ESSAYS ON FIRM ORGANIZATION AND INTERNATIONAL TRADE

CHENG CHEN

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF
ECONOMICS
ADVISER: PROFESSOR GENE M. GROSSMAN

SEPTEMBER 2014
© Copyright by Cheng Chen, 2014.

All rights reserved.
Abstract

This dissertation consists of three essays at the intersection of organizational economics and international trade. In the first essay, I investigate how the quality of management technology (MT) to monitor and incentivize employees affects aggregate economic outcomes. The key economic insight is that a common improvement in MT across all firms favors big firms, since these firms use management more intensively by adopting management hierarchies with more layers. This heterogeneous impact on firms with different numbers of layers creates a selection effect that the smallest firms exit the market, and the biggest firms expand. As a result, average firm size and aggregate productivity increase. In the second essay, I extend the baseline model developed in the first essay into the international context and investigate how an improvement in MT interacts with trade liberalization. Two theoretical results deserve particular attention. First, countries with better MT trade more with each other conditional on other factors. Second, a better MT amplifies the welfare gains from trade under certain conditions. Quantitative exercises show that an improvement in MT has quantitatively significant impacts on average firm size, aggregate productivity, and the welfare gains from trade. The final essay develops a general equilibrium model featuring an agency problem inside the firm (i.e., the separation of ownership and control) and points out a new channel through which trade liberalization leads to within-firm productivity gains. In the closed economy, managers working in the least productive firms exert effort higher than the second-best level to induce their owners to produce. After trade liberalization, a fraction of these managers is incentivized to exert more effort, since they still want to induce their owners to produce and continue to receive rents. Therefore, the least productive surviving firms whose ownership is separate from control receive productivity gains after trade liberalization.
Acknowledgements

This thesis would have never been written without the love and continuous support of my family. The support from my parents and wife helped me go through the tough months when I was on the job market.

I am grateful to my professors at Hitotsubashi University for teaching me to think as an economist and, more importantly, for encouraging me to keep learning about economics. In particular, I thank Taiji Furusawa, Jota Ishikawa and Hideshi Itoh for their help in finding my way to Princeton.

At Princeton, I have been incredibly fortunate to work under the supervision of Gene Grossman, Stephen Redding and Esteban Rossi-Hansberg. Gene was encouraging and supportive since my first day at Princeton. His comments and suggestions were always helpful, and his professionalism impressed me as well. Steve has been a wonderful mentor and a friend. His constant faith in me and insightful advice changed the course of my research and life. He will always be my role model for an advisor. Esteban's comments and suggestions were always instrumental in transforming my random thoughts into research projects and, later, into actual research papers. Other than my dissertation committee members, Richard Rogerson provided excellent suggestions that enormously improved this dissertation. I am also grateful to professors and students who attended my presentations at either the trade lunch seminar or the macro lunch seminar. Special thanks are given to Mark Aguiar, Mike Golosov, Oleg Itskhoki, Greg Kaplan, Nobuhiro Kiyotaki, Eduardo Morales, Benjamin Moll, and Felix Tintelnot. Discussions with professors from other institutions and participants of Hitotsubashi Conference on International Trade and FDI, 10 Years of the World Management Survey Conference, and Midwest International Trade Meeting 2014 helped me improve my dissertation substantially. Special thanks are given to Pol Antràs, Alexis Antoniades, Costas Arkolakis, Davin Chor, Kala Krishna, Nuno
Limao, John Mclaren, Guido Menzio, Andy Newman, Peter Neary, John Van Reenen, and Hongsong Zhang.

My classmates and friends at Princeton were constant sources of laughter, joy, and support. Special thanks go to Qingqing Cao, Tom Carter, Judd Cramer, Wei Cui, Liang Dai, Cecile Gaubert, Ji Huang, Zongbo Haung, Shaowei Ke, Mariano Somale, Chang Sun, Sharon Traiberman, Tom Winberry, and Yu Zhang who made my study and research at Fisher Hall infinitely more pleasant.

Finally, I want to thank my dearest one, Liwen Duan, who has given me strength and love throughout the years of this project.
To Sunny Y. Chen and Liwen Duan
Contents

Abstract ................................................................. iii
Acknowledgements ....................................................... iv
List of Tables ........................................................... x
List of Figures .......................................................... xi

1 Management Quality and Firm Organization in the Aggregate Economy 1
1.1 Introduction ......................................................... 1
1.2 Partial Equilibrium Analysis ...................................... 10
  1.2.1 Environment .................................................. 10
  1.2.2 The Organization of Production ............................... 12
  1.2.3 Endogenous Selection into the Hierarchy with Different Numbers of Layers ........................................... 19
  1.2.4 The Spans of Control, Wages, and Relative Wages ............. 24
  1.2.5 Firm Productivity ............................................. 26
1.3 General Equilibrium Analysis ..................................... 30
  1.3.1 Product Market Equilibrium ................................. 30
  1.3.2 Labor Market Equilibrium .................................... 32
  1.3.3 Equilibrium and Unemployment ................................ 34
1.4 Management Technology, Institutional Quality, and Firm-Level Outcomes ............................................ 38
List of Tables

1.1 Change in Welfare when MT Improves and $\sigma$ is Big ............... 45
1.2 Change in Welfare when MT Improves and $\sigma$ is Small ............. 46
1.3 Labor Market Rigidity, Management Scores, and Firm Size .......... 52
1.4 Average Management Score and Firm Size: Industry-Level Analysis .. 53
1.5 Management Score Decomposition ..................................... 57
1.6 Difference in the Quality of Monitoring and Incentives between China and the U.S. ..................................................... 58
1.7 The Cutoffs ..................................................................... 112

2.1 Moments from the Data and the Model ................................... 146
2.2 Parameter Values .................................................................. 147
2.3 Moments Related to Internal Firm Organization ......................... 147
2.4 Firm Size, Aggregate Productivity and Welfare .......................... 148
2.5 Trade Share and the Welfare Gains from Trade ......................... 149
# List of Figures

1.1 A Firm with Two Layers .................................................. 15  
1.2 A Firm with Three Layers ................................................ 16  
1.3 Average Variable Cost and Marginal Cost ............................ 21  
1.4 Average Variable Cost and the Demand Draw ....................... 23  
1.5 Firm Size and Productivity ............................................. 29  
1.6 Gains in Aggregate Productivity ................................. 43  
1.7 Management Quality and Firm Performance ....................... 56  
1.8 Management Quality and the Firm Size Distribution .......... 58  
1.9 Lower and Upper Bounds on Layer-Switching from $T = 1$ to $T = 2$ 71  
1.10 Optimal Choice of the Number of Layers ....................... 72  
1.11 The Modified AVC Curve and MC curve ............................. 111  
1.12 The Firm Size Distribution ........................................ 113  
1.13 Aggregate Productivity and the Number of Layers ............ 114  
3.1 The Timing of the Game ............................................. 164  
3.2 The Initial Quality and the Optimal Effort Choice ............ 170  
3.3 The Initial Quality and Firm Productivity ...................... 171  
3.4 The Optimal Effort Choice in the Open Economy .............. 178  
3.5 Opening Up to Trade and the Optimal Effort Choice .......... 182
Chapter 1

Management Quality and Firm Organization in the Aggregate Economy

1.1 Introduction

Recent empirical research using firm-level survey data from various countries has substantiated the existence of large variation in the quality of management technologies across countries. Furthermore, the quality of management technologies has been shown to have considerable impact on firm performance and organization (Bloom and Van Reenen (2007, 2010), and Bloom, Sadun, and Van Reenen (2012a)). However, relatively little is known about the aggregate implications of differences in management technologies. More specifically, how does the quality of management technologies affect the firm size distribution, the organizational structure of firms with different efficiency levels and the average productivity of firms in an economy? To study these

\footnote{For a discussion of management as a technology, see Bloom, Sadun, and Van Reenen (2012b). Management technologies defined in this paper are the same as management practices defined in Bloom and Van Reenen (2007, 2010). Examples include good management rules to remove poor performers and check employees’ behavior effectively.}
questions, I develop a general equilibrium model of heterogeneous firms in differentiated product markets that incorporates one type of management technology and endogenous managerial organization. I focus on the canonical approach to modeling endogenous firm organization based on effort and incentives, which have been empirically shown to be important dimensions of management practice. I show that a management technology that allows firms to better monitor and incentivize employees generates a selection effect that facilitates resource reallocation from less efficient and smaller firms to more efficient and bigger firms. As a result, firms are bigger, more decentralized, and more productive on average, when the management technology improves.

This paper focuses on the quality of a particular type of management technology: the ability to monitor and incentivize employees given a firm’s organizational choices. From now on, for the sake of simplicity I use the term management technology (MT) to refer to the management technology used to monitor and incentivize employees. I focus on this dimension of MT because it is an important component of overall management technology and affects firm performance substantially.\(^2\) This type of MT is soft technology that consists of various management rules. Management rules that either specify regular performance tracking and review or remove poor performers help a firm monitor workers and punish shirking employees.

Large differences in the quality of MT are beyond the control of the firm. For instance, low-quality institutions such as rigid labor markets and weak law enforcement negatively affect the ability of firms to punish misbehaving employees (e.g., Bloom and Van Reenen (2010) and Bloom et al. (2013)). Moreover, better management rules diffuse slowly across borders and do not exist in many countries because of information barriers.\(^3\) Hence, I treat MT as exogenous from the perspective of a firm,

\(^2\)For details on the overall management quality and effects of monitoring and incentives on firm performance, see Appendix 1.7.1.

\(^3\)Bloom et al. (2013) pointed out that one major reason why Indian firms are poorly managed is that their managers do not know about the existence of better management technologies.
but allow firms to make endogenous choices about management organization subject to this technology.

The economic objects I want to analyze interact together, and MT seems to play a role in determining them. First, differences in the firm size distribution across countries have implications for resource misallocation and aggregate productivity (Hsieh and Klenow (2009)). Second, the organizational structure of firms matters for firm performance and intra-firm wage inequality (Caliendo, Monte, and Rossi-Hansberg, 2014). Most importantly, all of these are systematically related to MT. For instance, Bloom et al. (2013) argued that one major reason for why efficient firms can’t expand fast in India is that they are unwilling to decentralize the production processes due to bad MT. Because of the slow expansion of efficient firms, many small and inefficient firms survive in India, which is one of the reasons why the aggregate productivity of firms is low in India. In summary, MT is a candidate to explain differences in aggregate-level and firm-level outcomes across economies.

This paper develops a general equilibrium model with two sectors. One sector is a homogeneous sector. It is a perfectly competitive sector with a constant returns to scale technology producing a homogeneous good. I assume that there are no monitoring and incentive issues inside firms of this sector. The other sector which is the main focus of my analysis is a monopolistically competitive sector. It comprises a continuum of differentiated products with a constant elasticity of substitution (CES) à la Dixit and Stiglitz (1977). For simplicity I refer to the monopolistically competitive sector henceforth as the CES sector. The purpose of having the homogeneous sector is to endogenize the expected wage of workers in the CES sector in a tractable way. The demand for these products varies depending on their individual characteristics. An entrepreneur can enter this sector by paying a fixed cost, and then she receives a random draw of demand (or quality) for her product. The demand draw and the quality draw are isomorphic in this framework, hence I will refer to them
interchangeably. Once the entrepreneur observes the quality, she decides whether or not to stay in the market as there is a fixed cost to produce as well. In equilibrium, entrepreneurs in the monopolistically competitive sector earn an expected payoff that is equal to their exogenous outside option due to free entry.\footnote{The expected payoff equals expected profit minus the disutility to exert effort.}

Firms in the CES sector need to monitor and incentivize employees, as production requires both time and effort, and the latter is costly for firms to observe. Following the canonical approach to modeling monitoring and incentive problems within a firm (i.e., Calvo and Wellisz (1978, 1979) and Qian (1994)), I assume that the firm sets up a hierarchy to monitor workers and provide incentives. A hierarchy is an organization with multiple layers, and a layer is a group of workers who have the same level of seniority. More specifically, the firm allocates workers into different layers to make supervisors monitor their direct subordinates and offer incentive-compatible wage contracts to workers. In equilibrium, production workers (i.e., workers in the bottom layer) and non-production workers are incentivized to exert effort to produce output and monitor subordinates respectively.

Firms whose products have greater demand set up a hierarchy with more layers. In addition to output and price, firms choose the number of layers as well as the span of control at each layer. The span of control is defined as the ratio of the number of supervisors to the number of their direct subordinates. When the firm wants to produce more, it has to increase the span of control owing to the constraint of managerial talent at the top. A larger span of control implies that less attention is paid to monitor each subordinate which has to be compensated by a higher wage, since the firm needs to prevent workers from shirking. As a result, the marginal cost (MC) increases. The firm can add a layer and decrease the span of control to save wage payments to workers at existing layers, which makes the MC drop. However, this comes at the cost of extra wage payments to workers at the new layer. In short,
adding a layer is like an efficiency-enhancing investment with a fixed cost. Firms whose goods are more preferred by consumers have an incentive to set up a hierarchy with more layers, as they produce more in equilibrium.

In order to study the selection effect of improved MT, I consider a scenario in which the quality of MT, which is common across firms, improves. This occurs when labor markets are deregulated, or better management rules are introduced into an economy. Such an improvement benefits all firms by reducing their labor costs. Furthermore, this benefit is heterogeneous across firms. Firms with more layers (or bigger firms) gain disproportionately more, since their average variable costs (AVCs) increase less rapidly with output. Intuitively, firms that choose to have more layers expand more aggressively after an improvement in MT. This aggressive expansion by bigger firms creates competitive pressure on smaller firms. As a result, firms with the worst demand draws are forced to leave the market, and firms with the best demand draws receive more profit and revenue. In total, the selection effect appears after an improvement in MT.\footnote{In a hypothetical world, if all firms were forced to have the same number of layers due to an infinitely high cost of adding and dropping layers, the heterogeneous impact (and the exit of the smallest firms) would disappear after MT improves. This is because all firms have the same AVC function. Therefore, endogenous selection into the hierarchy with different numbers of layers is the key to generating a heterogenous impact of an improvement in MT on firms with different demand draws.}

The selection effect discussed above yields three implications for resource reallocation inside the CES sector. First, the resulting firm size distribution moves to the right in the first-order-stochastic-dominance (FOSD) sense after an improvement in MT. In other words, firms are bigger on average in economies with superior MT. This result is consistent with the finding from Klenow and Olken (2014) that Mexico and India whose firms have lower management scores than American firms have more small firms and fewer big firms. Moreover, average firm size is much bigger in the U.S. compared with India and Mexico. Second, all surviving firms either increase the number of layers or make the span of control larger after an improvement in
MT. Therefore, firms are more decentralized in economies with superior MT, which is consistent with the findings in Bloom, Sadun, and Van Reenen (2012a) and Bloom et al. (2013). Finally, the aggregate productivity of firms increases as a result of an improvement in MT as well. Gains in aggregate productivity come from three sources. First, the least productive firms exit the market after an improvement in MT, which is due to the selection effect. Second, the market shares of the most productive firms increase after an improvement in MT, which is due to the selection effect as well. Third, the productivity of surviving firms increases, as improved MT reduces firm costs. In total, these three implications are the aggregate implications of an improvement in MT and are the key contributions of this paper compared to the literature. The key economic insight emphasized in this essay (i.e., the selection effect arising from an improvement in MT) is pointed out for the first time in the literature and deserves more attention in future research.

Beyond the macro-level implications, the model has micro-level predictions as well. First, although wages at all layers increase with firm size given the number of layers, they fall at existing layers when the firm adds a layer. This reduction occurs because the addition of a new layer reduces the span of control at existing layers. This result shows that employees might lose from firm’s expansion. Second, when wages increase, they increase disproportionately more at upper layers, which leads to a bigger wage ratio between two adjacent layers. Similarly, when wages fall, they fall disproportionately more at upper layers, which leads to a smaller wage ratio between two adjacent layers. These results imply a distributional effect of firm expansion on workers’ wages. Third, in the theory, firms that are bigger or more efficient have more

---

6Powell (2013) investigated how contract enforcement affects the dispersion of the distribution of firm productivity in a perfectly competitive product market. One key result from that paper is that weak enforcement of laws hurts unproductive firms more due to the existence of a fixed production cost and a dynamic-enforcement constraint. As a result, the distribution of firm productivity is more dispersed in economies with weaker enforcement of laws. Both Powell (2013) and my paper emphasize the role of institutions in shaping aggregate productivity. Powell (2013) focused on the second-order moment of the distribution of firm productivity, while my paper focuses on the first-order moment of it.
layers. This is because adding a layer is like an investment that requires a “fixed” organization cost and reduces the firm’s MC. In total, all the above results on firm-level outcomes are consistent with the evidence presented in Caliendo, Monte, and Rossi-Hansberg (2012) and have implications for intra- and inter-firm wage inequality.

This paper contributes to the literature on incentive-based hierarchies in three ways. First, I treat the number of layers as a discrete variable and manage to derive the optimal number of layers for the firm by characterizing the firm’s cost functions. This is for the first time in the research of incentive-based hierarchies that such a difficult problem has been solved. Treating the number of layers as a discrete variable is empirically realistic and important for the model’s predictions on wages and firm productivity. Second, firm size is endogenously determined in my model, as each firm faces a downward-sloping demand curve. The determination of optimal firm size is a central theoretical obstacle for the research on incentive-based hierarchies. This paper manages to solve this problem by considering the firm’s problem in a monopolistically competitive market environment. Finally, I incorporate the canonical model of incentive-based hierarchies into a general equilibrium setting. By doing so, I can analyze the impact of improved MT on the firm size distribution and weighted average of firm productivity, which are general equilibrium objects. Furthermore, these implications can be readily contrasted with the data and help explain the stylized patterns observed in the real world.

The literature on knowledge-based hierarchies (e.g., Garicano (2000), Garicano and Rossi-Hansberg (2004, 2006, 2012), Caliendo and Rossi-Hansberg (2012)) has been successful at providing a framework to analyze how the information and commu-
nication technology (ICT) affects various economic outcomes such as wage inequality and economic growth. However, it is silent on the role of MT in determining firm-level outcomes and aggregate economic variables such as the firm size distribution and aggregate productivity. A paper that is closely related to mine is Caliendo and Rossi-Hansberg (2012). There are three major differences between these two papers. First, the moral hazard problem which is absent in Caliendo and Rossi-Hansberg (2012) is a key ingredient of this paper. I focus on a different mechanism (effort and incentives), and it yields some different predictions such as what matters for the firm size distribution is not the ICT but institutions that affect the ability of firms to monitor and incentivize employees. Second, the focus of my paper is the selection effect of better MT in the closed economy, while Caliendo and Rossi-Hansberg (2012) focused on how trade liberalization affects firm organization and productivity in the open economy. Finally, this paper yields some important micro-level predictions that Caliendo and Rossi-Hansberg (2012) did not have. For example, the result that firms increase the number of layers when MT improves is unique to my model, since improvements in the ICT would imply flattened firm hierarchies (i.e., delayering). This difference is important, since the number of layers has been shown to be an important factor determining firm performance and workers’ wages (Caliendo, Monte, and Rossi-Hansberg (2012)).

