.29
2005	0.145	0.084	0.58	0.122	0.095	0.78	0.290	0.081	0.28	0.387	0.066	0.17	0.365	0.081	0.22
2006	0.050	0.050	1.00	0.135	0.069	0.51	0.337	0.068	0.20	0.356	0.062	0.17	0.306	0.053	0.17
2007	0.208	0.172	0.83	0.090	0.068	0.76	0.283	0.054	0.19	0.214	0.068	0.32	0.059	0.057	0.97
2008	0.234	0.100	0.43	0.081	0.081	1.00	0.297	0.079	0.27	0.363	0.070	0.19	0.417	0.109	0.26

Sector 22

Group	1-9			10-49			50-99			100-199			>200		
Year	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR
2000	0.327	0.317	0.97	0.366	0.261	0.71	0.167	0.141	0.84	0.629	0.130	0.21	0.275	0.102	0.37
2001	0.195	0.120	0.62	0.204	0.158	0.77	0.076	0.074	0.98	0.283	0.070	0.25	0.506	0.056	0.11
2002	0.108	0.098	0.91	0.093	0.064	0.69	0.137	0.086	0.63	0.287	0.076	0.27	0.594	0.077	0.13
2003	0.168	0.169	1.00	0.285	0.145	0.51	0.314	0.163	0.52	0.507	0.139	0.27	0.586	0.138	0.24
2004	0.089	0.082	0.92	0.085	0.076	0.89	0.290	0.056	0.19	0.347	0.064	0.18	0.670	0.087	0.13
2005	0.077	0.077	1.00	0.204	0.167	0.82	0.415	0.056	0.14	0.463	0.068	0.15	0.533	0.081	0.15
2006	0.142	0.142	1.00	0.233	0.138	0.59	0.322	0.128	0.40	0.538	0.144	0.27	0.600	0.124	0.21
2007	0.128	0.127	0.99	0.111	0.099	0.89	0.362	0.101	0.28	0.528	0.110	0.21	0.504	0.119	0.24
2008	0.203	0.119	0.59	0.215	0.176	0.82	0.381	0.104	0.27	0.109	0.660	0.60	0.362	0.241	0.67

Sector 24

Group	1-9			10-49			50-99			100-199			>200		
Year	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR
2000	0.403	0.212	0.53	0.265	0.232	0.88	0.305	0.209	0.69	0.372	0.261	0.70	0.561	0.260	0.46
2001	0.306	0.183	0.60	0.070	0.068	0.98	0.419	0.070	0.17	0.477	0.092	0.19	0.517	0.089	0.17
2002	0.180	0.103	0.57	0.153	0.153	1.00	0.365	0.185	0.51	0.464	0.126	0.27	0.455	0.172	0.38
2003	0.207	0.165	0.80	0.102	0.101	0.99	0.182	0.114	0.63	0.434	0.093	0.21	0.500	0.125	0.25
2004	0.158	0.094	0.60	0.092	0.092	1.00	0.146	0.088	0.60	0.428	0.113	0.27	0.409	0.094	0.23
2005	0.130	0.110	0.85	0.202	0.202	1.00	0.412	0.229	0.56	0.437	0.244	0.56	0.424	0.192	0.45
2006	0.103	0.082	0.80	0.186	0.181	0.97	0.354	0.089	0.25	0.521	0.095	0.18	0.457	0.072	0.16
2007	0.170	0.121	0.71	0.112	0.112	1.00	0.380	0.111	0.29	0.302	0.116	0.39	0.500	0.117	0.23
2008	0.211	0.149	0.70	0.092	0.090	0.98	0.213	0.094	0.44	0.489	0.099	0.20	0.470	0.119	0.25

Sector 25

Group	1-9			10-49			50-99			100-199			>200		
Year	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR
2000	0.153	0.150	0.98	0.131	0.125	0.95	0.206	0.107	0.52	0.474	0.093	0.20	0.436	0.088	0.20
2001	0.301	0.242	0.80	0.212	0.211	1.00	0.287	0.202	0.70	0.482	0.276	0.57	0.358	0.193	0.54
2002	0.232	0.177	0.77	0.102	0.102	1.00	0.267	0.090	0.34	0.354	0.118	0.33	0.431	0.109	0.25
2003	0.084	0.060	0.71	0.072	0.062	0.87	0.208	0.049	0.23	0.168	0.073	0.43	0.342	0.046	0.13
2004	0.215	0.151	0.70	0.077	0.077	0.99	0.158	0.074	0.46	0.113	0.078	0.69	0.336	0.075	0.22
2005	0.158	0.091	0.57	0.064	0.063	1.00	0.223	0.059	0.26	0.298	0.059	0.20	0.386	0.053	0.14
2006	0.265	0.182	0.69	0.088	0.088	1.00	0.261	0.082	0.31	0.229	0.112	0.49	0.384	0.098	0.25
2007	0.153	0.135	0.88	0.140	0.133	0.95	0.268	0.112	0.42	0.281	0.163	0.58	0.448	0.117	0.26
2008	0.200	0.155	0.78	0.164	0.132	0.81	0.127	0.125	0.98	0.276	0.143	0.52	0.361	0.110	0.30