This paper contributes to the recent macro-development literature that discusses cross-country differences in the firm size distribution and resource misallocation (e.g., Hsieh and Klenow (2009, 2012) and Hsieh and Olken (2014)) in two ways. First, this paper provides a new explanation for why firms in developing countries are smaller and less productive on average compared to those in developed countries. As pointed out by Hsieh and Olken (2014), some prevailing arguments that are used to explain cross-country differences in the firm size distribution face empirical challenges. For instance, size-dependent policies which are commonly used in many developing countries would
imply bunching of firms around certain thresholds on firm size. However, this type of bunching does not seem to be quantitatively important in the data, at least for India, Mexico and Indonesia (Hsieh and Olken (2014)). Another popular view is that financial constraints hurt smaller firms more than bigger firms in developing countries. This argument would imply that developing countries have “missing middle” in the size distribution of firms. That is, developing countries have both more small firms and more big firms compared with developed countries. Moreover, there are fewer medium-sized firms in developing countries compared with developed countries as well (i.e., a bimodal distribution of firm size). Unfortunately, this feature does not seem to exist in the data as well (Hsieh and Olken (2014)). On the contrary, the theory proposed in this paper does not have the above counterfactual predictions. Second, the key economic insight of this paper that bigger and more efficient firms are constrained more in developing countries (i.e., countries with worse MT) is consistent with conjectures and suggestive evidence provided by Bloom et al. (2013) and Hsieh and Olken (2014). In summary, this essay points out a new channel through which differences in quality of MT affect the firm size distribution and resource allocation. Furthermore, aggregate predictions of the theory are consistent with the suggestive evidence.

The remainder of the paper is organized as follows. Section two solves the individual firm’s optimization problem. Section three solves the problem of resource allocation in general equilibrium. Section four investigates how differences in MT across economies (and firms) affect various economic activities. Section five presents evidence to support the main theoretical predictions derived in Section four. Section six concludes.
1.2 Partial Equilibrium Analysis

In this section, I develop a model of the hierarchical firm that features firms’ endogenous selection of a hierarchy with a specific number of layers. The key elements are the firm’s decisions on the span of control as well as the number of layers. I will subsequently introduce the model into a general equilibrium setting and solve the problem of resource allocation in both the product and labor markets in the next section.

1.2.1 Environment

The economy comprises two sectors, $L$ units of labor and $N$ potential entrepreneurs, where $N$ is sufficiently large that the free entry (FE) condition discussed below will hold with equality. One sector produces a homogeneous good and is perfectly competitive, while the other sector produces horizontally differentiated goods and features monopolistic competition.

A representative agent demands goods from both sectors and has the following Cobb-Douglas utility function:

$$U = \left( \frac{C_c}{\gamma} \right) \gamma \left( \frac{C_h}{1-\gamma} \right)^{1-\gamma} - I\psi(a_i),$$  \hspace{1cm} (1.1)

where $C_h$ is the consumption of the homogeneous good and $C_c$ is an index of consumption of differentiated goods defined as

$$C_c = \left( \int_{\Omega} \theta^{\frac{1}{\sigma}} y(\theta) \frac{\sigma-1}{\sigma} M \mu(\theta) d\theta \right)^{\frac{\sigma}{\sigma-1}},$$  \hspace{1cm} (1.2)

$y(\theta)$ is the consumption of variety $\theta$, $M$ denotes the mass of products available to the consumer, $\mu(\theta)$ indicates the probability distribution over the available varieties in $\Omega$, and $\sigma > 1$ is the constant elasticity of substitution. Note that $\theta$ is a demand shifter.
for a variety produced by a firm, so agents demand more of goods with higher $\theta$ at a given price. $I$ and $\psi(a_i)$ are, respectively, an indicator function and a disutility to exert effort that will be discussed later. The final composite good is defined as

$$C \equiv \left( \frac{C_c}{\gamma} \right)^{\gamma} \left( \frac{C_h}{1-\gamma} \right)^{1-\gamma},$$

which is the first part of terms appearing in the right hand side of equation (1.1). I choose the price of it to be the numeraire, so

$$P^\gamma P_h^{1-\gamma} \equiv 1,$$  \hspace{1cm} (1.3)

where

$$P = \left( \int_{\Omega} \theta p(\theta)^{1-\sigma} M \mu(\theta) d\theta \right)^{\frac{1}{1-\sigma}}$$  \hspace{1cm} (1.4)

is the ideal price index of the differentiated goods. $p_h$ is the price of the homogeneous good, and $p(\theta)$ is the price of variety $\theta$.

The homogeneous sector features no frictions, and the perfectly competitive market structure implies that firms receive zero profit. Labor is the only factor used in production, and the production technology implies that output equals the number of workers employed. The price of the homogeneous good is also the wage offered in this sector. There is no unemployment among workers who enter this sector in equilibrium owing to the absence of frictions.

The CES sector produces a continuum of differentiated products. The demand for these products varies depending on their individual characteristics. There is a large pool of potential entrepreneurs who have managerial ability to set up firms in this sector. An entrepreneur can enter this sector and receive a random draw of quality for her product after paying a fixed cost $f_1$ to design it. Given the existence of a fixed cost $f_0$ to produce, the entrepreneur decides whether or not to stay in the market after...
she observes the draw. Both the entry cost and the fixed cost are paid in the form of
the final composite good, as in Atkeson and Burstein (2010). The entrepreneur has
to employ workers and organize the production process if she decides to produce.

Workers choose the sector in which they seek employment, while entrepreneurs
choose whether or not to operate a firm. Both types of agents are risk neutral. In
equilibrium, workers’ expected payoff obtained from entering both sectors must be
the same since they can freely move between sectors. I assume that the outside option
(or reservation utility) of an entrepreneur is forgone, if she chooses to enter the CES
sector. Thus, the expected payoff of entrepreneurs who choose to enter the CES sector
equals their exogenous outside option \( f_2 \) due to free entry of firms in equilibrium.\(^9\)

Workers cannot choose to be entrepreneurs, as they don’t have managerial talents.
Furthermore, I assume that \( f_2 \) is big enough that the expected payoff of workers is
strictly smaller than \( f_2 \) in equilibrium. Therefore, entrepreneurs have no incentives
to become workers.

1.2.2 The Organization of Production

I follow the literature on incentive-based hierarchies (e.g., Calvo and Wellisz (1978,
1979)) in modeling the organization of production. More specifically, I assume that
each firm has to employ workers at various layers and incentivize them to exert effort
in order to produce. Production workers only produce output, while non-production
workers only monitor their direct subordinates in order to incentivize them to work.\(^{10}\)

\(^9\)Essentially, I assume that there is another sector which is not explicitly modelled here. The
entrepreneur can enter and receive the payoff of \( f_2 \) by working in that sector.

\(^{10}\)Most firms monitor and incentivize their employees in reality. For real-world exam-
examples, see [http://matthewoudendyk.blogspot.com/2006/11/how-monitoring-is-being-do-
11626015505227748.html](http://matthewoudendyk.blogspot.com/2006/11/how-monitoring-is-being-do-
11626015505227748.html) and [http://management.about.com/cs/people/a/
MonitorEE062501.htm](http://management.about.com/cs/people/a/MonitorEE062501.htm). Management activities used to monitor and incentivize employees
include a variety of jobs done by supervisors. First, supervisors monitor their subordinates
using information technology such as video surveillance, e-mail scanning and location monitoring
(Hubbard (2000, 2003)). Second, monitoring also happens when supervisors communicate with
their subordinates and try to check whether or not the subordinates are working hard. Finally,
business meetings in which supervisors evaluate subordinates’ performance and decide whether or
Production requires effort and time of workers. The worker’s effort choice $a_i$ is assumed to be a binary variable between working and shirking (i.e., $a_i \in \{0, 1\}$) for reasons of tractability\(^{11}\). The input of workers’ time equals the number of workers. Production workers produce output, and shirking results in defective output that cannot be sold. Thus, the production function is

$$q = \int_0^{m_T} a(j) dj,$$

where $m_T$ is the measure of production workers and $a(j)$ is the effort level of the $j$-th unit of labor inputs. Here I assume that labor inputs are divisible, because there is a continuum of workers. Non-production workers at layer $i$ monitor their direct subordinates at layer $i + 1$ and need to be monitored by supervisors at layer $i - 1$ as well\(^{12}\). Layer $T$ is the lowest layer in the hierarchy which is occupied by production workers. In short, production workers and non-production workers have different roles in the production process.

 Workers must be monitored if the firm wants them to exert effort. The firm cannot fire a shirking worker unless it is able to detect his misbehavior. A worker at layer $i$ is induced to work for wage $w_i$, if and only if

$$w_i - \psi \geq (1 - p_i)w_i,$$

where $p_i(\leq 1)$ is the probability of catching and firing a shirking worker, and $\psi$ is the disutility of exerting effort. A worker’s utility differs from the utility of consuming goods only when he works in the CES sector and exerts effort in the production process not to fire poor performers are important parts of monitoring and incentivizing activities as well. Admittedly, monitoring and incentivizing subordinates are parts of what non-production workers do in reality. I focus on this dimension of non-production workers’ jobs in order to distill the key economic impact of an improvement in MT on economic outcomes.

\(^{11}\)As shown in Appendix [1.7.2], it is straightforward to generalize the analysis to allow effort to be a continuous variable.

\(^{12}\)A smaller $i$ denotes a higher layer in the firm’s hierarchy, and the entrepreneur is at layer zero.
(i.e., $I = 1$ in equation (1.1)). The above inequality is the incentive compatibility constraint that the payoff obtained from exerting effort must be greater than or equal to that of shirking.

The probability of catching and firing a shirking worker depends on two factors: the adjusted span of control and MT. First, the bigger the adjusted span of control, the less frequently a subordinate’s behavior is checked by his supervisor (less time or monitoring effort is spent on him). This implies a lower probability of catching a shirking worker. Second, the quality of MT affects the probability of detecting workers’ misbehavior. Management rules that clarify performance measures lead to easier detection of workers’ misbehavior. Finally, the quality of MT also affects the probability of successfully firing shirking workers. Firms that are located in economies with either rigid labor markets or weak law enforcement are found to be worse at using good management rules to remove poor performers (Bloom and Van Reenen (2010)). I capture these effects by assuming the following functional form for $p(b, x_i)$:

$$p(b, x_i) = \frac{1}{bx_i(\theta)},$$  

where $x_i(\theta) \equiv \frac{m_i(\theta)}{\int_0^{m_i(\theta)} a(j) dj}$ is the span of control adjusted by supervisors’ effort inputs.\textsuperscript{13} Parameter $b$ reflects the inefficiency of MT. More specifically, the worse the MT is, the bigger the value of $b$.

A firm may want to hire non-production workers, since it wants to economize on the cost to incentivize workers. I use Figures 1.1 and 1.2 to clarify the economic intuition behind this choice.\textsuperscript{14} Consider firm $A$ that receives a low demand draw $\theta_A$ and wants to produce two units of goods as illustrated in Figure 1.1. The span of control of the entrepreneur is small for this firm, which implies a low incentive-

\textsuperscript{13}As what I will show, the entrepreneur allocates the monitoring intensities evenly across workers at the same layer. A more flexible functional form is $p(b, x_i) = \frac{1}{bx_i(\theta)^\nu}$ where $\nu$ can be different from one. Allowing $\nu$ to differ from one does not affect qualitative results of the paper. Detailed derivations are available upon request.

\textsuperscript{14}These two figures only serve for illustrative purposes.
compatible wage paid to production workers. Thus, it is optimal to have production workers only, since non-production workers do not produce output. Next, consider firm $B$ that receives a high demand draw $\theta_B$ and wants to produce six units of goods, which is illustrated in Figure 1.2. The incentive-compatible wage paid to production workers would be too high, if the firm did not hire non-production workers between the entrepreneur and production workers.\footnote{Remember that there is a fixed number of entrepreneurs (i.e., one entrepreneur) at the top.} If the firm hires non-production workers who monitor production workers, the incentive-compatible wage paid to production workers will be reduced, which makes the labor costs lower. Obviously, this comes at the cost of extra wage payment to non-production workers. Therefore, it is optimal for the firm to add non-production workers only when the output level is high, which will be shown rigourously in Subsection 1.2.3. The above logic also explains why the firm wants to have \textit{multiple} layers of non-production workers when the output level is high. In total, the trade-off between lower wages paid to existing workers and extra wages paid to workers employed at the new layer determines the optimal choice of the number of layers for a firm.
I characterize two optimal choices of the firm before solving the firm’s optimal decisions on the other variables (e.g., output and employment etc.), as these two choices are independent of the firm’s decisions on the other variables. In equilibrium, the firm chooses to incentivize all workers to work and allocate the monitoring intensities evenly across workers at a given layer. Intuitively, the firm can always reduce the cost by doing so, if these choices are not made. Lemma 1 proves and summarizes the above results.

**Lemma 1** The firm incentivizes all workers to work \((i.e., a_i = 1)\) and equalizes the monitoring intensity across workers at a given layer.

Proof. See Appendix 1.7.3.

The entrepreneur faces the same incentive problem as her employees. She sits at the top of the hierarchy and is monitored by nobody. However, the entrepreneur of any firm that chooses to stay in the market is incentivized to exert effort in equilibrium.
(i.e., monitor her subordinates). First, staying in the market and shirking result in zero output and negative payoff for the entrepreneur. Second, the net payoff must be non-negative for the entrepreneur, if she decides to stay in the market and exert effort. This is because the entrepreneur would exit the market, if the ex-post net payoff that is the difference between profit and the cost to exert effort were negative. Therefore, the entrepreneur exerts effort if she decides to stay in the market.

Now I characterize the firm’s optimization problem. By substituting equation (1.7) into inequality (1.6), I derive the minimum incentive-compatible wage for layer \( i \) as follows:

\[
w_i(\theta) = \frac{\psi}{p_i(b, x_i(\theta))} = \psi bx_i.
\]  

(1.8)

The key feature of the above equation is that the minimum incentive compatible wage \( w_i(\theta) \) is negatively related to the supervision intensity. This relationship finds support in the data; see, for example, Rebitzer (1995) and Groshen and Krueger (1990).\footnote{First, Rebitzer (1995) finds empirical evidence of a trade-off between supervision intensity and wage payment. Workers get paid less if they are under intensive monitoring. More importantly, Groshen and Krueger (1990) and Ewing and Payne (1999) find evidence on a negative relationship between the span of control and the wage paid to subordinates. This finding directly supports the basic trade-off of wage determination in the current model. Namely, a bigger span of control results in higher wage payment to subordinates.}

Based on equations (1.1), (1.5), (1.8) and Lemma 1, the optimization problem for a firm with the quality draw \( \theta \) conditional on its staying in the market can be stated as

\[
\max_{\{m_i\}_{i=1}^T} A^{\theta \frac{1}{\sigma}} m_{T - 1}^{\frac{\sigma - 1}{\sigma}} - \sum_{i=1}^T b\psi m_i x_i
\]

s.t. \[ x_i = \frac{m_i}{m_{i-1}}, \]
\[ m_0 = 1. \]

(1.9)

where the first part of the above equation is the firm’s revenue and the second part denotes the variable cost. The demand shifter \( A \) captures market size adjusted by
the ideal price index and takes the following form:

\[ A \equiv \left( \frac{\gamma E}{P^{1-\sigma}} \right)^{1/\sigma}, \]  

(1.10)

where \( E \) is the total income of the economy. The number of entrepreneurs per firm is normalized to one, or, \( m_0 = 1 \). A big enough \( b \) is chosen to ensure that the probability of being monitored for any worker is always smaller than or equal to one.

The firm’s optimal decisions given the number of layers can be solved in two steps. First, given an output level \( q \), the first order conditions (FOCs) with respect to \( m_i \)'s imply

\[ w_T m_T = 2 w_{T-1} m_{T-1} = ... = 2^{T-1} w_1 m_1, \]  

(1.11)

where \( m_0 = 1 \) and \( m_T = q \), as the number of production workers equals output \( q \). This leads to the solution that

\[ m_i(q, T) = 2^i \left( \frac{q}{2^T} \right) \frac{2^T - 2^{T-i}}{2^{T-1}}, \]  

(1.12)

which is the number of workers at layer \( i \). Thus, the firm’s span of control at layer \( i \) is

\[ x_i(q, T) = \frac{m_{i+1}(q, T)}{m_i(q, T)} = 2 \left( \frac{q}{2^T} \right) \frac{2^T - 2^{T-i}}{2^{i+1}}, \]  

(1.13)

which increases with \( q \) given the number of layers. Equation \ref{1.12} shows that employment at each layer increases with output given the number of layers. Moreover, equation \ref{1.13} indicates that the number of workers increases disproportionately more at upper layers, which leads to a bigger span of control at each layer. This is due to the fixed number of entrepreneurs at the top.
Second, optimizing over output yields

\[ q(\theta, T) = m_T(\theta, T) = \left[ \frac{A\beta \theta^{\frac{1}{2}}}{b\psi 2^{2-T} - 2T - 1} \right] \frac{\sigma(2T-1)}{\sigma + (2T-1)}, \] (1.14)

which is the firm’s output level as well as the number of production workers. Substituting equations (1.12) and (1.14) into equation (1.9) leads to the firm’s operating profit (i.e., profit before paying the fixed cost) and revenue as

\[ \pi(\theta, T) = (1 - \frac{\beta(2T - 1)}{2^T})(A\theta^{\frac{1}{2}}) \frac{2^T}{\sigma + (2T - 1)} \left( \frac{\beta/b\psi}{2^{2T+1-2T}} \right)^{\frac{(\sigma-1)(2T-1)}{\sigma + (2T-1)}} \] (1.15)

and

\[ S(\theta, T) = (A\theta^{\frac{1}{2}}) \frac{2^T}{\sigma + (2T - 1)} \left( \frac{\beta/b\psi}{2^{2T+1-2T}} \right)^{\frac{(\sigma-1)(2T-1)}{\sigma + (2T-1)}}, \] (1.16)

which will be used later. The firm’s employment, output, and revenue increase continuously with the quality draw \( \theta \) given the number of layers. More importantly, all of these variables increase **discontinuously** when the firm adds a layer as shown below.

With the firm’s optimal decisions on employment and output in hand, I can solve for the optimal number of layers, which is the final step to solve the firm’s optimization problem.

### 1.2.3 Endogenous Selection into the Hierarchy with Different Numbers of Layers

This subsection characterizes a firm’s cost functions in order to solve for the optimal number of layers in a firm’s hierarchy. The key result is that firms with better quality draws choose to have more layers and produce more in equilibrium. Consider a firm that produces \( q \) units of output. The variable cost function of such a firm is given
by\[17\]

\[ TVC(q, b) = \min_{T \geq 1} TVC_T(q, b), \]

where \(TVC(q, b)\) is the minimum variable cost of producing \(q\) and \(TVC_T(q, b)\) is the minimum variable cost of producing \(q\) using a management hierarchy with \(T + 1\) layers. Based on equations (1.12) and (1.13), \(TVC_T(q, b)\) is derived as

\[
TVC_T(q, b) = \sum_{i=1}^{T} m_i(q, T)w_i(q, T) = \sum_{i=1}^{T} b \psi \frac{m_i^2(q, T)}{m_{i-1}(q, T)} = (2 - \frac{1}{2^{T-1}})b \psi 2^{1-\frac{T}{2}} q^{\frac{2^T}{2^{T-1}}}. \tag{1.17}
\]

Variable cost given the number of layers increases with output. Better MT, which is denoted by a smaller value of \(b\), pushes down the variable cost given any number of layers proportionately.

With the firm’s cost functions given different numbers of layers in hand, I can characterize the AVC curve and the MC curve using the following proposition.

**Proposition 1** Given the number of layers, both the average variable cost and the marginal cost increase continuously with output. The average variable cost curve kinks and its slope decreases discontinuously at the output level where the firm adds a layer. As a result, firms that produce more have more layers. The marginal cost falls discontinuously when the firm adds a layer.

Proof. See Appendix 1.7.4.

Figure 1.3 illustrates the AVC curve and the MC curve. The AVC curve denoted by the bold green curve is the lower envelope of all AVC curves given different numbers of layers. The MC curve does not increase with output monotonically. The span of control increases at all layers when the firm increases output and keeps the

\[17\]For firms that have one layer (i.e., self-employed entrepreneurs), management hierarchies are not needed. As this paper focuses on management hierarchies, I do not consider these firms in the paper.
number of layers unchanged. Therefore, wages increase at all layers, which implies that both the AVC and the MC increase with output given the number of layers. Wages fall at existing layers when the firm adds a layer owing to the smaller span of control. This leads to a discontinuous decrease in the MC. Because of this drop, the AVC curve kinks and its slope decreases *discontinuously* at the output level where the firm adds a layer.

Proposition 1 establishes a positive relationship between output and the optimal number of layers. When the output level is low, it is ideal to have a smaller number of layers. This is because adding a layer is *like* an investment that reduces the MC at the expense of a fixed cost. This property of the AVC curve is evident in Figure 1.3, as the AVC curve given a bigger number of layers has a smaller slope and a larger intercept on the y axis. Similarly, it is optimal to have more layers when output is high. In summary, the number of layers and output increase hand in hand in equilibrium.
What is the relationship between the firm’s demand draw and the optimal number of layers? The key observation is that it is more profitable for a firm with a better demand draw to add a layer. This is due to the key feature of the AVC curve that adding a layer is like investing a fixed amount of money to reduce the MC. In other words, there is a complementarity between the level of the firm’s demand draw and its incentives to add a layer. Proposition 2 characterizes a positive relationship between the firm’s demand draw and the optimal number of layers by proving this complementarity.

**Proposition 2** *Firms that receive better demand draws have more layers.*

Proof. See Appendix I.7.5.

Proposition 2 and the distinctive feature of the AVC curve discussed above are the keys to understanding the selection effect of an improvement in MT. Improved MT reduces labor costs and incentivizes firms to grow. Furthermore, it incentivizes firms with better demand draws to expand more and benefits them disproportionately more. This is because firms with better demand draws have more layers, and the elasticity of the AVC with respect to output is smaller for these firms.

The theoretical results proven in Proposition 2 are consistent with empirical findings from Caliendo, Monte, and Rossi-Hansberg (2012). First, firms that are bigger in terms of either employment or value added are found to have more layers in the data set of French firms. Second, firms that are bigger are found to have more layers as well. All these evidence supports the key result of this paper: firms with better demand draws have more layers.

I close this subsection by discussing how price and firm size respond to a change in the firm’s demand draw. These results are useful, since I am going to analyze the firm size distribution in the next section. Proposition 3 summarizes the results.
**Proposition 3** Given the number of layers, output, employment, and price increase continuously with the firm’s demand draw. When the firm adds a layer, output and employment increase discontinuously, while price falls discontinuously.

Proof. See Appendix 1.7.6.

The strategy of a firm to grow depends crucially on whether or not the production is reorganized. When the firm grows and keeps the number of layers unchanged, price increases as the MC increases. However, when one layer is added, the MC falls, which leads to lower prices. Exactly because of this discontinuous decrease in the MC, firm size increases discontinuously when production is reorganized. As a result, the AVC as a function of the demand draw jumps when the firm adds a layer, as shown in Figure 1.4.\textsuperscript{18}

\textsuperscript{18}The optimal output level is substituted into the firm’s AVC function for calculating the AVC as a function of the demand draw.
1.2.4 The Spans of Control, Wages, and Relative Wages

The incentive-based hierarchy proposed above has predictions for firm-level outcomes. This subsection presents predictions on the spans of control, wages, and relative wages. The first two variables increase with the demand draw given the number of layers, and they decrease discontinuously when firms add a layer as in Caliendo and Rossi-Hansberg (2012). In addition, the relative wage, defined as the ratio of the supervisor’s wage to his direct subordinate’s wage, behaves in a way consistent with the findings of Caliendo, Monte, and Rossi-Hansberg (2012).

What happens to the firm-level outcomes when the firm expands due to an improvement in the quality of its product and keeps the number of layers unchanged? Proposition 4 summarizes the results.

**Proposition 4** Given the number of layers, both the span of control and wages increase with the firm’s quality draw at all layers. Furthermore, relative wages increase with the firm’s quality draw at all layers as well.

Proof. See Appendix 1.7.7.

The change in the span of control is the key to understanding this proposition. When the firm is constrained to keep the number of layers unchanged, the only way to expand is to increase the span of control at all layers. When the span of control is larger, monitoring is less effective, which implies that higher wages are needed to incentivize workers. Furthermore, wages increase disproportionately more at upper layers. The share of workers at upper layers in total employment decreases, when the firm grows without adjusting the number of layers. Thus, the firm tolerates disproportionately more increases in wages at upper layers while keeping increases in wages at lower layers relatively small.
The results of Proposition 4 are consistent with the evidence presented in Caliendo, Monte, and Rossi-Hansberg (2012). First, given the number of layers, wages are found to increase with firm size in both cross-sectional and time-series regressions. These results are what the model predicts, as the bigger demand draw leads to bigger firm size. Second, and more importantly, relative wages are shown to increase with firm size at all layers when the firm does not change the number of layers. This prediction implies that, although workers all gain when a firm expands without reorganization, workers at higher layers gain more. This unambiguous prediction is a unique prediction of my model, as the model presented in Caliendo and Rossi-Hansberg (2012) is silent on how relative wages change when the firm expands.

When the firm chooses to add a layer owing to an improvement in the quality of its product, the firm-level outcomes move in the opposite direction, as summarized by the following proposition.\footnote{As in Caliendo and Rossi-Hansberg (2012), I assume that the firm adds a layer from above. As the entrepreneur is at layer zero, layer $i$ becomes layer $i + 1$ where $i \geq 1$, when the entrepreneur adds a layer.}

**Proposition 5** When the firm adds a layer owing to a marginal improvement in the quality of its product, both the span of control and wages fall at existing layers. Furthermore, relative wages decrease at existing layers as well.

Proof. See Appendix 1.7.8.

The change in the span of control is again the key to understanding this proposition. When the firm expands by adding a layer, the constraint at the top (i.e., the fixed supply of entrepreneurs) is relaxed. Thus, the firm can expand and economize on its labor cost at the same time. As a result, the span of control decreases at existing layers, which leads to lower wages paid to employees at existing layers. On top of that, wages fall disproportionately more at upper layers. The share of workers at upper layers in total employment increases after an addition of a layer, since the
span of control is reduced. Consequently, it is an efficient way to economize on labor cost by reducing their wages disproportionately more.

The results of Proposition 5 are consistent with the evidence presented in Caliendo, Monte, and Rossi-Hansberg (2012) as well. First, wages are found to decrease at existing layers when firms expand by adding a layer. Second, relative wages fall at existing layers as well for firms that expand and add a layer. In total, the model’s predictions on wages and relative wages are consistent with the empirical findings presented in Caliendo, Monte, and Rossi-Hansberg (2012).

1.2.5 Firm Productivity

I discuss firm productivity and its relationship with output in this subsection. Following Caliendo and Rossi-Hansberg (2012), I use the inverse of unit costs (i.e., output divided total costs) to measure firm productivity and analyze how the unit costs vary with output as well as the number of layers.

Formally, the unit costs given a number of layers and the inefficiency of MT are defined as

\[
UC_T(q, b) \equiv \frac{TVC_T(q, b) + f_0}{q} = AVC_T(q, b) + AFC(q),
\]

(1.18)

where \(f_0\) is the fixed production cost, and \(AFC(q)\) is the average fixed cost. The unit costs given the inefficiency of MT are defined as

\[
UC(q, b) \equiv \frac{TVC(q, b) + f_0}{q} = AVC(q, b) + AFC(q).
\]

(1.19)

As the fixed production cost does not affect the choice of the number of layers, equation (1.19) can be restated as

\[
UC(q, b) = UC_T(q, b), \quad \forall q \in [q_{T-1}, q_T),
\]

(1.19)
where \( q_T \) is defined as the solution to \( AVC_T(q_T, b) = AVC_{T+1}(q_T, b) \). In what follows, I discuss how \( UC_T(q, b) \) and \( UC(q, b) \) vary with output.

First, the curve of unit costs given the number of layers and the inefficiency of MT is “U”-shaped. In other words, it decreases first and increases afterwards. Note that the average fixed cost (AFC) always decreases with output, while the AVC always increases with output. The decrease in AFC dominates the increase in AVC when output increases from a low level and vice versa. Thus, the unit costs given a number of layers decrease until output exceeds a threshold. Furthermore, the slope of the curve approaches zero when output goes to infinity, as both the decrease in AFC and the increase in AVC triggered by an increase in output become infinitesimally small.

Second, I discuss the relationship between curves of the unit costs given various numbers of layers. I define the minimum efficient scale (MES) given a number of layers as the scale of production at which a firm minimizes unit costs given the number of layers, and the minimum unit costs (MUC) given a number of layers as the unit costs when the scale of production is at its MES. Mathematically, the MES given \( b \) and \( T \) is defined as

\[
q_{Tm}(b) \equiv \arg \min_q UC_T(q, b).
\]

(1.20)

And the MUC given \( b \) and \( T \) is defined as

\[
MUC_T(b) \equiv UC_T(q_{Tm}(b), b).
\]

(1.21)

What is the relationship between various \( MUC_T(b) \) given different numbers of layers? The following assumption assures that \( MUC_T(b) \) decreases with \( T \), which implies that the hierarchy with more layers has a lower MUC.

**Assumption 1** \( f_0 > 4b\psi \).

Under this assumption, firm productivity has an increasing overall trend with respect to output. On the contrary, firm productivity has an decreasing overall trend with
respect to output, if Assumption 1 is violated.\footnote{One implication here is that the positive correlation between firm productivity and size is stronger in economies with better MT, and this correlation may be negative when the quality of MT is sufficiently low. Interestingly, Bartelsman, Haltiwanger, and Scarpetta (2013) found that the positive correlation between firm productivity and size is much stronger in developed countries such as the U.S. and Germany compared with developing countries such as Hungary and Slovenia. In Romania, the covariance between firm productivity and size is even negative. Although there are no management data for firms in Hungary, Slovenia and Romania, firms in Poland which is also a central European country have lower management scores compared with firms in the U.S. and Germany. Therefore, the finding in Bartelsman, Haltiwanger, and Scarpetta (2013) supports my productivity measure.} Economically, Assumption 1 requires that MT is efficient enough (i.e., $b$ is small enough). I assume that Assumption 1 holds in what follows, since the estimated parameters from a calibrated model presented in the next chapter satisfy this constraint.

Now, I characterize properties of $q_{Tm}(b)$, $MUC_T(b)$, and $UC_T(q,b)$ using the following proposition.

**Proposition 6** Given the number of layers and the inefficiency of MT, the curve of unit costs is “U”-shaped, and the slope of it approaches zero when output goes to infinity. Under Assumption 1, the MES given the number of layers increases with the number of layers; the MUC given the number of layers decreases with the number of layers.

Proof. See Appendix 1.7.9.

The inefficiency of MT is the key to understanding this proposition. Improved MT makes the MES increase and the MUC decrease given the number of layers. Moreover, firms with more layers gain disproportionately more from such an improvement, as the share of the total variable cost in total costs is bigger for these firms. Thus, the MES increases more and the MUC decreases more for firms with more layers after an improvement in MT. As a result, the MES increases with the number of layers, and the MUC decreases with the number of layers when MT is efficient enough.
As firm productivity is simply the inverse of its unit costs, I characterize the overall shape of the curve of firm productivity using the following corollary.

**Corollary 1** Given the number of layers, the curve of firm productivity is inversely “U”-shaped. Under Assumption 1, the maximum value of firm productivity given the number of layers increases with the number of layers.

Figure 1.5 illustrates how firm productivity varies with output. The overall trend is that firm productivity increases with output. This implies that small firms are less productive on average. Aggregate productivity increases if the smallest firms exit the market after some shocks. This is one impact of improved MT on aggregate productivity which I will analyze in the next section.
1.3 General Equilibrium Analysis

I close the model by aggregating across firms and solve for the general equilibrium in this subsection. There are two product markets and one labor market. Entrepreneurs decide whether or not to enter the CES sector and must be indifferent between entering or not in equilibrium due to the large pool of potential entrepreneurs. Workers choose which sector and which labor submarket to enter. They must be indifferent between sectors and various labor submarkets in equilibrium, since they can freely move across sectors and firms.

1.3.1 Product Market Equilibrium

There are two equilibrium conditions for the CES sector: the zero cutoff payoff (ZCP) condition and the free entry (FE) condition. They are used to pin down two equilibrium variables: the exit cutoff for the quality draw (i.e., $\bar{\theta}$) and the mass of active firms in equilibrium (i.e., $M$). First, the ZCP condition that firms with the quality draw $\bar{\theta}$ earn zero payoff can be written as

$$\Pi(\bar{\theta}, A) = 0,$$

(1.22)

where $\Pi(\bar{\theta}, A) \equiv \pi(\bar{\theta}, T(\bar{\theta}, A), A) - (f_0 + \psi)$ is the entrepreneur’s payoff. This condition pins down the exit cutoff $\bar{\theta}$ given the adjusted market size $A$. Note that the ZCP condition here incorporates both the fixed cost to produce and the cost of exerting effort, as entrepreneurs of active firms exert effort to monitor their subordinates in equilibrium. For simplicity, I use $f \equiv f_0 + \psi$ to denote the overall “fixed cost” to produce.

---

21I will explain what labor submarkets mean in what follows.
The FE entry condition implies that the expected payoff obtained from entering the CES sector equals the outside option of entrepreneurs, or

\[ \int_{\bar{\theta}}^{\infty} \Pi(\theta, A)g(\theta)d\theta = f_e, \]

where \( f_e \equiv f_1 + f_2 \) is the overall opportunity cost to enter the CES sector, and \( g(\theta) \) is the probability density function (PDF) of the quality draw \( \theta \). This equation determines the adjusted market size \( A \) given the exit cutoff \( \bar{\theta} \).

The mass of firms is undetermined in the homogeneous sector, and the managerial talent is not needed for firms in the homogeneous sector. Given the assumptions of a linear production technology and perfect competition in the homogeneous sector, firm boundaries are not defined in that sector. Therefore, I assume that entrepreneurs choose whether or not to enter the CES sector. In equilibrium, the FE condition holds with equality if and only if

\[ N \geq \frac{M}{1 - G(\bar{\theta})}, \]

where \( M \) is the mass of active firms in equilibrium, and \( G(\theta) \) is the cumulative distribution function (CDF) of \( \theta \). A sufficiently large \( N \) ensures that the above inequality holds.

The equilibrium condition for the homogeneous sector is that supply of the homogeneous good equals the demand for it, or

\[ p_h L_h = (1 - \gamma)E, \]

where \( L_h \) is the number of workers in the homogeneous sector. This condition pins down \( p_h \), which is the price of the homogeneous good as well as workers’ wages in this sector.
1.3.2 Labor Market Equilibrium

The labor market in the CES sector is characterized by competitive search. Firms demand workers for each layer, and a worker chooses one type of job to apply for in order to maximize the expected payoff. Firms randomly select workers among those who come to apply for jobs to employ. A type of job corresponds to a firm-layer pair \((\theta, i)\), as different firms offer different wages for various positions (i.e., layers). In other words, there are labor submarkets indexed by \((\theta, i)\) in the CES sector. As workers are homogeneous and can freely choose which type of job to apply for, the expected payoff from applying for any type of job must be the same in equilibrium. Moreover, this uniform expected payoff must be equal to the wage offered in the homogeneous sector, which is the outside option of workers entering the CES sector. In total, I have

\[
\frac{m_i(\theta)}{Q(\theta, i)}(w_i(\theta) - \psi) = \frac{m_i'(\theta')}{Q(\theta', i')}(w_i'(\theta') - \psi) = p_n \quad \forall (i, i') \forall (\theta, \theta'),
\]

(1.25)

where \(m_i(\theta)\) is the firm’s labor demand at layer \(i \geq 1\), and \(Q(\theta, i)\) is the number of workers who come to apply for this type of job. \(\frac{m_i(\theta)}{Q(\theta, i)}\) is the probability of being employed in labor submarket \((\theta, i)\), and \((w_i(\theta) - \psi)\) is the net payoff of being employed. Different job turn-down rates across labor submarkets (i.e., \(\frac{Q(\theta, i) - m_i(\theta)}{Q(\theta, i)} \geq 0\)) are needed to equalize the expected payoff obtained from entering various labor submarkets. As a result, there is unemployment in equilibrium.

I derive the labor-market-clearing condition in two steps. First, the number of workers who choose to enter the CES sector (i.e., \(L_c\)) can be derived from the worker’s
indifference condition in equation (1.25), or

\[
L_c = \int_{\theta=\bar{\theta}}^{\infty} \sum_{i=1}^{\infty} Q(\theta, i) \frac{Mg(\theta)}{1 - G(\bar{\theta})} d\theta
\]

\[= \frac{WP(\bar{\theta}, A, M) - \psi LD(\bar{\theta}, A, M)}{p_h}, \tag{1.26}\]

where \(WP(\bar{\theta}, A, M)\) is the total wage payment in the CES sector, and \(LD(\bar{\theta}, A, M)\) is the number of workers employed in the CES sector\(^{22}\), or

\[
LD = \int_{\theta=\bar{\theta}}^{\infty} T(\theta, A) \sum_{i=1}^{\infty} m(\theta, i) \frac{Mg(\theta)}{1 - G(\bar{\theta})} d\theta. \tag{1.27}\]

Equation (1.26) has an intuitive explanation. It says that the total expected payoff of workers entering the CES sector (i.e., \(p_h L_c\) due to the indifference condition) is equal to the difference between the total wage payment and the total disutility to exert effort.

Second, the labor-market-clearing condition indicates that the number of workers employed in the homogeneous sector is the difference between the endowment of labor and the number of workers who choose to enter the CES sector, or

\[L_h = L - L_c. \tag{1.28}\]

Equations (1.26) and (1.28) are two labor market equilibrium conditions that are used to determine the allocation of labor between sectors.