Sector 26

Group	1-9			10-49			50-99			100-199			>200		
Year	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR
2000	0.321	0.188	0.59	0.091	0.091	0.99	0.151	0.064	0.43	0.360	0.060	0.17	0.440	0.086	0.19
2001	0.145	0.121	0.83	0.003	0.003	1.00	0.235	0.101	0.43	0.320	0.001	0.00	0.271	0.000	0.00
2002	0.258	0.137	0.53	0.093	0.093	0.99	0.348	0.114	0.33	0.257	0.145	0.56	0.342	0.154	0.45
2003	0.258	0.158	0.61	0.215	0.144	0.67	0.203	0.195	0.96	0.249	0.210	0.84	0.340	0.147	0.43
2004	0.202	0.100	0.50	0.094	0.094	1.00	0.283	0.129	0.46	0.299	0.134	0.45	0.249	0.139	0.56
2005	0.074	0.058	0.78	0.100	0.072	0.72	0.158	0.081	0.51	0.270	0.091	0.33	0.305	0.082	0.27
2006	0.291	0.117	0.40	0.097	0.094	0.97	0.282	0.105	0.37	0.334	0.122	0.36	0.369	0.105	0.29
2007	0.033	0.032	0.96	0.081	0.045	0.56	0.192	0.019	0.10	0.240	0.020	0.08	0.381	0.025	0.07
2008	0.169	0.085	0.50	0.085	0.085	1.00	0.254	0.112	0.44	0.319	0.131	0.41	0.253	0.115	0.45

Sector 28

Group	1-9			10-49			50-99			100-199			>200		
Year	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR
2000	0.183	0.183	1.00	0.236	0.189	0.80	0.365	0.187	0.51	0.467	0.205	0.44	0.379	0.154	0.41
2001	0.255	0.242	0.95	0.177	0.165	0.93	0.254	0.170	0.67	0.445	0.147	0.33	0.199	0.115	0.57
2002	0.137	0.119	0.87	0.173	0.116	0.67	0.325	0.121	0.37	0.452	0.111	0.24	0.364	0.089	0.25
2003	0.158	0.119	0.75	0.112	0.100	0.90	0.254	0.101	0.40	0.465	0.109	0.23	0.215	0.091	0.42
2004	0.057	0.047	0.82	0.079	0.037	0.46	0.111	0.048	0.43	0.278	0.048	0.17	0.311	0.039	0.13
2005	0.087	0.063	0.73	0.068	0.067	0.98	0.265	0.063	0.24	0.350	0.068	0.20	0.146	0.060	0.41
2006	0.223	0.108	0.49	0.101	0.100	1.00	0.289	0.113	0.39	0.395	0.134	0.34	0.306	0.107	0.35
2007	0.167	0.101	0.60	0.127	0.124	0.98	0.226	0.130	0.58	0.271	0.142	0.52	0.190	0.116	0.61
2008	0.155	0.048	0.31	0.089	0.089	1.00	0.156	0.062	0.40	0.152	0.075	0.49	0.215	0.067	0.31

Sector 29

Group	1-9			10-49			50-99			100-199			>200		
Year	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR
2000	0.229	0.209	0.91	0.145	0.142	0.98	0.493	0.179	0.36	0.435	0.156	0.36	0.582	0.167	0.29
2001	0.340	0.313	0.92	0.394	0.213	0.54	0.366	0.251	0.69	0.460	0.214	0.46	0.220	0.200	0.91
2002	0.371	0.295	0.80	0.235	0.223	0.95	0.334	0.293	0.88	0.396	0.208	0.52	0.213	0.196	0.92
2003	0.268	0.236	0.88	0.236	0.190	0.81	0.441	0.188	0.43	0.279	0.232	0.83	0.231	0.160	0.69
2004	0.172	0.135	0.79	0.207	0.139	0.67	0.346	0.148	0.43	0.396	0.208	0.53	0.312	0.178	0.57
2005	0.133	0.133	1.00	0.239	0.140	0.59	0.313	0.130	0.42	0.312	0.194	0.62	0.582	0.166	0.28
2006	0.185	0.166	0.90	0.246	0.231	0.94	0.355	0.247	0.70	0.491	0.437	0.89	0.386	0.241	0.62
2007	0.114	0.095	0.83	0.005	0.005	0.98	0.265	0.008	0.03	0.426	0.007	0.02	0.347	0.007	0.02
2008	0.138	0.128	0.93	0.223	0.121	0.54	0.300	0.130	0.43	0.424	0.142	0.33	0.362	0.145	0.40

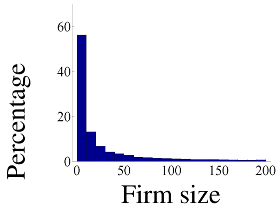
Sector 36

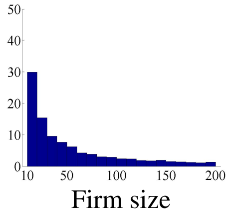
Group	1-9			10-49			50-99			100-199			>200		
Year	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR	IE_group	IE_Meta	MTR
2000	0.173	0.144	0.83	0.177	0.168	0.95	0.305	0.166	0.55	0.395	0.238	0.60	0.489	0.150	0.31
2001	0.143	0.137	0.96	0.181	0.081	0.45	0.217	0.083	0.38	0.285	0.085	0.30	0.325	0.080	0.25
2002	0.181	0.136	0.75	0.102	0.091	0.89	0.120	0.107	0.89	0.394	0.185	0.47	0.293	0.103	0.35
2003	0.184	0.119	0.64	0.080	0.080	1.00	0.174	0.085	0.49	0.249	0.088	0.35	0.221	0.091	0.41
2004	0.157	0.134	0.85	0.043	0.043	1.00	0.314	0.042	0.13	0.222	0.055	0.25	0.121	0.054	0.45
2005	0.092	0.086	0.94	0.115	0.086	0.75	0.190	0.094	0.50	0.261	0.094	0.36	0.266	0.089	0.34
2006	0.148	0.094	0.64	0.123	0.122	1.00	0.291	0.163	0.56	0.226	0.147	0.65	0.343	0.159	0.46
2007	0.114	0.052	0.45	0.017	0.017	1.00	0.134	0.017	0.12	0.262	0.016	0.06	0.254	0.017	0.07
2008	0.163	0.078	0.48	0.074	0.062	0.84	0.145	0.063	0.43	0.077	0.065	0.85	0.269	0.058	0.21