\(^{22}\)For further discussion of the labor market equilibrium in the CES sector, see Appendix 1.7.10.
1.3.3 Equilibrium and Unemployment

The market-clearing condition of the final composite good implies that

\[ E = \int_{\theta}^{\infty} LC(\theta) \frac{Mg(\theta)}{1 - G(\theta)} d\theta + p_h L_h + \left[ f_0 + f_1 \left( \frac{\theta}{\theta_{\text{min}}} \right)^k \right] M + \left[ \psi + f_2 \left( \frac{\theta}{\theta_{\text{min}}} \right)^k \right] M, \]  

(1.29)

where \( LC(\theta) \) is the total wage payment of firms with demand draw \( \theta \). The third part of the right hand side (RHS) of equation (1.29) is the demand for the final composite good by firms, and the last part of the RHS of equation (1.29) is the consumption of active entrepreneurs who earn profit in equilibrium.\(^{23}\) Total income of the economy equals total expenditure which includes two parts: demand from workers and demand from firms. Note that only workers and active entrepreneurs demand goods in equilibrium. Entrepreneurs who choose not to enter the CES sector receive their outside option without consuming goods; entrepreneurs who enter the CES sector and choose not to produce don’t consume goods as their income is zero.

The general equilibrium of this economy is characterized by the quality threshold of the firm that obtains zero payoff, \( \bar{\theta} \), the mass of firms that operate \( M \), the price of the homogeneous good \( p_h \), the labor allocation between two sector, \( L_c \) and \( L_h \), and the aggregate income \( E \). These six equilibrium variables are obtained by solving the six equations (i.e., equations (1.22), (1.23), (1.24), (1.26), (1.28) and (1.29)). Obviously, one equilibrium condition is redundant due to Walras’ law, and I normalize the price of the final composite good to one.

One implicit assumption for the existence of the equilibrium is that the probability of being employed implied by equation (1.26) is smaller than or equal to one in every labor submarket in equilibrium. In other words, wages offered in the CES sector must

\(^{23}\)The equilibrium condition stated in equation (1.24) has used the FE condition described above. Ex post profit per active firm must compensate both the cost of exerting effort and the forgone outside option. The overall forgone utility is \( f_2 \left( \frac{\theta}{\theta_{\text{min}}} \right)^k M \), and total utility cost of exerting effort is \( \psi M \). Therefore, profit per active firm that compensates these costs is \( \psi + f_2 \left( \frac{\theta}{\theta_{\text{min}}} \right)^k \).
satisfy
\[ w_i(\theta) - \psi \geq p_h \quad \forall \, (i, \theta), \tag{1.30} \]
where \( w_i(\theta) \) is determined in equation (1.8). The above inequality would be violated if \( \psi \) were zero. Firms do not need to pay incentive-compatible wages to workers, if exerting effort does not generate any cost to them. At the same time, there is no unemployment in all labor submarkets, and every worker in the CES sector receives the same wage. I don’t consider this case, since information friction and MT do not matter in this case. In the paper, I focus on the case in which unemployment exists in every labor submarket, and the incentive-compatible wage determined in equation (1.8) satisfies the constraint specified in equation (1.30) in every labor submarket.

There are three reasons why I want to investigate this type of equilibrium. First, the model yields clean insights and testable implications in this case. Second, the testable implications at the micro-level (e.g., wages, relative wages, and the optimal number of layers etc.) have been shown to be consistent with the evidence presented in Caliendo, Monte, and Rossi-Hansberg (2012). Finally, as shown below, the model’s predictions at the aggregate level (i.e., effects of improved MT on the firm size distribution and firm organization) are consistent with the evidence presented in Hsieh and Klenow (2009, 2012) and Bloom et al. (2013) as well. The following proposition discusses the existence and uniqueness of an equilibrium with unemployment in every labor submarket.

**Proposition 7** When \( \frac{\sigma - 1}{\sigma} > \gamma \), there exists a unique equilibrium with unemployment in every labor submarket, if the labor endowment (i.e., \( L \)) is small enough; When

\[ \frac{\sigma - 1}{\sigma} \leq \gamma \]

the outside option of workers (i.e., \( p_h \)) is not too small in equilibrium, the equilibrium has the property that some labor submarkets have unemployment and the others don’t. In this case, firms that would offer lower incentive-compatible wages in the absence of the outside option of workers are forced to raise wages up to \( p_h + \psi \). I discuss this case in Appendix 1.7.1.4 and show that qualitative results of the model in this case are the same as the ones we are going to derive in the paper.
\[ \frac{\sigma - 1}{\sigma} < \gamma, \] 
there exists a unique equilibrium with unemployment in every labor submarket if the labor endowment is big enough.

Proof. See Appendix 1.7.10

The labor endowment affects the outside option of workers through two channels. First, a bigger labor endowment reduces workers’ wage in the homogeneous sector as a result of the supply-side effect. Second, a bigger labor endowment increases total income of the economy, which leads to a larger demand for labor and increases the wage in the homogeneous sector. The relative strength of these two effects depends on parameter values. When \( \frac{\sigma - 1}{\sigma} > \gamma \), the demand-side effect dominates. Thus, a sufficiently small labor endowment insures that the workers’ outside option is small enough in equilibrium, which validates the existence of a unique equilibrium with unemployment in every labor submarket and vice versa.\(^{25}\)

Admittedly, the condition assuring the existence of a unique equilibrium with unemployment in every labor submarket involves endogenous variables.\(^{26}\) This is because both wages offered in the CES sector and the wage offered in the homogeneous sector cannot be solved analytically. However, when I treat the number of layers as a continuous variable à la Keren and Levhari (1979) and Qian (1994), the condition can be stated using exogenous parameters only. Readers are referred to Appendix 1.7.13 for more details.

Aggregate labor demand that takes into account the product-market-clearing conditions either increases or decreases in \( p_h \). There are two countervailing effects on the aggregate labor demand, when the workers’ outside option increases. On the one hand, the homogeneous sector demands less labor when \( p_h \) goes up, and the number of job applicants that equalizes the expected payoff obtained from entering the two

\(^{25}\)The wage determination in the CES sector features that buyers (i.e., firms) have all the bargaining power when the wage contracts are offered. This implies that changes in labor endowment do not affect incentive-compatible wages, as long as there is unemployment in the labor submarkets.

\(^{26}\)The condition assuring the existence and uniqueness of the equilibrium is that \( \frac{\sigma - 1}{\sigma} \neq \gamma \) and \( p_h \leq w_{\min}(\theta) - \psi \), where \( w_{\min}(\theta) \) is the lowest wage among wages offered in the CES sector. This is a sufficient and necessary condition for the existence and uniqueness of the equilibrium.
sectors goes down due to the higher outside option. On the other hand, the higher expected wage increases the total income of the economy, which makes both sectors demand more labor. The relative strength of these two offsetting effects depends on parameter values. When \( \frac{\sigma - 1}{\sigma} < \gamma \), the first effect dominates and vice versa. \( \frac{\sigma - 1}{\sigma} = \gamma \), the aggregate labor demand does not respond to the change in \( p_h \). Thus, there is either no equilibrium or infinitely many equilibria depending on values of other parameters in this knife edge case. Thanks to the uniqueness of the equilibrium under restrictions on parameter values, I can analyze how an improvement in MT affects various economic activities.

I close this section by discussing the role of the unemployment rate in this model. Although the wage determination in the current model is similar to the one used in the efficiency wage theory (e.g., Shapiro and Stiglitz (1984)), the role of the unemployment rate is different. In the efficiency wage theory, the aggregate unemployment rate feeds back to the incentive-compatible wage in a dynamic setup, and unemployment is present only when there exists an exogenous separation rate between firms and workers\(^{27}\). In this paper, unemployment (or being fired) still serves as a disciplinary device to incentivize workers to exert effort. However, unemployment rates (more precisely, job-turn-down rates) across different labor submarkets are used to equalize the expected payoff obtained from entering different labor submarkets. Essentially, the role of unemployment rate in my model is the same as in Harris and Todaro (1970) and is similar to the role played by the labor market tightness in the literature on competitive search (e.g., Moen (1997) etc.).\(^{28}\)

\(^{27}\)Remember that every agent is incentivized to work in Shapiro and Stiglitz (1984).

\(^{28}\)More specifically, the difference in the role of unemployment rate between the efficiency wage theory and my model comes from the assumption of how the firm punishes misbehaving workers. In my static model, the firm punishes shirking workers whose misbehavior has been detected by firing them and reducing the wage payments. In the efficiency wage theory, the firm punishes shirking workers whose misbehavior has been detected by firing them but not reducing the wage payments. Thus, the incentive to work comes from a decrease in future income due to unemployment in the efficiency wage theory. It is probably true that workers do get wage cuts when their performance does not meet some goals that have been preset in practice.
1.4 Management Technology, Institutional Quality, and Firm-Level Outcomes

This section investigates how an improvement in MT affects firm characteristics as well as welfare. Ample evidence suggests that there are substantial differences in the quality of MT across countries due to factors that are beyond the control of firms. Furthermore, Hsieh and Klenow (2009, 2012) showed that China and India whose firms receive low management scores have more firms of a small size and fewer efficient firms with large market shares than the U.S. Finally, firm organization differs across countries due to differences in MT and affects firm size and performance as well. More specifically, firms in India are, compared with those in the U.S., less decentralized owing to worse MT and weak enforcement of laws, and the low level of decentralization impedes Indian firms’ expansion (e.g., Bloom, Sadun, and Van Reenen (2012a), Bloom et al. (2013)). The purpose of this section is to show that there is a link between the quality of MT and the firm characteristics discussed above.

This section is divided into two subsections. In the first subsection, I consider the case in which the quality of MT only differs across countries and investigate how an improvement in management quality affects economic outcomes. In the second subsection, I consider the case in which MT differs both across firms and across countries in order to derive a testable hypothesis that will be tested in Section 1.5.

1.4.1 Selection Effect of Better Management Technology

I consider a scenario in which the MT that is common across all firms improves. Such an improvement is equivalent to a decrease in \( b \) in the model, since it becomes easier for the firm to catch and fire shirking workers after the change. As a result, firms’ labor costs decrease, since workers’ wages are determined by the incentive compatibility constraint.
An improvement in MT generates a pro-competitive effect that reallocates resources toward more efficient firms. This improvement favors more efficient firms, since they have more layers. More specifically, an improvement in MT benefits all firms since it reduces firms’ labor cost. Moreover, firms with more layers gain disproportionately more, as their AVCs increase less rapidly with output. More precisely, the AVC functions of firms with more layers have smaller elasticities with respect to output. As a result, firms with the worst demand draws are forced to leave the market; firms whose demand draws are in the middle receive shrinking revenue and profit; and firms with the best demand draws expand. In other words, an improvement in MT facilitates inter-firm resource allocation through benefitting bigger firms more, which is exactly what Bloom et al. (2013) argued in their paper.

Endogenous selection of a management hierarchy with a specific number of layers is the key to understanding the selection effect of an improvement in MT. In a hypothetical world, if all firms were forced to have the same number of layers, the uneven effect would disappear. This is because all firms would have the same AVC function in such a world. As a result, the exit cutoff for the demand draw would be unaffected by an improvement in MT. Furthermore, firms’ revenue and profit would be unchanged as well. In short, the pro-competitive effect (i.e., the selection effect) is present only when firms endogenously choose to have different numbers of layers.

How does the internal organization of firms evolve when MT improves? Bloom et al. (2013) found that Indian firms are unwilling to decentralize their production processes (i.e., constrained span of control), because it is hard to catch and punish misbehaving employees in India. Furthermore, they argued that poor monitoring and weak enforcement of laws are reasons for why Indian firms can’t catch and punish misbehaving workers easily. Finally, they argued that low level of decentralization is one reason for why Indian firms are small on average. My model gives economic reasons rationalizing these findings. First, when firms are able to monitor their employees
more easily, the span of control increases. Second, and more importantly, the number of layers also increases weakly for each firm because of better monitoring. Firms expand when monitoring becomes more effective, and the expansion incentivizes firms to have more layers. As a result, firms have fewer layers or constrained span of control in economies with worse MT. Furthermore, less decentralized production processes are associated with smaller average firm size as shown below. In total, firms are less decentralized in economies with poor MT, and the low level of decentralization is one reason for why firms are small in these economies.

The predictions on the internal structure of firms are unique predictions of my model. In knowledge-based hierarchy models (e.g., Garicano (2000)), an improvement in ICT flattens firms’ hierarchies. That is, firms delayer when ICT improves (See Garicano and Van Zandt (2012) for details). The intuition is that the impact of an improvement in ICT is heterogeneous across layers conditional on output. When communication becomes more efficient, firms increase the knowledge learned by workers at upper layers and reduce the knowledge learned by workers at lower layers. Furthermore, the number of layers is reduced as well. When the costs of learning knowledge become cheaper, firms increase the knowledge learned by workers at all layers and decrease the number of layers. However, an improvement in MT does not have such a heterogeneous effect across layers in this paper, since monitoring is done layer by layer. In other words, two non-adjacent layers do not interact with each other in the management hierarchy considered in this paper. Therefore, better MT affects different layers evenly. As a result, firm size (i.e., output or employment) is a sufficient statistic to determine the optimal number of layers.

In order to derive analytical results on the firm size distribution and the distribution of the number of layers, I assume that \( \theta \) follows a Pareto distribution with a

\[ \text{Note that output and employment go up for all firms. However, revenue and operating profit fall for small firms after MT improves.} \]
coefficient \( k \), or

\[
G(\theta) = 1 - \left( \frac{\theta_{\min}}{\theta} \right)^k,
\]

where \( G(\theta) \) is the CDF of the demand draw \( \theta \). The following proposition summarizes the changes in firm characteristics due to an improvement in MT. Note that the above distributional assumption is only needed for the results on the firm size distribution and the distribution of the number of layers.

**Proposition 8** Suppose management technology that is common across all firms improves. Consider the case in which the minimum number of layers among active firms is unchanged. For the economy as a whole, the exit cutoff for the quality draw increases. At the firm level, all surviving firms either increase the number of layers (weakly) or make the span of control bigger and keep the number of layers unchanged. Finally, if the quality draw follows a Pareto distribution, both the firm size distribution and the distribution of the number of layers move to the right in the First-Order-Stochastic-Dominance (FOSD) sense.

Proof: See Appendix 1.7.11.

I focus on the case in which the minimum number of layers of active firms is unchanged when MT improves, although similar results emerge in the other cases. The reason why I focus on this case is that there are always some extremely small firms that have only two layers (i.e., \( T = 1 \)) in every economy of the world. Therefore, the case considered in the paper is empirically more relevant. Furthermore, I prove that all the results of Proposition 8 except for the prediction on the span of control hold, when the number of layers is treated as a continuous variable à la Keren and Levhari (1979) and Qian (1994). In total, the aggregate-level predictions (i.e., changes in the exit cutoff, the firm size distribution, and the number of layers) derived in the

\[30\text{The span of control is a constant which is not affected by the quality of MT and the demand draw in the continuous case. Readers are referred to the Appendix 1.7.13 for details.}\]
case of a continuous number of layers are qualitatively the same as those derived in
the case of a discrete number of layers.

The FOSD results have important implications for resource allocation in the e-
conomy and are consistent with the data. First, the FOSD result for the firm size
distribution implies that there are fewer small firms in economies with superior MT.
Furthermore, firms with better demand draws have bigger sales, and average firm size
is bigger in such economies as well. These theoretical predictions show the key role
played an improvement in MT in determining resource allocation. That is, resources
are reallocated toward more efficient firms. Second, the FOSD result for the distri-
bution of the number of layers implies that firms are less decentralized in economies
with worse MT, which is one important reason for why efficient firms don’t have large
enough market shares in such economies. Finally, these theoretical predictions are
consistent with several key findings and conjectures of a number of recent papers. For
instance, Hsieh and Klenow (2009) showed that India and China are worse at getting
efficient firms to obtain big market shares compared with the U.S. Bloom, Sadun,
and Van Reenen (2012b) found that a given amount of increase in the management
score has a bigger positive impact on firm size in the U.S. than in other countries.
As management scores are highly correlated with firm productivity, this implies that
more productive firms gain more in economies with better MT such as in the U.S.
This finding is exactly the key prediction of my model. Next, Hsieh and Olken (2014)
argued that it is the big firms that are constrained more in developing countries, not
the small firms. This conjecture is also consistent with the key result of this paper.
In short, the key results of this chapter seem to square well with existing evidence.

Other than changes in the firm size distribution and the internal organization
of firms, the weighted average of firm productivity also increases as a result of an
improvement in MT. The weighted average of firm productivity is defined as the sum
of the product of firm productivity and its market share across all firms.\textsuperscript{31} Gains in the weighted average of firm productivity come from three sources. First, firms with the worst demand draws which are less productive on average exit the market after an improvement in MT (i.e., the between-firm effect). Second, market shares of more productive firms increase, because improved MT favors more productive firms. This makes the weighted average of firm productivity increase as well (i.e., the between-firm effect). Finally, the productivity of all surviving firms increases, as improved MT reduces firms’ costs (i.e., the within-firm effect). In total, these three effects together increase the weighted average of firm productivity, as shown in Figure 1.6.

Other than firm-level outcomes, I am also interested in how improved MT affects the worker’s welfare. As workers can freely move between two sectors, the expected

\textsuperscript{31}The market share is either the firm’s share in the sector’s total output or its share in the sector’s total sales.
payoff obtained from entering the CES sector must equal the wage offered in the homogeneous sector. Therefore, the worker’s expected payoff obtained from entering the CES sector is a sufficient statistic to evaluate welfare. In what follows, I discuss how this changes when MT improves.

Better MT can either increase or decrease welfare due to multiple frictions in the model. First, there is a moral hazard problem inside the firm due to information frictions. Second, there is monopolistic distortion in one of the two sectors of this economy. Finally, there is a labor market friction due to random search. As a result, unemployment exists in labor submarkets. Therefore, a reduction in one friction does not necessarily increase welfare. It turns out that the factor governing the direction of the change in welfare is the elasticity of substitution between products in the CES sector, since it determines whether or not the CES sector expands after an improvement in MT.

Workers in the CES sector face a trade-off between lower wages and higher probabilities of being employed. When MT improves, employed workers receive lower wages and payoffs (i.e., wages minus the disutility to exert effort) on average. However, firms expand and demand more labor due to better MT. On top of that, the elasticity of substitution determines the sensitivity of the firm’s expansion (i.e., the increase in average employment) with respect to an improvement in MT. When products are more substitutable, this sensitivity is higher. Thus, the increase in employment per firm is bigger. Moreover, the bigger increase in average employment eventually increases the aggregate income of the economy which makes the market size bigger. As a result, the CES sector accommodates more firms, and its aggregate labor demand increases. This increase reduces the risk of being unemployed for workers in the CES sector. In summary, the increase in the average probability of being employed dominates wage

\(^{32}\)Remember that average firm size in terms of employment increases when MT improves.
loss, when MT improves and the elasticity of substitution is high. As a result, the worker’s expected payoff obtained from entering the CES sector increases.