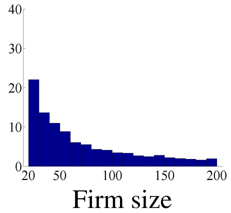
Figure 7.1: Distribution of Firm Size Measured by Number of Employees

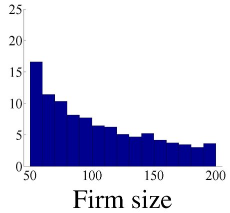
• Year 2000

$0 < \text{Size} \leq 200$ | $10 < \text{Size} \leq 200$ | $20 < \text{Size} \leq 200$ | $50 < \text{Size} \leq 200$

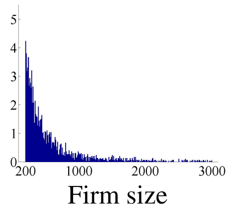






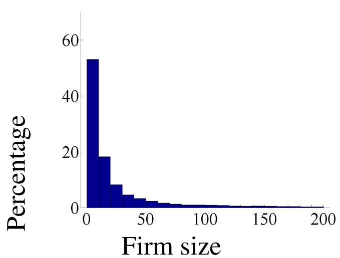


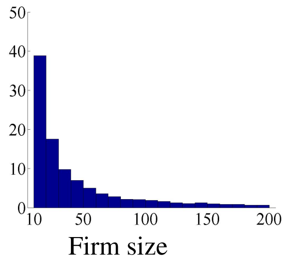
$200 < \text{Size} \leq 3000$

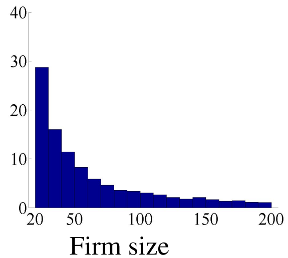


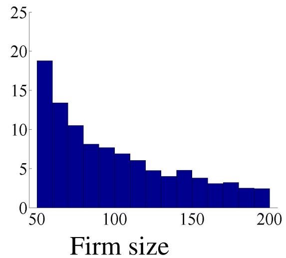
• Year 2004

$0 < \text{Size} \leq 200$ | $10 < \text{Size} \leq 200$ | $20 < \text{Size} \leq 200$ | $50 < \text{Size} \leq 200$









$200 < \text{Size} \leq 3000$

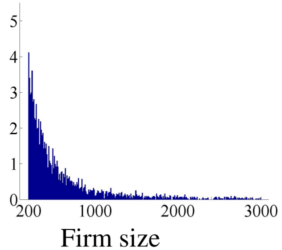


Figure 7.2: Distribution of Employment Share by Firm Size

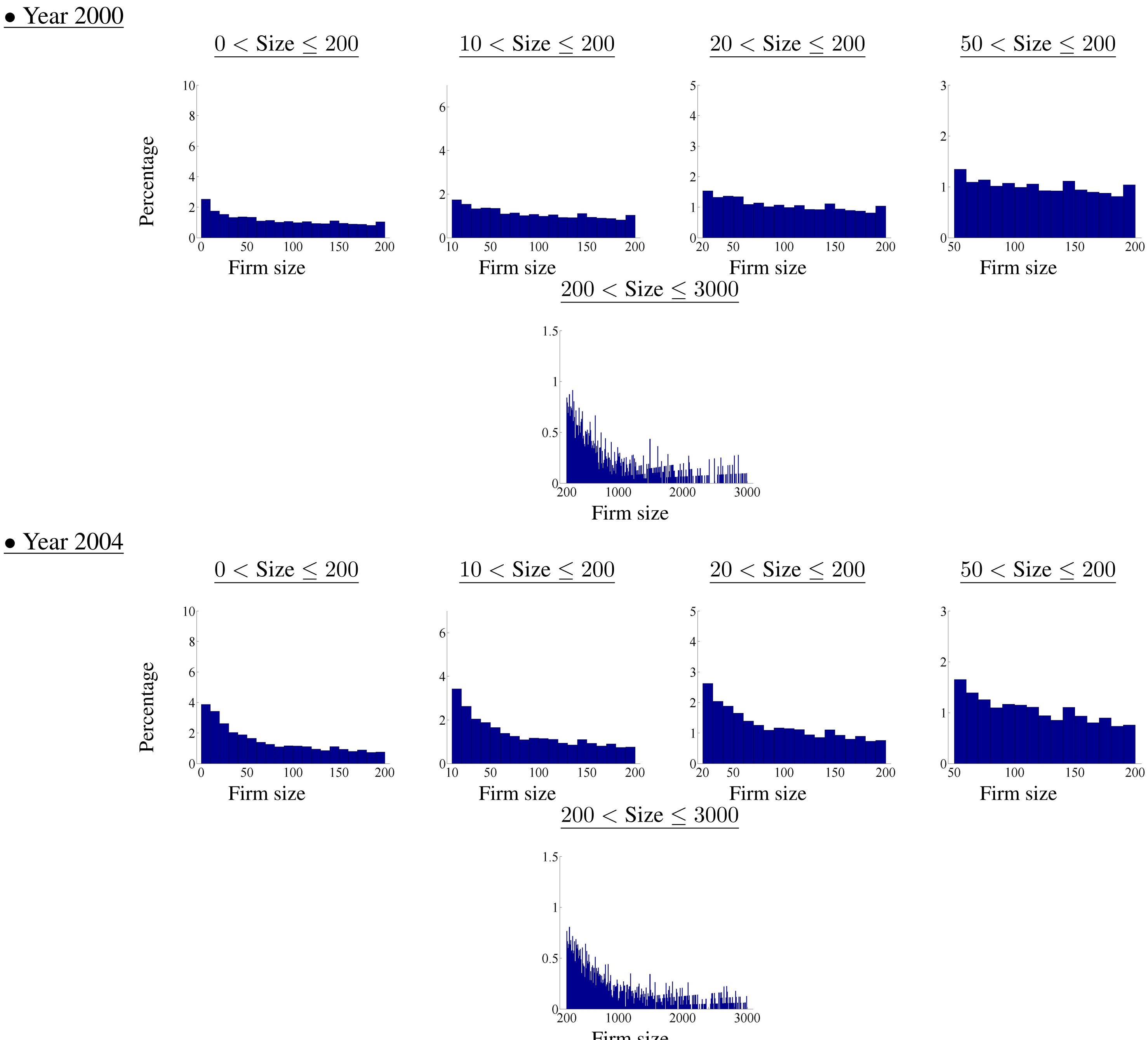
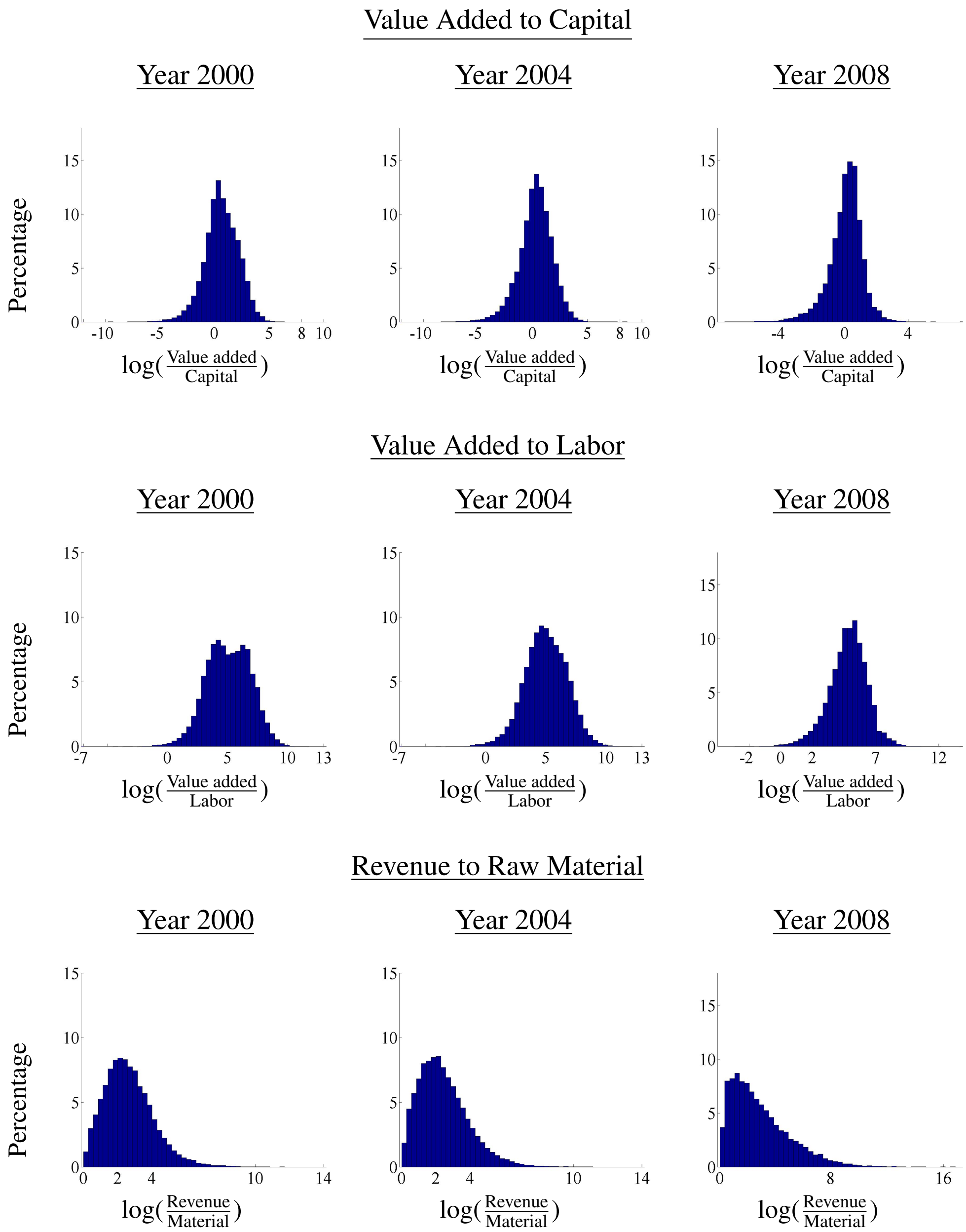
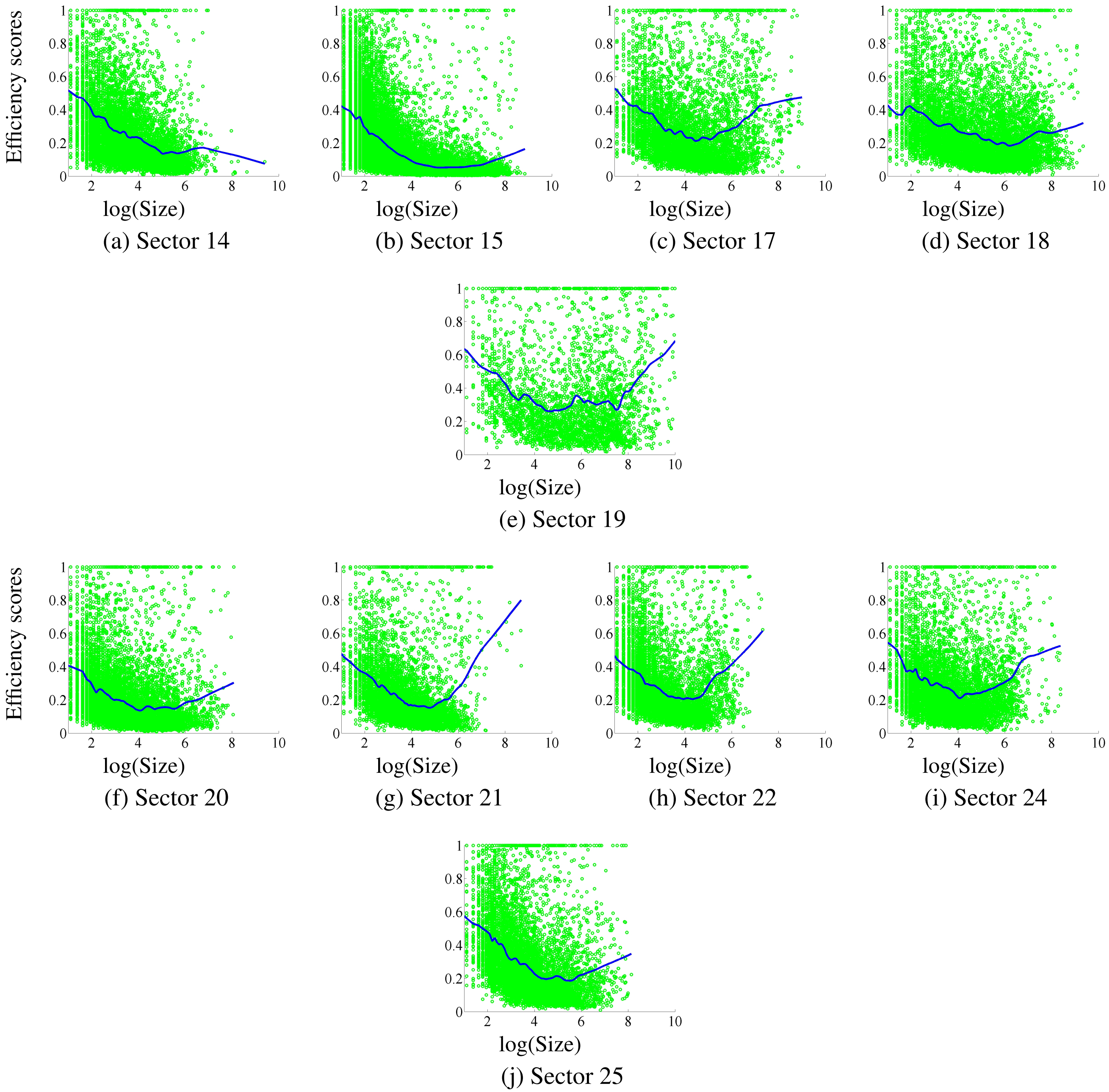