The opposite story happens when the elasticity of substitution is low. In this case, the increase in employment per firm is small when MT improves. This small increase in employment per firm and the decrease in the average wage eventually push down the aggregate income of the economy, which makes the market size smaller. As a result, the CES sector accommodates fewer firms and its aggregate labor demand decreases. Therefore, workers obtain a lower expected payoff by entering the CES sector, as both the average wage and the average employment rate in the CES sector decrease when the elasticity of substitution is lower. Note that although firms with the worst demand draws are driven out of the market when the MT improves, the ideal price index for the CES sector increases. This is because the decrease in the mass of firms dominates the decrease in the average price charged by active firms. In total, welfare decreases when the elasticity of substitution is sufficiently small.

The above discussions on welfare are not analytical results, although simulation results do show that welfare can either increase or decrease after an improvement in MT. Table 1.1 presents an example in which welfare increases when MT improves, while Table 1.2 shows an example in which welfare decreases when MT improves. Importantly, when the number of layers is treated as a continuous variable, welfare increases after an improvement in MT if and only if \( \frac{\sigma-1}{\sigma} > \gamma \). This qualitative result is shown in Appendix 1.7.13.

Table 1.1: Change in Welfare when MT Improves and \( \sigma \) is Big

<table>
<thead>
<tr>
<th></th>
<th>Welfare</th>
<th>Ave(wage)</th>
<th>Ave(ur)</th>
<th>M</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>b=1.6</td>
<td>0.29</td>
<td>0.96</td>
<td>0.56</td>
<td>1.02</td>
<td>25.66</td>
</tr>
<tr>
<td>b=1.5</td>
<td>0.35</td>
<td>0.90</td>
<td>0.42</td>
<td>1.15</td>
<td>31.33</td>
</tr>
</tbody>
</table>

ur: unemployment rate; M: the mass of active firms; E: total income
\( \sigma = 3.8, \gamma = 0.6, \psi = 0.3 \)
Table 1.2: Change in Welfare when MT Improves and $\sigma$ is Small

<table>
<thead>
<tr>
<th>b</th>
<th>Welfare</th>
<th>Ave(wage)</th>
<th>Ave(ur)</th>
<th>M</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.6</td>
<td>0.38</td>
<td>0.95</td>
<td>0.41</td>
<td>2.61</td>
<td>53.49</td>
</tr>
<tr>
<td>1.5</td>
<td>0.29</td>
<td>0.90</td>
<td>0.51</td>
<td>1.90</td>
<td>41.07</td>
</tr>
</tbody>
</table>

ur: unemployment rate; M: the mass of active firms; E: total income
$\sigma = 2.8, \gamma = 0.75, \psi = 0.3$

1.4.2 Two-Dimensional Heterogeneity in Management Quality

The data from the World Management Survey (WMS) show that there is substantial heterogeneity in the quality of MT both across countries and across firms. Thus, I now consider a case in which MT differs both across economies and across firms. More specifically, I assume that firms receive both demand draws (i.e., $\theta$) and draws of the inefficiency of MT (i.e., $b$) after entry. Furthermore, these two draws are assumed to be independent. Under these assumptions, I show that the results established in Proposition 8 hold in this case as well. The detailed proof is relegated to Appendix 1.7.15 as it is similar to the proof of Proposition 8. Moreover, because of the variation in MT across firms, the model yields the following testable prediction which will be tested in Section 1.5.

**Proposition 9** Suppose management technology differs both across economies and across firms. An improvement in a firm’s own management technology increases its firm size, the number of layers (weakly), and the span of control given the number of layers. However, an improvement in all other firms’ management technology holding constant the firm’s own management technology reduces the firm’s size.

Proof. See Appendix 1.7.12

---

$^{33}$Now, there are one FE condition and $n$ ZCP conditions, if there are $n$ possible outcomes of the draw of $b$. More specifically, the exit cutoff for the quality draw depends on both the adjusted market size, $A$, and the draw of $b$. 

46
The key to understanding the above testable prediction is the pro-competitive effect of an improvement in MT. On the one hand, when the MT improves only for a single firm, its firm size and profit unambiguously increase as better MT reduces labor costs. Furthermore, the firm becomes more decentralized by increasing either the span of control or the number of layers. On the other hand, when MT improves on average for all firms except for one firm, that firm’s revenue and profit shrink because of the intensified competition. In total, variations in the quality of MT across firms and economies yield testable predictions that can be contrasted with the data.

1.5 Evidence

In this section, I present econometric evidence to confirm the model’s predictions for firm size using the data sets from the WMS.

1.5.1 Data

Two main data sets I use were obtained from the WMS, and they were originally used in Bloom and Van Reenen (2010). The first data set is a cross-sectional data set that contains roughly 5700 firms across 16 countries. Each firm was interviewed once between 2004 and 2008, and the score on each management practice was given after the firm had been interviewed. The second data set is a panel data set that contains accounting information (e.g., employment, assets and sales etc.) for firms whose management scores were given in the first data set. The time span for the second data set is from 2003 to 2008. I obtained country-level data from the website of either the World Bank or the Penn World Table. In particular, the labor market rigidity index that measures the easiness of firing workers is included in the first data set.

\footnotesize{For details of the empirical work, see both Appendix 1.7.1 and Appendix 1.7.16}
set I obtained from the WMS and is also available from the website of the World Bank.

I constructed a variable called *moinc* to measure the quality of MT to monitor and incentivize workers. There are eighteen management practices defined in Bloom and Van Reenen (2010) in total, and these practices are grouped into three categories: monitoring, targets and incentives. I picked up seven management practices from them and argue that these seven management practices are closely related to the quality of MT to monitor and incentivize workers. For instance, management practices such as “performance tracking” and “performance review” are related to whether or not the firm can successfully catch misbehaving employees. Management practices such as “rewarding high performance” and “remove poor performers” are related to whether or not the firm can credibly reward hard-working employees and punish shirking workers. I calculated the average score on these seven management practices and treat it as a measure for the quality of MT.

I used the Olley-Pakes approach to estimate firms’ total factor productivity (TFP). One complication is that I cannot estimate the production function using all firms in a given industry, since firms that have different numbers of layers have different production functions. One solution is to obtain information on the number of layers for each firm and group firms into several bins based on the number of layers they have. Unfortunately, such information is not available in the data sets I obtained from the WMS. I circumvented this problem by using firm size (i.e., deflated sales or employment) as a *proxy* for the number of layers, as the theory predicts that bigger firms have more layers. More specifically, I grouped firms in the same sector into two bins based on their deflated sales or employment and estimated the production function for firms in each bin separately using the Olley-Pakes approach.
1.5.2 Specification

I test two predictions of Proposition 9. First, the management score on monitoring and incentives at the firm level (i.e., \( moinc \)) positively affects firm size, ceteris paribus. Second, the average quality of MT of firms in an economy negatively affects a firm’s size, ceteris paribus. The second prediction is the general-equilibrium effect of an improvement in MT. I use two variables to measure the average quality of MT in an economy. First, the average value of \( moinc \) across firms within an industry is used to capture the average quality of MT in an industry. Second, as the labor market rigidity index negatively affects firms’ ability to incentivize workers, I use it as a proxy for the average quality of MT in a country.\(^{35}\)

More specifically, I run two regressions. I consider the following specification first.\(^{36}\)

\[
\log(size_{it}) = \beta_0 + \beta_1 \log(TFP)_{it} + \beta_2 moinc_{it} + \beta_3 \text{rigidity}_c + \text{yearind}_{tj} + \text{controls}_{ct} + \epsilon_{it},
\]

(1.32)

where \( i \) is the firm ID; \( j \) indicates the industry the firm belongs to; \( c \) represents the country the firm is from; and \( t \) is the year. For the left-hand-side variable, I use both employment and deflated sales as measures for firm size. For the right-hand-side variables, \( \log(TFP)_{it} \) captures variations in any firm-level variable that affects firm productivity; \( moinc_{it} \) captures variations in the quality of MT; \( \text{yearind}_{tj} \) (i.e., the year-industry fixed effect) captures any shock that is specific to an industry, but common across countries in a given year (e.g., a drop in the world price of steel for

\(^{35}\)Based on the theory, what I need here are exogenous variations in the average quality of MT from the perspective of the firm. As the labor market rigidity index is probably out of a single firm’s control, the variation in it is what I need. As it is hard to find an exogenous variable that affects the average quality of MT for firm in an industry, I treat the variation in the average value of \( moinc \) across firms within an industry as the variation in the average quality of MT.

\(^{36}\)Since the random errors might be correlated for firms within an industry of a certain country, I cluster the standard error at the country-industry level.
the steel industry in a given year). I cannot include the country fixed effect into the regression since the labor market rigidity index does not vary across years. Instead, I include a set of country-level controls that affect firm size not through affecting the quality of MT into the regression. \( \log(GDP) \) is used to capture the market size effect; \( \log(GDP_{percapita}) \) is used to capture variations in the quality of physical capital, human capital, and scientific technology across countries; and the Doing Business index is used to capture variations in the quality of political and financial institutions. For this regression, the theory predicts that both the TFP and the management score (i.e., \( moinc_{it} \)) at the firm level positively affects firm size, ceteris paribus. On top of that, the labor market rigidity index should positively affect firm size after I have controlled for firm-level characteristics as well as country-level characteristics.

Second, I run a regression of firm size on firm-level characteristics and the average score of \( moinc \). More specifically, the specification is

\[
\log(size_{it}) = \beta_0 + \beta_1 \log(TFP)_{it} + \beta_2 moinc_{it} + \beta_3 ave \log(TFP)_{jct} + \beta_4 avemoinc_{jct} + yearctry_{tc} + yearind_{tj} + \epsilon_{it}.
\] (1.33)

For the right-hand-side variables, \( yearctry_{tc} \) (i.e., the year-country fixed effect) captures any macro-level shock that is specific to a country, but common across industries in a given year (e.g., an economic recession happening in a given year); \( ave \log(TFP)_{jct} \) (i.e., the weighted average of \( \log(TFP) \)) captures variations in the overall market competitiveness; and \( avemoinc_{jct} \) (i.e., the weighted average of \( moinc \)) captures variations in the market competitiveness due to changes in the average quality of MT. Similarly, the theory predicts that both the TFP and the management
score positively affect firm size, ceteris paribus. Moreover, the average management score should negatively affect firm size, ceteris paribus.

1.5.3 Results

Regression results reported in Tables 1.3 and 1.4 confirm the theoretical results established in Proposition 9. First, the coefficient in front of moinc is positively significant in all regressions. This is consistent with the prediction of the effect of the individual firm’s management score on firm size. Second, the coefficient in front of the labor market rigidity index is positively significant in most regressions reported in Table 1.3 even when I control for a variety of country-level variables. This is consistent with the theoretical prediction that firm size is smaller in countries with less rigid labor market conditional on firm-level characteristics. Third, the coefficient in front of the average management score is negatively significant in all regressions reported in Table 1.4 even when I control for the weighted average of log(TFP). This result is again consistent with my theoretical prediction. Namely, firm size is smaller in an economy that has better managed firms conditional on firm-level efficiencies. Finally, the impact of the average quality of MT on firm size is quantitatively important. For instance, an increase in the average management score from 2.86 (i.e., India) to 3.50 (i.e., the U.S.) leads to about 7.7% decrease in a firm’s size, ceteris paribus. In short, the model gains support from the data.

Finally, I did several robustness checks for results reported in Tables 1.3 and 1.4. First, I re-estimated the TFP by grouping firms in each sector into three bins and implementing the Olley-Pakes productivity estimation. After having obtained the new TFP estimates, I rerun the regressions specified in equations (1.32) and (1.33) and found that regression results are qualitatively similar to what are reported in Tables

---

37One potential concern here is that firms might not be randomly selected for the survey. The description of the data set in Bloom and Van Reenen (2010) shows that they randomly sampled medium-sized firms, employing between 100 and 5,000 workers.
Second, I run the regressions specified in equations (1.32) and (1.33) for firms in each bin separately, as the effect of the quality of MT on firm size might be heterogeneous across firms that have different numbers of layers. The estimation results for both groups of firms are consistent with the predictions in Proposition 9 as well. For more details on the robustness checks, see Appendix 1.7.16.

Table 1.3: Labor Market Rigidity, Management Scores, and Firm Size

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(empl)</td>
<td>log(sales)</td>
<td>log(empl)</td>
<td>log(sales)</td>
</tr>
<tr>
<td>log(TFP1)</td>
<td>0.140***</td>
<td>1.007***</td>
<td>0.0180</td>
</tr>
<tr>
<td></td>
<td>(0.0404)</td>
<td>(0.0342)</td>
<td>(0.0220)</td>
</tr>
<tr>
<td>log(TFP2)</td>
<td>0.0180</td>
<td>0.146***</td>
<td>0.223***</td>
</tr>
<tr>
<td></td>
<td>(0.0220)</td>
<td>(0.0286)</td>
<td>(0.0271)</td>
</tr>
<tr>
<td>moinc</td>
<td>0.223***</td>
<td>0.253***</td>
<td>0.236***</td>
</tr>
<tr>
<td></td>
<td>(0.0271)</td>
<td>(0.0286)</td>
<td>(0.0271)</td>
</tr>
<tr>
<td>Rigidityempl</td>
<td>0.00625**</td>
<td>0.00424</td>
<td>0.00669***</td>
</tr>
<tr>
<td></td>
<td>(0.00255)</td>
<td>(0.00273)</td>
<td>(0.00252)</td>
</tr>
<tr>
<td>Constant</td>
<td>8.231***</td>
<td>6.848***</td>
<td>7.799***</td>
</tr>
<tr>
<td></td>
<td>(1.082)</td>
<td>(0.965)</td>
<td>(1.078)</td>
</tr>
</tbody>
</table>

MNE dummies Yes Yes Yes Yes
Industry-year F.E. Yes Yes Yes Yes
N 3329 3329 3339 3339
R² 0.269 0.671 0.258 0.436
adj. R² 0.208 0.644 0.197 0.389

TFP1: estimated TFP when firms are grouped into two bins based on employment.
TFP2: estimated TFP when firms are grouped into two bins based on deflated sales.
empl: employment; Rigidityempl: employment rigidity index.
Controls: log(GDP), log(GDP/per capita), log(price index) and the Doing Business index.
Standard errors are in parentheses and clustered at the country-industry level.
* p < 0.10, ** p < 0.05, *** p < 0.01
## Table 1.4: Average Management Score and Firm Size: Industry-Level Analysis

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>log(empl)</td>
<td>log(sales)</td>
<td>log(empl)</td>
<td>log(sales)</td>
</tr>
<tr>
<td>log(TFP1)</td>
<td>0.129***</td>
<td>1.001***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0411)</td>
<td>(0.0349)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(TFP2)</td>
<td></td>
<td>0.0151</td>
<td>0.156***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0240)</td>
<td>(0.0332)</td>
<td></td>
</tr>
<tr>
<td>moinc</td>
<td>0.247***</td>
<td>0.272***</td>
<td>0.258***</td>
<td>0.363***</td>
</tr>
<tr>
<td></td>
<td>(0.0318)</td>
<td>(0.0333)</td>
<td>(0.0321)</td>
<td>(0.0409)</td>
</tr>
<tr>
<td>ave log(TFP1)</td>
<td>-0.00679</td>
<td>-0.0208</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0292)</td>
<td>(0.0260)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ave log(TFP2)</td>
<td></td>
<td></td>
<td>0.0224</td>
<td>-0.0123</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0252)</td>
<td>(0.0283)</td>
</tr>
<tr>
<td>ave(moinc)</td>
<td>-0.116*</td>
<td>-0.116**</td>
<td>-0.117*</td>
<td>-0.131*</td>
</tr>
<tr>
<td></td>
<td>(0.0607)</td>
<td>(0.0580)</td>
<td>(0.0608)</td>
<td>(0.0752)</td>
</tr>
<tr>
<td>Constant</td>
<td>5.950***</td>
<td>5.094***</td>
<td>6.302***</td>
<td>9.381***</td>
</tr>
<tr>
<td></td>
<td>(0.371)</td>
<td>(0.349)</td>
<td>(0.352)</td>
<td>(0.451)</td>
</tr>
<tr>
<td>MNE dummies</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Industry-year F.E.</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Country-year F.E.</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>3113</td>
<td>3113</td>
<td>3121</td>
<td>3121</td>
</tr>
<tr>
<td>R²</td>
<td>0.288</td>
<td>0.686</td>
<td>0.280</td>
<td>0.456</td>
</tr>
<tr>
<td>adj. R²</td>
<td>0.228</td>
<td>0.660</td>
<td>0.219</td>
<td>0.410</td>
</tr>
</tbody>
</table>

* TFP1 estimated TFP when firms are grouped into two bins based on employment.
* TFP2 estimated TFP when firms are grouped into two bins based on deflated sales.
* ave(A): weighted average of variable A at the industry-country-year level.
* Firms in industry-country-year pairs that contain only one observation are excluded.
* Standard errors are in parentheses and clustered at the country-industry level.
* p < 0.10, ** p < 0.05, *** p < 0.01
1.6 Conclusions

This paper uses one canonical approach to modeling the incentive problem inside the firm and incorporates an incentive-based hierarchy into a general equilibrium framework to show the aggregate implications of an improvement in MT. By investigating how the quality of MT affects aggregate economic outcomes, this paper rationalizes several key findings in the macro-development literature and the organizational economics literature. The key role played by improved MT is to intensify market competition and make more efficient firms obtain bigger market shares. As a result, firms are smaller, less decentralized and less productive on average in countries whose firms have worse MT. These key theoretical results are consistent with what the recent empirical literature has found from firm-level data.

Undoubtedly, much more research remains to be done. First, it is extremely important to use the current framework to quantitatively evaluate how the quality of MT affects average productivity of firms in the economy. Second, integrating the knowledge-based hierarchy and the incentive-based hierarchy into a unified framework is an interesting idea, since each approach reflects only one part of the function of the hierarchy. Finally, investigating how the quality of MT affects firms’ organizational choice (e.g., outsourcing or in-house production) is also an interesting topic to explore.

1.7 Appendix

1.7.1 Empirical Motivation

In this subsection, I discuss the details of my empirical motivation. Three parts come in order. First, I discuss the content of MT used in this paper. Next, I argue that MT is highly correlated with firm performance by showing some motivating evidence. Third, I show that quality of monitoring and incentives is an important
part of the overall management quality, and there is substantial heterogeneity on these management practices across firms. Finally, I argue that the quality of MT differs across economies and is systematically correlated with firm size distribution.

First, I discuss what MT used in this paper means. In Bloom and Van Reenen (2010), eighteen management practices are grouped into three categories; monitoring, targets and incentives. “Monitoring” refers to “how companies monitor what goes on inside their firms and use this for continuous improvement”. “Targets” refers to “how companies set the right targets, track the right outcomes, and take appropriate action if the two are inconsistent”. “Incentives” refers to “how companies promote and reward employees based on performance, and whether or not companies try to hire and keep their best employees”. This paper focuses on the first type and a part of the third type of management practices defined in Bloom and Van Reenen (2010).