Figure 7.3: Distribution of Average Products



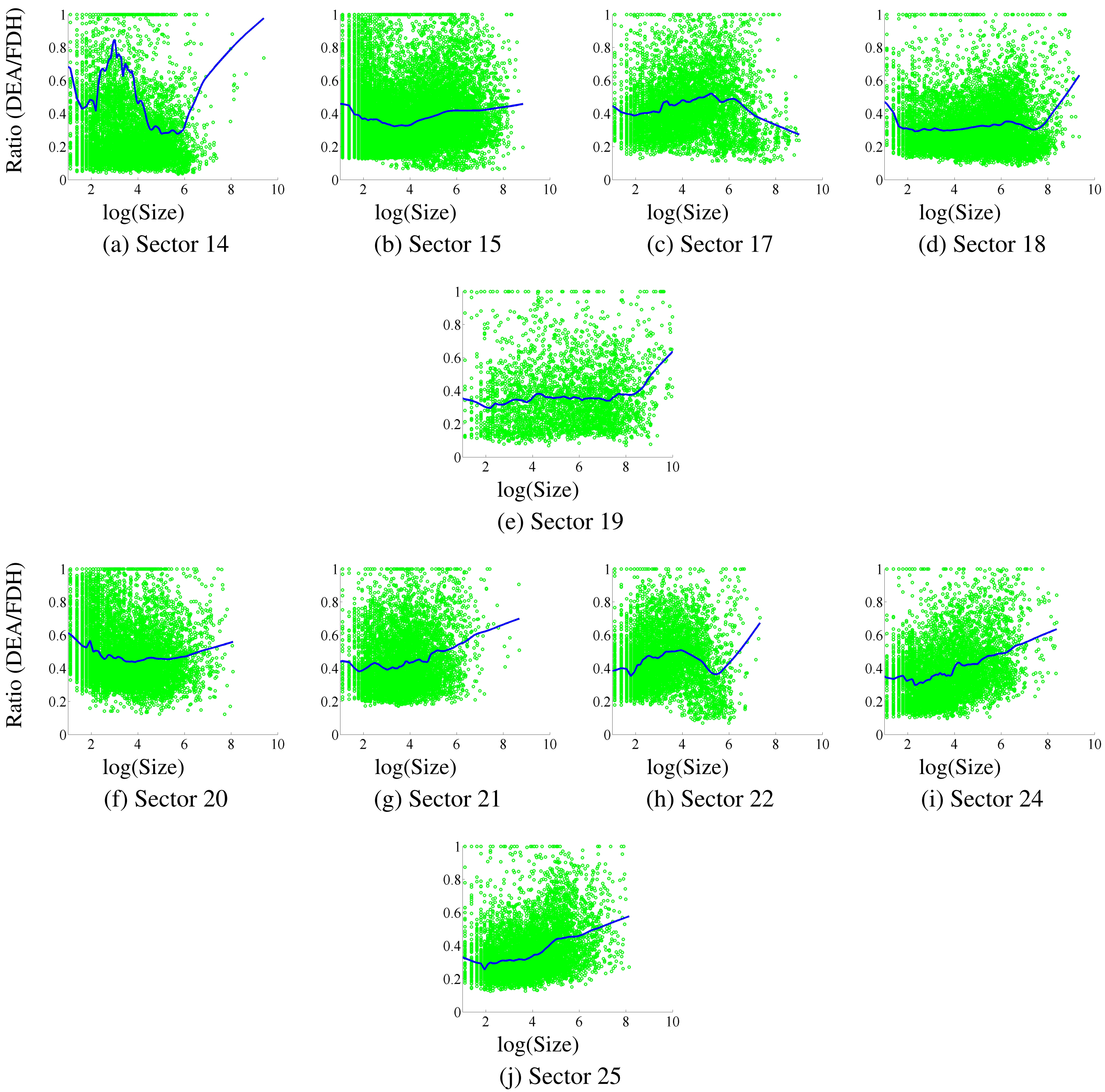
Note: We take the maximums and the minimums of $\log(\frac{\text{Value added}}{\text{Capital}})$ (the first row), $\log(\frac{\text{Value added}}{\text{Capital}})$ (the second row) and $\log(\frac{\text{Revenue}}{\text{Material}})$ (the third row), respectively and truncate the intervals into 50 bins. Each bin contains the upper bound and not the lower bound.

Figure 7.4: Efficiency Scores in Each Industry



(a) Sector 14
(b) Sector 15
(c) Sector 17
(d) Sector 18
(e) Sector 19
(f) Sector 20
(g) Sector 21
(h) Sector 22
(i) Sector 24
(j) Sector 25

Figure 7.5: Ratio of the Two Efficiency Scores in Each Industry



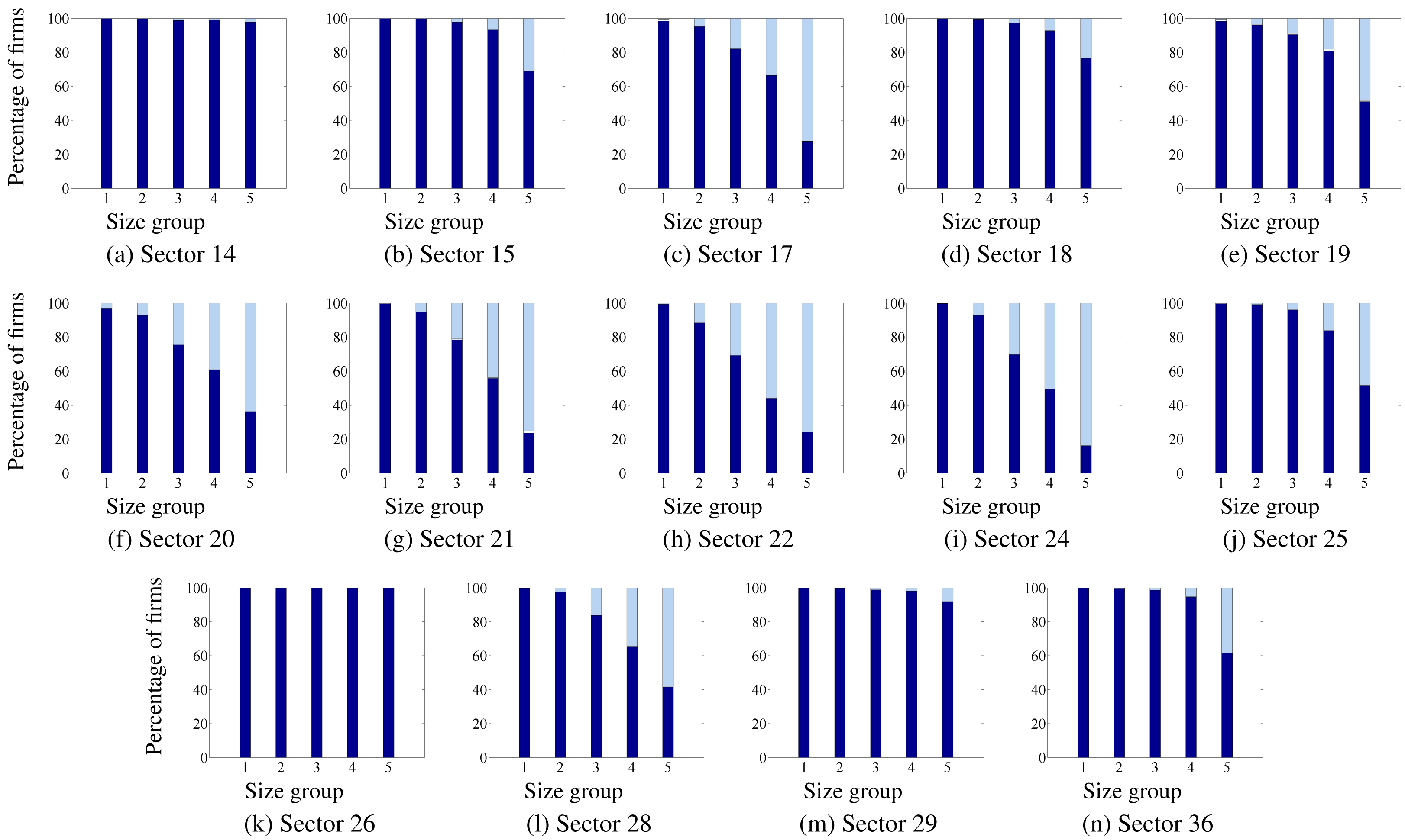
(a) Sector 14 (b) Sector 15 (c) Sector 17 (d) Sector 18