I pick up seven management practices and argue that they are closely related to the concept of MT used in this paper. Among the seven practices I pick up, the first four items are “performance tracking”, “performance review”, “performance dialogue” and “performance clarity”, which are related to whether or not firms can successfully find and catch misbehaving employees. The other three items are “consequence management”, “rewarding high performance”, “remove poor performers”, which are related to whether or not the firm can credibly punish (and reward) shirking employees (and hard working employees). I calculate the average score on these seven items and treat it as the measure for the quality of MT. The average score on these seven items is defined as $moinc$.

Next, I show that the quality of MT is highly correlated with firm performance by presenting simple scatter plots in Figure 1.7. As it is evident in the figure, the quality of MT is positively associated with firm performance such as sales per employee or total employment.
Third, I implement a variance and covariance decomposition exercise and show that there is substantial heterogeneity for scores on monitoring and incentives across firms. Note that the average management score is the sum of two parts or

\[ ms_i = moinc_i + nonmoinc_i = \frac{1}{18} \sum_{j \in moinc} Score_{ij} + \frac{1}{18} \sum_{j \in nonmoinc} Score_{ij}, \]

where \( ms_i \) is the average management score for firm \( i \), and \( j \) is the \( j \)-th management practice. The set \( moinc \) (or \( nonmoinc \)) is the set of management practices that are (or are not) related to monitoring and incentives. Next, I decompose the variation in
the average management score into the following three parts:

\[
\frac{1}{n} \Sigma_i (ms_i - \overline{ms})^2 = \frac{1}{n} \left[ \Sigma_i (moinc_i - \overline{moinc})^2 + \Sigma_i (nonmoinc_i - \overline{nonmoinc})^2 + 2\Sigma_i (moinc_i - \overline{moinc})(nonmoinc_i - \overline{nonmoinc}) \right],
\]

where \( \overline{ms} = \frac{1}{n} \Sigma_i ms_i \) and \( n \) is the number of firms in the data set. The first and the second terms above are the variations in management scores coming from \( moinc \) and \( nonmoinc \) respectively. The last term reflects the correlation between these two scores across firms. Table 1.5 shows that there is substantial variation in the score of \( moinc \) across firms.

**Table 1.5: Management Score Decomposition**

<table>
<thead>
<tr>
<th>Contributions</th>
<th>overall variation</th>
<th>var(moinc)</th>
<th>var(nonmoinc)</th>
<th>cross term</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.442</td>
<td>0.071</td>
<td>0.187</td>
<td>0.184</td>
</tr>
</tbody>
</table>

\( moinc \): average score on monitoring and incentives.

\( nonmoinc \): average score on other management practices.

\( crossterm \): correlation between the above two scores across firms.

Finally, I show that quality of monitoring and incentives differs across countries and is associated with firm size distribution. First, Table 1.6 shows that scores on every management practice that is included in the set of \( moinc \) differ significantly between China and the U.S. Second, Figure 1.8 shows that the distribution of scores on \( moinc \) differs substantially between India and the U.S. \[38\] Average score on \( moinc \) is significantly higher for U.S. firms. Finally, average management score on \( moinc \) is positively associated with the average firm size in an economy as Figure 1.8 shows. In other words, U.S. firms that have good MT are larger than Indian firms on average, and there are much more small firms in India compared with the U.S.

\[38\] The upper graph uses data from the WMS. The lower graph comes from Hsieh and Klenow (2012).
Table 1.6: Difference in the Quality of Monitoring and Incentives between China and the U.S.

<table>
<thead>
<tr>
<th></th>
<th>Perf2</th>
<th>Perf3</th>
<th>Perf4</th>
<th>Perf10</th>
<th>Perf5</th>
<th>talent2</th>
<th>talent3</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S.</td>
<td>3.65</td>
<td>3.63</td>
<td>3.58</td>
<td>2.93</td>
<td>3.67</td>
<td>3.17</td>
<td>3.82</td>
</tr>
<tr>
<td>China</td>
<td>2.99</td>
<td>3.09</td>
<td>2.74</td>
<td>2.77</td>
<td>2.65</td>
<td>2.88</td>
<td>3.04</td>
</tr>
</tbody>
</table>

Range for scores: 1 − 5.
All differences in means are statistically significant at 1% level.

Figure 1.8: Management Quality and the Firm Size Distribution
In total, evidence presented in this subsection explains why cross-country differences in the management quality is a natural candidate to explain cross-country differences in aggregate variables such as the size distribution of firms.

### 1.7.2 Continuous Effort Choice

In this subsection, I setup up the model under the assumption that workers' effort choice is a continuous variable. The goal of this exercise is to show that the model under this alternative assumption is isomorphic to the model with a binary effort choice.

I discuss the wage determination under the alternative assumption for the effort choice. Now, non-production workers' effort choice \( a_i \) \((i \leq T - 1)\) is a continuous variable with an upper bound \( \bar{a} \). An increase in \( a_i \) raises the probability of catching and firing a misbehaving worker even when the span of control is unchanged. Under this specification, the worker’s incentive compatibility constraint is

\[
w_i - \psi(a_i) \geq (1 - p(b, x_{i-1}, a_{i-1}))w_i
\]

(1.34)

where \( p(b, x_i, a_{i-1}) = \frac{a_{i-1}}{b \cdot x_{i-1}} = \frac{a_{i-1}m_{i-1}}{m_i} \). Thus, the incentive compatible wage is

\[
w_i = \frac{b \psi(a_i)m_i}{a_{i-1}m_{i-1}}.
\]

(1.35)

For reasons of tractability, I choose the following functional form for the cost function of exerting effort:

\[
\psi(a) = \frac{\bar{\psi}}{\bar{a} - a}.
\]

(1.36)

The key feature of the above functional form is that disutility increases with the effort level at an increasing speed (i.e., a convex function).
I show why a firm wants to incentivize workers at a given layer to choose the same level of effort and allocates the monitoring intensity evenly across them. The basic logic is the same as in the discrete case. The firm can always reduce the wage payment and induce the same total amount of effort inputs by equalizing the effort level and the monitoring intensity across workers at a given layer. Lemma 2 summarizes the results.

**Lemma 2** The firm incentivizes workers at a given layer to choose the same level of effort and equalizes the monitoring intensity across them.

Proof. I prove this lemma in three steps. For any two units of effort inputs at layer $i$ and the corresponding monitoring intensities for them, (i.e., $a_{i1}$, $p_{i1}$) and ($a_{i2}$, $p_{i2}$), there are four possibilities of relationship between them in total.\(^\text{39}\) Namely, $a_{i1} = a_{i2}$ and $p_{i1} = p_{i2}$; $a_{i1} = a_{i2}$ and $p_{i1} \neq p_{i2}$; $a_{i1} \neq a_{i2}$ and $p_{i1} = p_{i2}$; $a_{i1} \neq a_{i2}$ and $p_{i1} \neq p_{i2}$.

I prove that the last three possibilities are not optimal in what follows.

First, for the case in which $a_{i1} = a_{i2}$ and $p_{i1} \neq p_{i2}$, I can use exactly the same approach used in the discrete case to prove that it is not optimal. Second, I discuss the case in which $a_{i1} \neq a_{i2}$ and $p_{i1} = p_{i2}$. Total wage payment that is used to induce effort levels $a_{i1}$ and $a_{i2}$ can be reduced, if the effort levels are equalized. Formally, it must be true that

$$b\frac{\psi\left(\frac{a_{i1} + a_{i2}}{2}\right)}{p_{i1}} + b\frac{\psi\left(\frac{a_{i1} + a_{i2}}{2}\right)}{p_{i2}} < b\frac{\psi(a_{i1})}{p_{i1}} + b\frac{\psi(a_{i2})}{p_{i2}},$$

as $\psi(a)$ is a convex function. Finally, for the case in which $a_{i1} \neq a_{i2}$ and $p_{i1} \neq p_{i2}$, let me make the following simplifying notations:

$$a_0 \equiv a_{i1} + a_{i2}$$

\(^{39}\)The monitoring intensity is defined as $p_i = \frac{a_{i-1}m_{i-1}}{m_i}$. 
and

\[ p_0 \equiv p_{i1} + p_{i2}. \]

Cost minimization requires that allocation of two units of effort inputs and the monitoring intensities given \((a_0, p_0)\) are optimal at \((a_{i1}, p_{i1})\) and \((a_{i2}, p_{i2})\). Thus, we must have

\[ \frac{\psi'(a_{i1})}{p_{i1}} = \frac{\psi'(a_{i2})}{p_{i2}} \]

and

\[ \frac{\psi(a_{i1})}{p_{i1}^2} = \frac{\psi(a_{i2})}{p_{i2}^2} \]

due to the FOCs with respect to \(a_{i1}\) and \(p_{i1}\). The above two equations lead to the result that

\[ a_{i1} = a_{i2} = \frac{a_0}{2} \]

and

\[ p_{i1} = p_{i2} = \frac{p_0}{2} \]

which contradicts that \(a_{i1} \neq a_{i2}\) and \(p_{i1} \neq p_{i2}\). Therefore, the cost structure that is consistent with the firm’s cost minimization is the case in which \(a_{i1} = a_{i2}\) and \(p_{i1} = p_{i2}\) for any two units of effort inputs. In other words, the firm incentivizes workers at a given layer to choose the same level of efforts and equalizes the monitoring intensities on them. QED.
I can characterize the firm’s optimization problem now. Based on equation (1.35) and Lemma 2, the firm’s optimization problem can be written as

$$\max_{\{a_i, m_i\}_{i=1,\ldots,T}} \ A(\theta)^{\frac{1}{2}} q^{\frac{s-1}{s}} - \sum_{i=1}^{T} b \psi(a_i) \frac{m_i^2}{a_{i-1}m_{i-1}}$$

(1.37)

s.t. $$x_i = \frac{m_i}{m_{i-1}},$$

$$x_0 = 1; \ a_0 = 1; \ m_T = N = \frac{q}{a_T}; \ m_0 = 1;$$

$$0 \leq a_i \leq \bar{a}.$$ 

As in the discrete case, I minimize the firm’s variable cost given an output level first and then solve for the optimal output level as well as the optimal number of layers.

First, the variable cost for a firm that produces $q$ units of output and has $T + 1$ layers is

$$\min_{\{a_i, m_i\}_{i=1,\ldots,T}} \sum_{i=1}^{T} b \psi(a_i) \frac{m_i^2}{a_{i-1}m_{i-1}},$$

(1.38)

where $m_T = q/a_T$. The FOC of equation (1.38) with respect to $m_i$ is

$$\frac{2m_i \psi'(a_i)}{a_{i-1}m_{i-1}} - \frac{\psi(a_{i+1})m_{i+1}^2}{a_im_i^2} = 0.$$ 

(1.39)

The FOC of equation (1.38) with respect to $a_i$ is

$$\frac{m_i^2 \psi'(a_i)}{a_{i-1}m_{i-1}} - \frac{\psi(a_{i+1})m_{i+1}^2}{a_i^2m_i} = 0.$$ 

(1.40)

Note that the second order conditions hold for both $a_i$ and $m_i$. By comparing equation (1.39) with equation (1.40), I have

$$\frac{2\psi(a_i)}{m_i \psi'(a_i)} = \frac{a_i}{m_i}.$$ 

62
or
\[
\frac{\psi'(a_i)a_i}{\psi(a_i)} = 2.
\]

Due to the functional form of \(\psi(a)\) given in equation (1.36), the optimal effort level \(a_i (i < T)\) is
\[
a_i^* = \frac{2\bar{a}}{3}.
\]  
(1.41)

Without loss of generality, I normalize \(\bar{a}\) to \(\frac{3}{2}\) which leads to
\[
a_i^* = 1
\]
for all \(i < T\). Thus, the total wage payment given \(q\) and \(T\) is
\[
\min_{a_T, \{m_i\}_{i=1}^T} \sum_{i=1}^{T-1} b\psi(1) \frac{m_i^2}{m_{i-1}} + b\psi(a_T) \left(\frac{q}{a_T}\right)^2, \quad (1.42)
\]
from which I solve for optimal \(a_T\). Using the FOC of equation (1.42) with respect to \(a_T\), I derive that
\[
\frac{\psi'(a_T)a_T}{\psi(a_T)} = 2 \quad (1.43)
\]
and
\[
a_T^* = \frac{2\bar{a}}{3} = 1.
\]

In total, the optimal effort choice is one for all workers, which is the same as in the discrete case\(^{40}\)

Second, the FOCs of equation (1.38) with respect to various \(m_i\)'s can be rewritten as
\[
w_T m_T = 2w_{T-1} m_{T-1} = ... = 2^{T-1} w_1 m_1,
\]
\(^{40}\)The key equations here are equations (1.40) and (1.43). The key feature of them is that the optimal effort choice is completely determined by the cost function to exert effort. Thus, neither firm-level characteristics nor market-level variables are needed when we derive it.
which are the same as the ones derived in the discrete case. Therefore, I obtain the same solution for optimal employment at layer $i$ as in the discrete case or

$$m_i(N, T) = 2^i \left( \frac{N}{2^T} \right)^{\frac{T - 2^T - i}{2^T - 1}}. \quad (1.44)$$

Third, I derive optimal output and operating profit for firms with the quality draw $\theta$ from equation $(1.37)$. The solutions are

$$q(\theta, T) = N(\theta, T) = \left[ \frac{A\beta \theta^{\frac{1}{2}}}{b\psi 2^{2 - \frac{T}{2^T - 1}}} \right]^{\sigma(2^T - 1)} \quad (1.45)$$

and

$$\pi(\theta, T) = (1 - \frac{\beta(2^T - 1)}{2^T})(A\theta^{\frac{1}{2}})^{\frac{T}{2^T - 1}} \left( \frac{\beta/b}{2^{\frac{T + 1}{2^T - 1} - 1}} \right)^{\frac{(\sigma - 1)(2^T - 1)}{\sigma + (2^T - 1)}}, \quad (1.46)$$

These two solutions are exactly the same as the ones derived in the discrete case.

Finally, the firm chooses the number of layers optimally. The proofs and results in the discrete case all apply here, since the cost structure derived in the continuous case is exactly the same as the one derived in the discrete case. Furthermore, all empirical predictions and propositions derived in the discrete case hold in the continuous case as well.

### 1.7.3 Proof of Lemma 1

Proof. First, suppose that there is a worker at layer $i$ who shirks in equilibrium (i.e., $a_i = 0$). If he is a production worker, removing him from the hierarchy does not affect the firm’s output and (weakly) reduces labor cost. If he is a non-production worker, his direct subordinates at layer $i + 1$ would shirk as a result of the absence of monitoring from above. Furthermore, all his direct and indirect subordinates would
shirk as well. Similar as before, removing them from the hierarchy does not affect
the firm’s output and (weakly) reduces the labor costs, which means excluding them
from the hierarchy is always optimal. Thus, all workers are incentivized to work
in equilibrium. Second, the reason why the firm wants to allocate the monitoring
intensities evenly across workers is that it could reduce wage payments by doing so,
if the monitoring intensities were not equalized. More specifically, suppose there are
two units of effort inputs that are monitored under different monitoring intensities \( p_1 \)
and \( p_2 \). As all workers are incentivized to work, the wage payment to these two units
equals
\[
b\psi \left( \frac{1}{p_1} + \frac{1}{p_2} \right),
\]
However, the firm can reduce this wage payment by equalizing the monitoring intensities across these two units of effort inputs as
\[
2b\psi \frac{1}{(p_1 + p_2)/2} < b\psi \left( \frac{1}{p_1} + \frac{1}{p_2} \right)
\]
for any \( p_1 \neq p_2 \). This means that the firm can elicit the two units of effort inputs
under a lower cost. Therefore, the firm’s optimal choice is to equalize the monitoring intensities across workers at a given layers. QED.

### 1.7.4 Proof of Proposition 1

The AVC function and the MC function given the number of layers are
\[
AVC(q, T) = (2 - \frac{1}{2^{T-1}})b\psi 2^{1-\frac{T}{2^{T-1}}}q^{\frac{1}{2^{T-1}}} \tag{1.47}
\]
and
\[
MC(q, T) = b\psi 2^{2-\frac{T}{2^{T-1}}}q^{\frac{1}{2^{T-1}}} = \frac{2^T}{2^T-1} AVC(q, T). \tag{1.48}
\]
From the expression of these two cost functions, it is straightforward to see that both of them increase with output $q$ given the number of layers $T + 1$. Thus, the first part of the proposition has been proved.

Next, I discuss the overall shape of the AVC curve. Before the discussion, let me make the following notation for future use.

**Definition 1** Let $q_T$ be the solution to the following equation:

$$AVC(q_T, T) = AVC(q_T, T + 1).$$

*In other words, the AVC of using $T + 1$ layers is equal to the AVC of using $T + 2$ layers at output level $q_T$.***

Now, I prove the following lemma which assures the monotonicity of $q_T$.