(e) Sector 19

(f) Sector 20 (g) Sector 21 (h) Sector 22 (i) Sector 24

(j) Sector 25

Figure 7.6: Inefficiency Associated with Scale in Each Industry



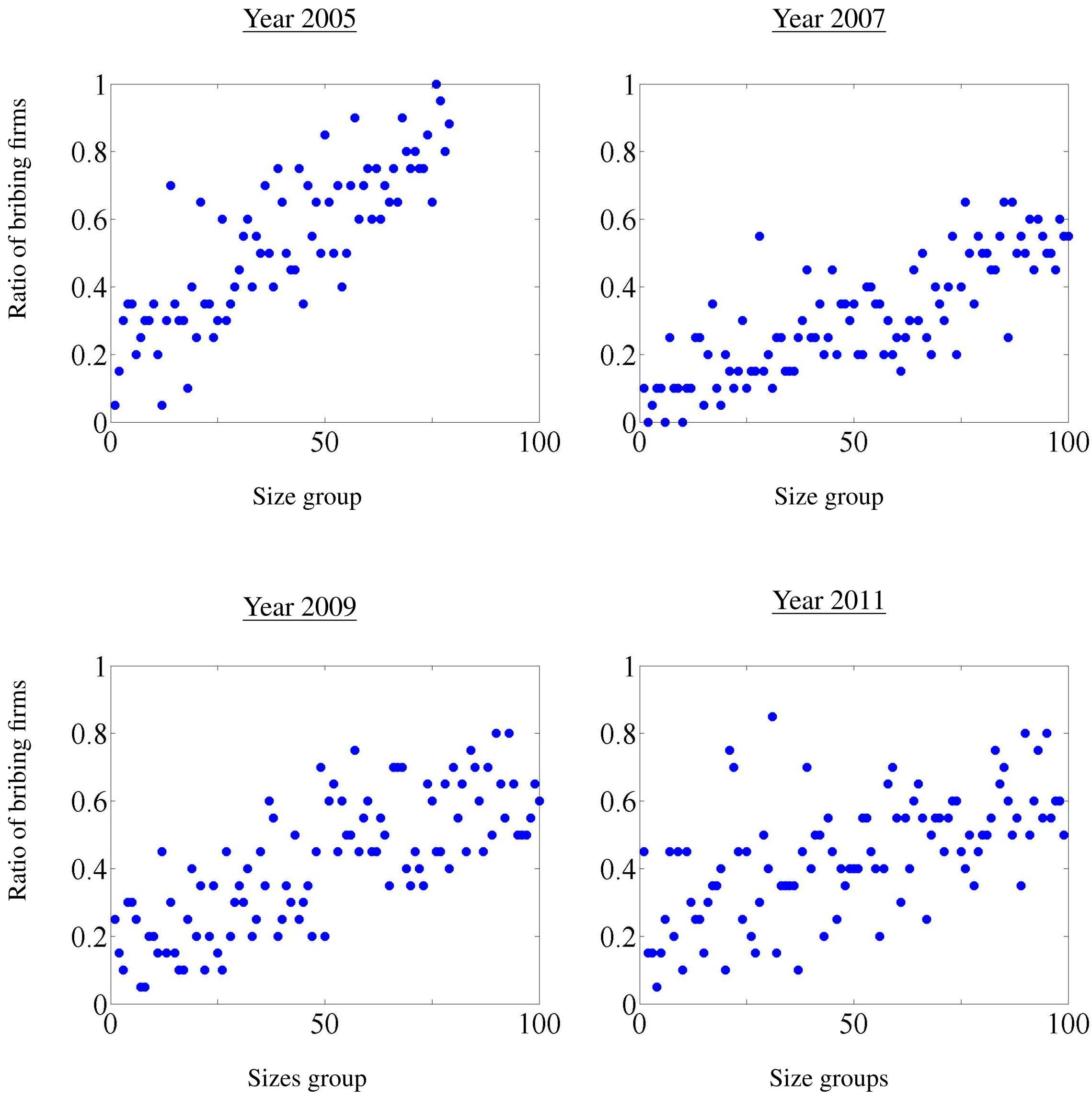
(a) Sector 14 (b) Sector 15 (c) Sector 17 (d) Sector 18 (e) Sector 19