**Lemma 3** $q_T$ increases in $T$.

Proof. I rewrite $AVC(q_T, T) = AVC(q_T, T + 1)$ as

$$\frac{(2 - \frac{1}{2^T})/2^{2^T-1}}{(2 - \frac{1}{2^{(T+1)-1}})/2^{2^{(T+1)-1}}} q_T^{1-1/2(T+1)-1} = 1.$$ 

Thus, the switching point $q_T$ can be rewritten as

$$q_T = \left[\frac{2^{T+1} - 1}{2^{T+1} - 2}\right]^{(2^{T-1})(2^{T+1-1})} 2^{(T-1)+\frac{T}{2T}} \equiv \Psi_1(T) \Psi_2(T).$$

Taking logs and calculating the first order derivative with respect to $T$ yields the following result:

$$\frac{d[\ln(\Psi_1(T)) + \ln(\Psi_2(T))]}{dT} = \ln 2(2^{T+1} - 2^{-T}) \ln \left(\frac{2^{T+1} - 1}{2^{T+1} - 2}\right) - \ln 2 + \ln 2 \left[1 - \ln \frac{2}{2^T}\right].$$
Thus, the sign of the above expression depends on

\[
Sign\left( \frac{d[\ln(\Psi_1(T)) + \ln(\Psi_2(T))]}{dT} \right) = Sign\left( (2^{2T+1} - 1) \ln\left( \frac{2^{T+1} - 1}{2^T - 1} \right) - \ln 2 \right)
\]

or

\[
Sign\left( \frac{d[\ln(\Psi_1(T)) + \ln(\Psi_2(T))]}{dT} \right) = Sign\left( (2^{2T+1} - 1) \ln\left( \frac{2^{T+1} - 1}{2^T - 1} \right) - 2^{2T+1} \ln 2 \right)
\]

or

\[
Sign\left( \frac{d[\ln(\Psi_1(T)) + \ln(\Psi_2(T))]}{dT} \right) = Sign\left( (1 - 2^{-(2T+1)}) \ln\left( \frac{2^{T+1} - 1}{2^T - 1} \right) - \ln 2 \right).
\]

I want to show that \((1 - 2^{-(2T+1)}) \ln\left( \frac{2^{T+1} - 1}{2^T - 1} \right) - \ln 2\) decreases in \(T\) for \(T \geq 1\). First, I have

\[
\frac{d}{dT} \left[ (1 - 2^{-(2T+1)}) \ln\left( \frac{2^{T+1} - 1}{2^T - 1} \right) - \ln 2 \right] = \ln 2 \left[ \ln\left( \frac{2^{T+1} - 1}{2^T - 1} \right) \frac{1}{2^T} - \left(1 - \frac{1}{2^{2T+1}}\right) \frac{1}{2^T - 1} - 3 + \frac{1}{2^T} \right] \\
\equiv \ln 2 (K_1(T) - K_2(T)).
\]

Second, I prove that

\[K_1(T) - K_2(T) < 0\]

for all \(T \geq 1\). I proceed in two steps. In the first step, calculation shows that \(K_1(1) - K_2(1) < 0\). In the second step, it is straightforward to see that for any \(T > 1\)

\[K_1(T) < \frac{1}{2^{2(T-1)}} K_1(1)\]

and

\[K_2(T) > \frac{1}{2^T - 1} K_2(1)\].
Thus, I have

\[ K_1(T) - K_2(T) < K_1(1) - K_2(1) < 0 \]

for all \( T > 1 \). Finally, due to the monotonicity of \((1 - 2^{-2(T+1)}) \ln(\frac{2^{T+1} - 1}{2^T - 1}) - \ln 2\) with respect to \( T \) that has just been proven, I conclude that

\[(1 - 2^{-2(T+1)}) \ln(\frac{2^{T+1} - 1}{2^T - 1}) - \ln 2 > \lim_{T \to \infty} (1 - 2^{-2(T+1)}) \ln(\frac{2^{T+1} - 1}{2^T - 1}) - \ln 2 = 0.\]

and

\[ \text{Sign} \left( \frac{d[\ln(\Psi_1(T)) + \ln(\Psi_2(T))]}{dT} \right) = \text{Sign} \left( (1 - 2^{-2(T+1)}) \ln(\frac{2^{T+1} - 1}{2^T - 1}) - \ln 2 \right) > 0. \]

Therefore, \( q_T \) must be an increasing function of \( T \). QED.

Having established the monotonicity of \( q_T \) in Lemma 3, I can characterize the overall shape of the AVC curve.

**Lemma 4** If the output produced in equilibrium \( q \in [q_{T-1}, q_T) \), the optimal number of hierarchical layers is \( T + 1 \). At the output level \( q_T \), the AVC curve kinks and its slope decreases discontinuously as the firms adds a layer. Finally, the switching point \( q_T \) does not depend on the firm’s quality draw (i.e., \( \theta \)), the inefficiency of MT (i.e., \( b \)) and the adjusted market size (i.e., \( A \)).

Proof. I proceed the proof in the following several steps. First, note that at \( q_{T-1} \), the slope of \( AVC(q, T) \) is smaller than the slope of \( AVC(q, T - 1) \) as \( AVC(q_{T-1}, T) = AVC(q_{T-1}, T - 1) \). This prove the second part of this lemma. Second, due to this property, \( AVC(q, T - 1) \) is below \( AVC(q, T) \) for \( q < q_{T-1} \) and above \( AVC(q, T) \) for \( q > q_{T-1} \). Thus, \( T + 1 \) layers is never chosen for \( q < q_{T-1} \). Similarly, \( T + 1 \) layers is never chosen for \( q > q_T \) as \( AVC(q, T) \) is above \( AVC(q, T + 1) \) for \( q > q_T \). Third, as \( AVC(q, T) \) is below \( AVC(q, T - 1) \) for \( q > q_{T-1} \) and \( q_T \) increases in \( T \), \( AVC(q, T) \) is
below \( AVC(q, t) \) for all \( t < T \) when \( q > q_{T-1} \). Similarly, as \( AVC(q, T + 1) \) is above \( AVC(q, T) \) for \( q < q_T \) and \( q_T \) increases in \( T \), \( AVC(q, t) \) is above \( AVC(q, T) \) for all \( t > T \) when \( q < q_T \). In total, \( AVC(q, T) \) is below \( AVC(q, t) \) for all \( t \neq T \) when \( q \in (q_{T-1}, q_T) \) which leads to the result that for \( q \in (q_{T-1}, q_T) \), the optimal choice of layers is \( T + 1 \). Of course, when \( q = q_{T-1} \), choosing either \( T \) layers or \( T + 1 \) layers is optimal. Finally, the third half of the above lemma follows from the expression of \( q_T \) directly. QED.

I prove the following claim that characterizes the overall shape of the MC curve.

**Claim 1** Given the number of layers \( T + 1 \), the MC increases with output. The final \( MC \) curve is

\[
MC(q) = MC(q, T)
\]

where \( q \in [q_{T-1}, q_T) \). This cost increases in interval \([q_{T-1}, q_T)\) for all \( T \) and decreases discontinuously at the point \( q_T \).

Proof. It is straightforward to see the first part of this proposition due to Lemma 4. The only thing that needs proof is the last part. First, it is straightforward to see that \( MC(q, T) \) increases in \( q \) for a given \( T \). Second, at \( q_T \), I have

\[
AVC(q_T, T) = AVC(q_T, T + 1).
\]

As

\[
MC(q, T) = \frac{2^T}{2^T - 1} AVC(q, T),
\]

it must be true that

\[
MC(q_T, T) > MC(q_T, T + 1).
\]

The fall in the marginal cost when the firm adds a layer comes from the reorganization inside the firm. QED.

In sum, I proved Proposition due to Lemma and Claim QED.
1.7.5 Proof of Proposition 2

The first part of this proposition is true because of the shape of the AVC curve shown in Proposition 1. I prove the second of this proposition in five steps. First, I define two demand thresholds for a given number of layers $T + 1$ for future use.

**Definition 2** For the number of layers $T + 1$, $\theta_{T1}$ is defined as the solution to

$$MR(\theta_{T1}, q_T) = A\beta\theta_{T1}^{\frac{1}{T+1}}q_T^{\frac{1}{T+1}} = MC(q_T, T + 1) = b\psi2^{2-\frac{T+1}{T+1-1}}q_T^{\frac{1}{T+1-1}}.$$  

In other words, firms with the quality draw $\theta_{T1}$ have their marginal revenue (MR) curve intersect the MC curve of using $T + 2$ layers at output level $q_T$. $\theta_{T3}(> \theta_{T1})$ is defined as the solution to

$$MR(\theta_{T3}, q_T) = A\beta\theta_{T3}^{\frac{1}{T+1}}q_T^{\frac{1}{T+1}} = MC(q_T, T) = b\psi2^{2-\frac{T}{T-1}}q_T^{\frac{1}{T-1}}.$$  

In other words, firms with the quality draw $\theta_{T3}$ have their MR curve intersect the MC curve of $T + 1$ layers at output level $q_T$.

The graphical representation of $\theta_{T1}$ and $\theta_{T3}$ is in Figure 1.9

Second, I show that only when the firm’s quality draw is between $[\theta_{T1}, \theta_{T3}]$, does it have incentive to switch from $T + 1$ layers to $T + 2$ layers in the following lemma.

**Lemma 5** For each $T$, firms having the quality draw smaller than or equal to $\theta_{T1}$ prefer $T + 1$ layers over $T + 2$ layers, while firms having the quality draw higher than or equal to $\theta_{T3}$ prefer $T + 2$ layers over $T + 1$ layers.

Proof. First, note that as $MC(q_T, T) > MC(q_T, T + 1)$ and $MR(\theta, q)$ is an increasing function of $\theta$ for a given $q$, it must be true that $\theta_{T1} < \theta_{T3}$.

Next, if a firm with $\theta < \theta_{T1}$ chose $T + 2$ layers, it must be true that $q(\theta, T + 1) < q_T$ which is not optimal for the firm as $AVC(q, T) < AVC(q, T + 1)$ for output levels
smaller than $q_T$. Thus, Firms with $\theta < \theta_{T1}$ prefer $T + 1$ layers over $T + 2$ layers.

Similarly, if a firm with $\theta > \theta_{T3}$ chose $T + 1$ layers, it must be true that $q(\theta, T) > q_T$
which contradicts that $AVC(q, T) > AVC(q, T + 1)$ for output levels bigger than $q_T$.

Thus, Firms with $\theta > \theta_{T3}$ prefer $T + 2$ layers over $T + 1$ layers.

Finally, when $\theta = \theta_{T1}$, choosing $T + 1$ layers yields more profit as

$$\pi(\theta_{T1}, T) \equiv \pi(\theta_{T1}, T, q(\theta_{T1}, T)) > \pi(\theta_{T1}, T, q_T) = \pi(\theta_{T1}, T + 1, q_T),$$

where I have used the result that $AVC(q_T, T) = AVC(q_T, T + 1)$. Similarly, when $\theta = \theta_{T3}$, choosing $T + 2$ layers yields more profit as

$$\pi(\theta_{T3}, T + 1) \equiv \pi(\theta_{T3}, T + 1, q(\theta_{T3}, T + 1)) > \pi(\theta_{T3}, T + 1, q_T) = \pi(\theta_{T3}, T, q_T).$$

QED.

Third, I use the following lemma to show the complementarity between the benefit of adding a layer and the quality draw $\theta$. 71
Lemma 6 For a given $T$, $\pi(\theta, T + 1) - \pi(\theta, T)$ increases continuously in $\theta$ for $\theta \in [\theta_{T1}, \theta_{T3}]$.

Proof. I use Figure 1.10 to prove this lemma. For any $\theta \in [\theta_{T1}, \theta_{T3}]$, suppose the quality draw $\theta$ increases by $\Delta > 0$ which corresponds to a shift of the MR curve from the red one to the green one. The difference between $\pi(\theta, T)$ and $\pi(\theta + \Delta, T)$ is represented by the red region, while the difference between $\pi(\theta, T + 1)$ and $\pi(\theta + \Delta, T + 1)$ is represented by the sum of the red region and the blue region. Thus, I have

$$\pi(\theta + \Delta, T + 1) - \pi(\theta + \Delta, T) - [\pi(\theta, T + 1) - \pi(\theta, T)] = [\pi(\theta + \Delta, T + 1) - \pi(\theta, T + 1)] - [\pi(\theta + \Delta, T) - \pi(\theta, T)]$$

which is the blue region. As the MR curve moves upward when $\theta$ increases and the MC curve of $T + 2$ layers lies below the MC curve of $T + 1$ layers when $q \geq q_{T1}$, the
area of the blue region increases as $\Delta$ increases. Thus, it must be true that

$$\pi(\theta + \Delta, T + 1) - \pi(\theta + \Delta, T) - [\pi(\theta, T + 1) - \pi(\theta, T)]$$

increases in $\Delta$ which means that $\pi(\theta, T + 1) - \pi(\theta, T)$ increases in $\theta$ for $\theta \in [\theta_{T1}, \theta_{T3}]$. The continuity of $\pi(\theta, T + 1) - \pi(\theta, T)$ in $\theta$ is straightforward to see. QED.

Fourth, I prove the following result which is the key step to prove this proposition. More specifically, there exists a threshold $\theta_{T2} \in (\theta_{T1}, \theta_{T3})$ such that firms with this level of efficiency is indifferent between having $T + 1$ layers and having $T + 2$ layers. Claim 3 summarizes the results.

**Claim 2** For each $T$, there exists a threshold $\theta_{T2} \in (\theta_{T1}, \theta_{T3})$ such that firms with this demand draw is indifferent between having $T + 1$ layers and having $T + 2$ layers. Moreover, firms with a level of the demand draw smaller than $\theta_{T2}$ strictly prefer $T + 1$ layers over $T + 2$ layers, while firms with a level of the demand draw bigger than $\theta_{T2}$ strictly prefer $T + 2$ layers over $T + 1$ layers.

**Proof.** From Lemma 5 I have

$$\pi(\theta_{T1}, T) > \pi(\theta_{T1}, T + 1),$$

and

$$\pi(\theta_{T3}, T) < \pi(\theta_{T3}, T + 1).$$

As $\pi(\theta, T + 1) - \pi(\theta, T)$ continuously increases in $\theta$ for $\theta \in [\theta_{T1}, \theta_{T3}]$ due to Lemma 6, there must exist a threshold $\theta_{T2} \in (\theta_{T1}, \theta_{T3})$ such that

$$\pi(\theta_{T2}, T) = \pi(\theta_{T2}, T + 1).$$
And for all $\theta < \theta_{T_2}$
\[ \pi(\theta, T) > \pi(\theta, T + 1), \]
while for all $\theta > \theta_{T_2}$
\[ \pi(\theta, T) < \pi(\theta, T + 1). \]

QED.

Now, I can prove this proposition by generalizing Claim 3 into the case of any two different values of the number of layers. First, I define the upper bound and the lower bound on the quality draw for the firm’s changing the number of layers from $T_0$ to $T_1 (> T_0)$, where $T_0$ and $T_1$ can be any positive numbers. The following definition is used for this purpose.

**Definition 3** For the numbers of layers $T_0$ and $T_1 (> T_0)$, $\theta_{T_0,T_1}$ is defined as the solution to
\[ MR(\theta_{T_0,T_1}, q_{T_0,T_1}) = MC(q_{T_0,T_1}, T_1), \]
where $q_{T_0,T_1}$ is the output level at which $AVC(q_{T_0,T_1}, T_0) = AVC(q_{T_0,T_1}, T_1)$. $\theta_{T_1,T_1}(> \theta_{T_0,T_1})$ is defined as the solution to
\[ MR(\theta_{T_1,T_1}, q_{T_0,T_1}) = MC(q_{T_0,T_1}, T_0). \]

Second, using the same approach used in the proof of Claim 3, one can prove that there exists a quality cutoff $\theta_{2T_0,T_1} \in (\theta_{T_0,T_1}, \theta_{T_1,T_1})$ such that firms with quality draws bigger than $\theta_{2T_0,T_1}$ prefer $T_1 + 1$ layers over $T_0 + 1$ layers can vice versa. Third, suppose there are two firms with quality draws $\theta_1$ and $\theta_0 > (\theta_1)$ such that the firm with quality draw $\theta_0$ has fewer layers than the firm with quality draw $\theta_1$. I use $T_1 + 1$ and $T_0 + 1(< T_1 + 1)$ to denote the number of layers for firms with quality draws $\theta_1$ and $\theta_0$ respectively. Form the above discussion, it is straightforward to see that this
supposition can’t be true, as firms with quality draws bigger than $\theta_{T_0,x_1}$ prefer $T_1 + 1$ layers over $T_0 + 1$ layers and vice versa. Therefore, firms with better demand draws have more layers. QED.

Thanks to this proposition, I only need to derive the sequence of $\theta_{T_2}$ for $T = 1, 2, 3...$ when solving the optimal number of layers for each firm. In other words, there is no need to solve the optimal number of layers for each firm respectively. Simulations become much less time-consuming because of this result.

1.7.6 Proof of Proposition 3

First, when the firm expands without changing the number of layers, both employment and output increase continuously due to equations (1.12) and (1.14). Second, the firm’s optimal pricing rule implies that

$$p(\theta) = \frac{\sigma}{\sigma - 1} MC(q(\theta)).$$

As the firm’s MC increases given the number of layers and decreases discontinuously when the firm adds a layer due to a marginal increase in $\theta$, the firm’s price follows the same pattern. Finally, the firm’s optimal output is

$$q(\theta) = \theta A^\sigma \left( \frac{\sigma}{\sigma - 1} MC(q(\theta)) \right)^{-\sigma}.$$

When the firm adds a layer due to a marginal increase in $\theta$, the output increases discontinuously as the price falls discontinuously. The span of control increases at all existing layers as well which will be shown later. Therefore, employment jumps up discontinuously due to both the decreasing span of control at existing layers and the employed workers at the new layer. QED.
1.7.7 Proof of Proposition 4

First, equation (1.13) implies that the span of control increases at all layers when \( \theta \) increases and its number of layers is unchanged. Second, as the wage defined in equation (1.8) is positively affected the span of control, wages increase at all layers. Third, the FOCs with respect to employment in equation (1.11) show that

\[
\frac{w_i(q(\theta, T(\theta)), T(\theta))}{w_{i+1}(q(\theta, T(\theta)), T(\theta))} = \frac{1}{2} \frac{m_{i+1}(q(\theta, T(\theta)), T(\theta))}{m_i(q(\theta, T(\theta)), T(\theta))} = \frac{1}{2} x_i(q(\theta, T(\theta)), T(\theta))
\]

for \( T(\theta) > i \geq 1 \). As the span of control increases at all layers, relative wages increase at all layers as well.

Finally, I prove the employment hierarchy that the number of workers is smaller in upper layers.\(^{41}\) As I consider the employment hierarchy for workers, the minimum value for \( T \) is two. Equation (1.12) shows

\[
\frac{m_{i+1}(q(\theta, T(\theta)), T(\theta))}{m_i(q(\theta, T(\theta)), T(\theta))} = 2 \left[ \frac{q(\theta, T(\theta))}{2^{T(\theta)}} \right]^{2^{T(\theta)-1}} \frac{2^{T(\theta)-1}}{2^{T(\theta)-1}} \geq \left[ \frac{q(\theta, T(\theta))}{2^{T(\theta)-1}} \right]^{2^{T(\theta)-1}-1},
\]

as \( T(\theta) \geq 2 \) and \( T(\theta) > i \geq 1 \). Now, I show the following property of \( q_{T-1} \) that is the key step to prove the result of the employment hierarchy.\(^{42}\)

\[
\frac{q_{T-1}}{2^{T-1}} = \left[ \frac{2^T - 1}{2^T - 2} \right]^{ \frac{2^{T-1}-2}{2^{T-1}} - 1 } > 1.
\]

This is because

\[
\left[ \frac{2^T - 1}{2^T - 2} \right]^{ \frac{2^{T-1}-2}{2^{T-1}} - 1 } > 1,
\]

\(^{41}\)This result will be used later.

\(^{42}\)\( q_T \) is defined in Definition.
increases in $T$ for $T \geq 2$ and achieves its minimum value of 1.299 when $T = 2$. In total,

$$\frac{m_{i+1}(q(\theta, T(\theta)), T(\theta))}{m_i(q(\theta, T(\theta)), T(\theta))} \geq \frac{q(\theta, T(\theta))^{\frac{T(\theta)-(i+1)}{2T(\theta)-1}}}{\frac{qT(\theta)-1}{2T(\theta)-1}} > 1.$$ 

Therefore, the employment hierarchy holds for workers. QED.

### 1.7.8 Proof of Proposition 5

First of all, keep in mind that I am considering a small change in $\theta$ from $\theta_{T_0,2} - \Delta$ to $\theta_{T_0,2} + \Delta$ that triggers the addition of one layer into the hierarchy. Note that $\theta_{T_0,2}$ is the demand threshold where the firm switches from having $T_0 + 1$ layers to having $T_0 + 2$ layers.