(f) Sector 20 (g) Sector 21 (h) Sector 22 (i) Sector 24 (j) Sector 25

(k) Sector 26 (l) Sector 28 (m) Sector 29 (n) Sector 36

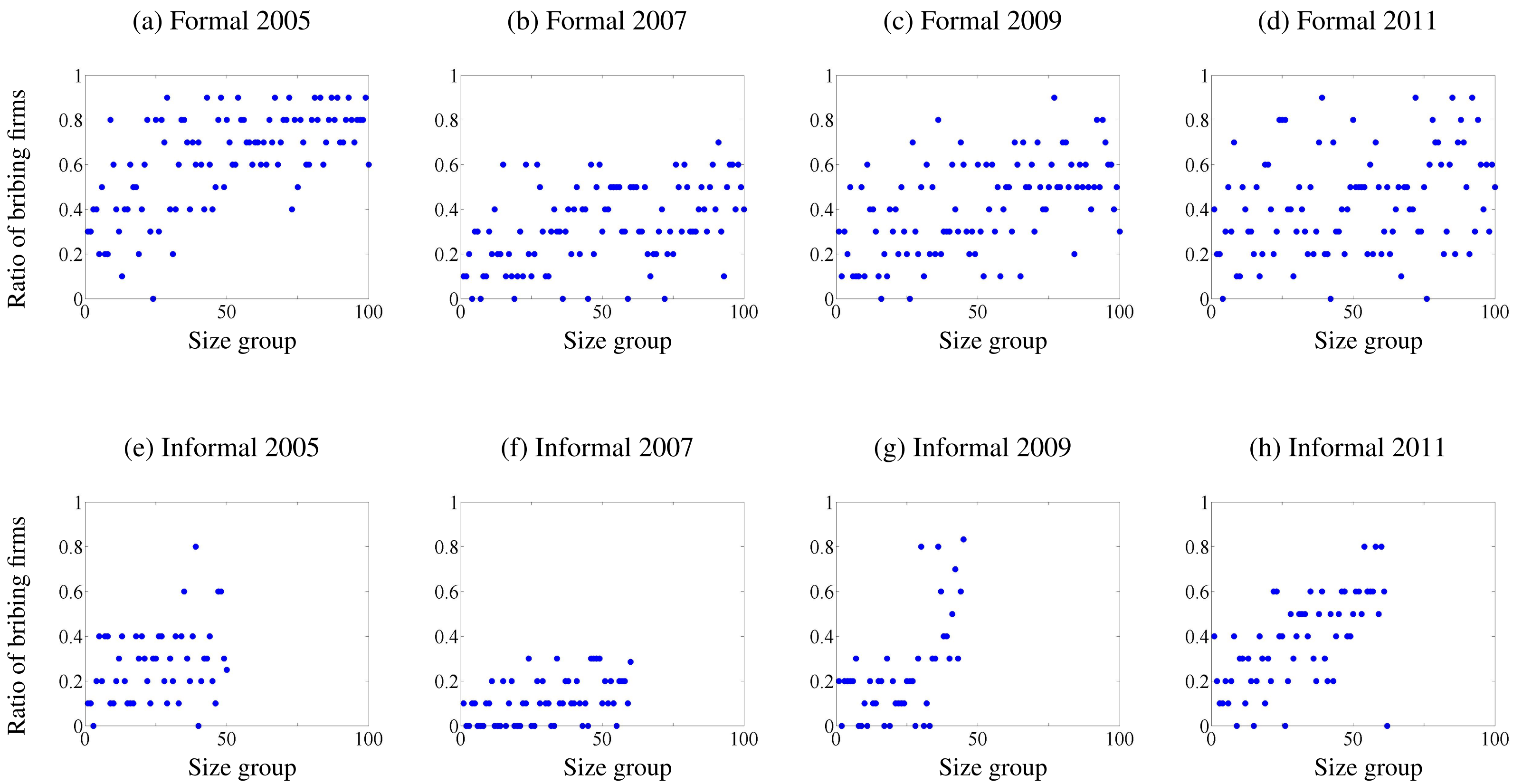
Note: A light color represents a portion of firms for $\lambda_i^- > 1$, a dark color represents a portion of firms with $\lambda_i^+ < 1$ and a white color represents the rest of firms which do not belong to either group. The x-axis represents each group, namely n is Group n for each $n = 1, \cdots, 5$.

Figure 7.7: Ratio of Bribing Firms for Every 20 Firms (Total)



Note: The x-axis represents firms' size groups. The n's group is the n-th largest group for $n = 1, \cdots, 100$.

Figure 7.8: Ratio of Bribing Firms for Every 10 Firms (Formal and Informal)



Note: The first row is the figures of formal firms who have tax codes. The second raw is the figures of informal firms who do not have tax codes. The x-axis represents firms' size groups. The n's group is the n-th smallest group for $n = 1, \cdots, 100$. Each dot represents a ratio of firms paying bribes compared to the total number of firms in each group.