As the change in the span of control is the key to prove this proposition, I prove that the span of control falls at all existing layers first. From equations (1.12) and (1.14), I have

$$|\frac{m_i^*}{m_{i-1}^*}|_{T_0} = 2 \left[ \frac{\beta A (\theta_{T_0,2} - \Delta)^{\frac{1}{2}}}{4b\psi 2^{\frac{T_0}{\sigma}}} \right]^{\sigma 2^{T_0-i}}_{\sigma (2^{T_0-i})}.$$

and

$$|\frac{m_i^*}{m_{i+1}^*}|_{T_0+1} = 2 \left[ \frac{\beta A (\theta_{T_0,2} + \Delta)^{\frac{1}{2}}}{4b\psi 2^{\frac{T_0+1}{\sigma}}} \right]^{\sigma 2^{T_0+i}}_{\sigma (2^{T_0+i+1})},$$

where $\Delta$ is infinitesimally small. Thus, what I have to prove is that

$$Z(\theta_{T_0,2}, T_0) = \left[ \frac{\beta A \theta_{T_0,2}^{\frac{1}{2}}}{4b\psi 2^{\frac{T_0}{\sigma}}} \right]^{\sigma (2^{T_0-1})}_{\sigma (2^{T_0-1})},$$

decreases with $T_0$ at $\theta_{T_0,2}$. Calculation shows that

$$\text{Sign}\left[ \frac{dZ(\theta_{T_0,2}, T_0)/dT_0}{dT_0} \right] = \text{Sign}\left[ \ln 2 \left[ -2^{T_0} \left( \frac{\ln \left( \frac{\beta A \theta_{T_0,2}^{\frac{1}{2}}}{4b\psi} \right)}{2^{T_0} + (\sigma - 1)} \right) - \frac{1}{\sigma} \right] \right].$$
Obviously, if
\[
\frac{\beta A \theta T^2_{0, 2}}{4b\psi 2\frac{T}{\sigma}} \geq 1,
\]
then the proof is done. So, I only need to consider the case where
\[
\frac{\beta A \theta T^2_{0, 2}}{4b\psi 2\frac{T}{\sigma}} < 1.
\]

For this case, there is a lower bound on the above term due to the result that \(\theta T_0, 2 > \theta T_0, 1\). Thus, I only have to prove that
\[
-2^{T_0} \left( \ln \frac{\beta A \theta T^2_{0, 1}}{4b\psi} - \ln \frac{2}{\sigma} \right) - \frac{1}{\sigma} < 0.
\]

Based on Definition 2, \(\theta T_0, 1\) can be rewritten as
\[
MR(\theta T_0, 1, q_T 0) = A\beta \theta T^2_{0, 1} q_T^0 = MC(q_T 0, T 0 + 1) = b\psi 2^{T_0 + \frac{T_0 + 1}{2} - 1} q_T^0\frac{1}{2^{T_0 + 1} - 1}.
\]

Thus, I can solve \(\theta T_0, 1\) as
\[
\theta T_0, 1 = \frac{(b\psi 2^{T_0 + \frac{T_0 + 1}{2} - 1} q_T^0)^{\frac{1}{\sigma + (2^{T_0 + 1} - 1)}}}{(A\beta)^\sigma}.
\]

Consequently, I have
\[
2^{T_0} \left( \ln \frac{\beta A \theta T^2_{0, 1}}{4b\psi} - \ln \frac{2}{\sigma} \right) = \frac{2^{T_0}(\sigma + (2^{T_0 + 1} - 1))}{(\sigma + (2^{T_0 - 1}))(2^{T_0 + 1} - 1)} \ln \left( \frac{q_T 0}{2^{T_0}} \right) - \frac{2^{T_0}}{(\sigma + (2^{T_0 - 1}))(2^{T_0 + 1} - 1)} \ln 2.
\]
As \( \frac{q_{T_0}}{q_T} > 1 \) for \( T_0 \geq 1 \) due to the proof in Appendix 1.7.1, I conclude that

\[
-2^{T_0} \left( \frac{\ln \left( \frac{\beta A \theta_{T_0,1}}{4bw} \right) - \ln \left( \frac{2^{T_0}}{\sigma} \right)}{2^{T_0} + (\sigma - 1)} \right) - \frac{1}{\sigma} < \frac{2^{T_0}}{(\sigma + (2^{T_0} - 1))(2^{T_0+1} - 1)} \ln 2 - \frac{1}{\sigma} < 0
\]

for all \( T_0 \geq 1 \). In total, I conclude that

\[
-2^{T_0} \left( \frac{\ln \left( \frac{\beta A \theta_{T_0,2}}{4bw} \right) - \ln \left( \frac{2^{T_0}}{\sigma} \right)}{2^{T_0} + (\sigma - 1)} \right) - \frac{1}{\sigma} < 0
\]

for all \( \theta_{T_0,2} \). As \( \Delta \) is infinitesimally small, It must be true that

\[
\frac{m^*_i}{m^*_{i-1}} \bigg|_{T_0} > \frac{m^*_{i+1}}{m^*_i} \bigg|_{T_0+1}
\]

for all \( i \) and \( T_0 \geq 1 \). Therefore, the span of control must fall at all existing layers when the firm adds a layer.

Next, as the wage at layer \( i \) is

\[
w_i(\theta) = bw_i \frac{m_i(\theta, T)}{m_{i-1}(\theta, T)},
\]

wages fall at all existing layers when the firm adds a layer.

Third, as the relative wage is proportional to the span of control or

\[
\frac{w_{i-1}(\theta)}{w_i(\theta)} = \frac{m_i(\theta, T)}{2m_{i-1}(\theta, T)},
\]

relative wages also fall at all existing layers when the firm adds a layer.

Finally, total employment increases discontinuously when the firm adds layer, as output increases discontinuously, and the span of control fall at existing layers. QED.
1.7.9 Proof of Proposition 6

Let me write out the expression of unit costs given \( T + 1 \) layers as follows:

\[
UC_T(q, b) = (2 - \frac{1}{2^{T-1}})b\psi 2^{1 - \frac{T}{2^{T-1}}} q^{\frac{1}{2^{T-1}}} + \frac{f_0}{q}.
\]

First, taking the FOC of the above equation with respect to \( q \) results in

\[
\frac{\partial UC_T(q, b)}{\partial q} = \frac{1}{2^{T-1}}b\psi 2^{1 - \frac{T}{2^{T-1}}} q^{\frac{-(2^T-2)}{2^{T-1}}} - \frac{f_0}{q^2}.
\]

(1.49)

There exists a unique \( q_{Tm}(b) \) given \( T \) and \( b \) such that the above equation equals zero. Moreover, \( \frac{\partial UC_T(q, b)}{\partial q} > 0 \) if and only if \( q > q_{Tm}(b) \) and vice versa. Therefore, the curve of unit costs given \( T \) and \( b \) is “U” shaped, which implies that firm productivity given \( T \) and \( b \) has an inverted “U” shape.

Next, it is straightforward to observe that

\[
\lim_{q \to \infty} \frac{\partial UC_T(q, b)}{\partial q} = 0
\]

given \( T \) and \( b \). Therefore, the slope of unit costs approaches zero given \( T \) and \( b \) when output goes to infinity.

Third, the MES given \( T \) and \( b \) can be solved as follows:

\[
q_{Tm}(b) = \left[ \frac{f_0}{4b\psi} \right]^{\frac{T}{2^{T-1}}} 2^T.
\]

(1.50)

A sufficiency and necessary condition for \( \{q_{Tm}(b)\}_{T=1,2,3,...} \) to be an increasing sequence is that

\[
4f_0 > b\psi.
\]
Finally, I need to derive the condition under which \( \{MUC_T(b)\}_{T=1,2,3...} \) is a decreasing sequence, or

\[
AVC_{T+1}(q_{T+1,m}(b), b) + \frac{f_0}{q_{T+1,m}(b)} < AVC_T(q_{T,m}(b), b) + \frac{f_0}{q_{T,m}(b)} \quad \forall T \geq 1 \quad (1.51)
\]

One key observation is that equation (1.49) implies

\[
\frac{1}{2T - 1} \frac{AVC_T(q_{T,m}(b), b)}{q_{T,m}(b)} = \frac{f_0}{q_{T,m}(b)^2}. \quad (1.52)
\]

From equation (1.52), I conclude that

\[
AVC_T(q_{T,m}(b), b) + \frac{f_0}{q_{T,m}(b)} = (2^T - 1) \frac{f_0}{q_{T,m}(b)} + \frac{f_0}{q_{T,m}(b)} = 2^T \frac{f_0}{q_{T,m}(b)}. \quad (1.53)
\]

Substituting equation (1.53) into equation (1.51) leads to

\[MUC_T(b) > MUC_{T+1}(b),\]

if and only if

\[q_{T+1,m}(b) > 2q_{T,m}(b),\]

where \( T = 1, 2, 3... \) The expression of \( q_{T,m}(b) \) in equation (1.50) implies that

\[q_{T+1,m}(b) > 2q_{T,m}(b),\]

if and only if

\[f_0 > 4b\psi.\]

In total, if \( f_0 > 4b\psi, \) \( \{q_{T,m}(b)\}_{T=1,2,3...} \) is an increasing sequence, and \( \{MUC_T(b)\}_{T=1,2,3...} \) is a decreasing sequence. QED.
1.7.10 Proof of Proposition \[7\]

The strategy to prove this proposition is the following. First, I assume that the incentive compatible wage defined in equation (1.8) satisfies the constraint indicated in equation (1.30) in every labor submarket and prove that there is a unique equilibrium with unemployment in every labor submarket. Second, I show that there is a non-empty set of parameter values within which the incentive compatible wage defined in equation (1.7) satisfies the constraint indicated in equation (1.30) in every labor submarket.

First, I redefine the equilibrium using three conditions. Substituting equation (1.10) into equation (1.24) leads to the homogeneous sector’s employment expressed as

\[ L_h = \frac{(1 - \gamma)A^\sigma P^{1-\sigma}}{\gamma p_h}. \]  

(1.54)

Substituting the above equation and equation (1.28) into equation (1.28) yields the following labor market clearing condition:

\[ \frac{WP(\bar{\theta}, A, M) - \psi LD(\bar{\theta}, A, M)}{p_h} + \frac{(1 - \gamma)A^\sigma P^{1-\sigma}}{\gamma p_h} = L. \]  

(1.55)

Now, the equilibrium of the economy can be solved using three equations (i.e., equations (1.22), (1.23) and (1.55)). As a result, I obtain three endogenous variables: \( \theta, A \) and \( p_h \).

Values of other equilibrium variables can be solved using \( \theta, A \) and \( p_h \) derived above. First, the ideal price index is

\[ P = \frac{1}{p_h}. \]  

(1.56)
due to equation (1.3). Second, the ideal price index defined in equation (1.4) can be reexpressed as

\[ P = \left( \int_{\theta=\bar{\theta}}^{\infty} \theta p(\theta) \frac{1-\sigma}{1-G(\bar{\theta})} d\theta \right)^{\frac{1}{1-\sigma}} \equiv P_1(\bar{\theta}, A)M^{\frac{1}{1-\sigma}}. \]  

(1.57)

This is because prices charged by various firms in the CES sector only depend on \( A \) and \( \theta \). Thus, the mass of firms \( M \) can be derived by using equations (1.56) and (1.57) and values of \( \bar{\theta} \), \( A \) and \( p_h \). Third, the aggregate income \( E \) can be derived by using equation (1.10) and value of \( A \) and \( P \). Finally, the allocation of labor can be obtained by using equations (1.28) and (1.54) and value of \( A \), \( P \) and \( p_h \).

Now I show why I can use three variables (i.e., \( \bar{\theta} \), \( A \) and \( M \)) to derive both the aggregate wage payment and the number of employed workers in the CES sector. In equation (1.9), only \( A \) and \( \theta \) affect firm’s optimal choices given values of exogenous parameters \( b \) and \( \psi \). As firms endogenously choose whether or not to stay in the market, wage payment per active firm and employment per active firm are functions of \( (A, \bar{\theta}) \) only. Therefore, I can use three variables (i.e., \( A \), \( \theta \) and \( M \)) to derive both the aggregate wage payment and the number of employed workers in the CES sector.

Next, the following claim shows the existence and uniqueness of the equilibrium in the CES sector.

**Claim 3** There exists a unique equilibrium for the CES sector characterized by a unique pair of \((\bar{\theta}, A)\).

Proof. I have two equilibrium conditions: the ZCP condition and the FE condition. I have two endogenous variables to be pinned down: the exit cutoff \( \bar{\theta} \) and the adjusted market size \( A \). Let us think about the ZCP condition first. The goal is to establish a negative relationship between \( \bar{\theta} \) and \( A \) from this condition. Suppose \( A \) increases from \( A_0 \) to \( A_1(> A_0) \) in equation (1.22). If the exit cutoff \( \bar{\theta} \) increased from \( \bar{\theta}_0 \) to \( \bar{\theta}_1(\geq \bar{\theta}_0) \),
the following contradiction would appear.

$$0 = \Pi(\bar{\theta}_1, A_1) \equiv \pi(\bar{\theta}_1, T(\bar{\theta}_1, A_1), A_1) - f$$
$$\geq \pi(\bar{\theta}_1, T(\bar{\theta}_0, A_0), A_1) - f$$
$$> \pi(\bar{\theta}_0, T(\bar{\theta}_0, A_0), A_0) - f = \Pi(\bar{\theta}_0, A_0) = 0.$$  

The first inequality comes from firm’s revealed preference on the number of layers, and the second inequality is due to the fact that firm’s profit function defined in equation (1.15) strictly increases with both $\theta$ and $A$. Therefore, equation (1.22) leads a negative relationship between $\bar{\theta}$ and $A$. Of course, when $\bar{\theta}$ approaches zero, $A$ determined from equation (1.22) approaches infinity. And when $\bar{\theta}$ goes to infinity, $A$ determined from equation (1.22) approaches zero.

Second, let me discuss the FE condition. The goal is to show that for all pairs of $(\bar{\theta}, A)$ that satisfy the ZCP condition, there is a positive relationship between these two variables determined by the FE condition. Suppose $\bar{\theta}$ decreases from $\bar{\theta}_0$ to $\bar{\theta}_1 (< \bar{\theta}_0)$ in equation (1.23). If the adjusted market size $A$ increased from $A$ to $A_1(\geq A_0)$, the following result must be true.

$$f_e = \int_{\bar{\theta}_1}^{\infty} \Pi(\theta, A_1)g(\theta)d\theta$$
$$= \int_{\bar{\theta}_1}^{\bar{\theta}_0} \Pi(\theta, A_1)g(\theta)d\theta + \int_{\bar{\theta}_0}^{\infty} \Pi(\theta, A_1)g(\theta)d\theta$$
$$> \int_{\bar{\theta}_0}^{\infty} \Pi(\theta, A_1)g(\theta)d\theta$$
$$> \int_{\bar{\theta}_0}^{\infty} \Pi(\theta, A_0)g(\theta)d\theta$$
$$= f_e,$$

which is a contradiction. In the above derivation, I have implicitly used the ZCP condition which implies $\Pi(\theta, A_1) \geq 0$ for all $\theta \in [\bar{\theta}_1, \bar{\theta}_0]$. In total, the downward sloping
ZCP curve and upward sloping FE curve intersects only once, and the intersection pins down a unique pair of $(\bar{\theta}, A)$ for the product market equilibrium.

Now, I prove the uniqueness. Suppose there were two pairs of $(\bar{\theta}, A)$ (i.e., $(\bar{\theta}_1, A_1)$ and $(\bar{\theta}_2, A_2)$) that satisfy both the ZCP condition and the FE condition. Without loss of generosity, let me assume that $\bar{\theta}_1 > \bar{\theta}_2$. Due to the property of the ZCP condition, it must be true that $A_1 < A_2$ which contradicts the positive relationship between $\bar{\theta}$ and $A$ implied by the FE condition. Therefore, the equilibrium must be unique.

Finally, I prove the existence. For any $A \in (0, \infty)$, there exists a unique $\bar{\theta}(A)$ with $\bar{\theta}'(A) < 0$ determined by the ZCP condition. Furthermore, $\bar{\theta}(A)$ decreases continuously in $A$, as the firm’s profit function with the optimal number of layers increases continuously with $\theta$ conditional on $A$. Therefore, among those $(A, \bar{\theta}(A))$ that satisfy the ZCP condition, there must be a pair of $(\bar{\theta}, A)$ that satisfies the FE condition. QED.

Third, the following claim shows that there is a unique $p_h$ that clears the labor market in general.

**Claim 4** When $\frac{\sigma - 1}{\sigma} \neq \gamma$ and parameter values satisfy certain conditions, there exists a unique wage $p_h$ that clears the labor market given that the product markets are cleared.

Proof. First, let me decompose the total wage payment of the CES sector and the number of workers employed in the CES sector into the following two parts:

\[
WP(\bar{\theta}, A, M) = WP_{\text{per}}(A, \bar{\theta}) \ast M
\]

and

\[
LD(\bar{\theta}, A, M) = LD_{\text{per}}(A, \bar{\theta}) \ast M,
\]
where “per” means per firm. Second, Substituting the above two expressions into equation \(1.55\) yields

\[
\frac{WC_{\text{per}}(A, \bar{\theta}) - \psi LD_{\text{per}}(A, \bar{\theta})}{p_h} M + \frac{(1 - \gamma)A^\sigma P^{1-\sigma}}{\gamma p_h} = L. \tag{1.58}
\]

Next, substituting equation \(1.57\) into equation \(1.3\) leads to the expression of \(M\) in terms of \(p_h\) and \(P_1(\bar{\theta}, A)\) as follows:

\[
M = \frac{(1 - \gamma)(\sigma - 1)}{\sigma - 1} p_h P_1(\bar{\theta}, A)^{\sigma - 1}. \tag{1.59}
\]

Finally, substituting equations \(1.57\) and \(1.59\) into equation \(1.55\) results in the following labor market clearing condition:

\[
\left[ WC_{\text{per}}(A, \bar{\theta}) - \psi LD_{\text{per}}(A, \bar{\theta}) \right] P_1(\bar{\theta}, A)^{\sigma - 1} + \frac{(1 - \gamma)A^\sigma}{\gamma} = \frac{(1 - \gamma)(\sigma - 1)}{\gamma} p_h L. \tag{1.60}
\]

There exists a unique \(p_h\) that satisfies the above equation, as long as \(\frac{(1 - \gamma)(\sigma - 1)}{\gamma} \neq 1\).\(^{43}\)

Moreover, equilibrium \(p_h\) must satisfy the condition that

\[
w_{\text{min}} \geq \psi(i) + p_h,
\]

where \(w_{\text{min}}\) is the minimum wage offered in the CES sector. This puts a constraint on parameter values, which I will discuss soon. QED.

There are three effects on the labor market when the price of the homogeneous good goes up. First, as \(p_h\) is the wage offered in the homogeneous sector, labor demand of firms in the homogeneous sector goes down. Second, as \(p_h\) is the outside option for workers entering the CES sector, the number of them must go down in

\(^{43}\)Note that I have implicitly used the product market equilibrium conditions to derive the above equation.
order to make the worker who chooses to enter the CES sector earn higher expected payoff. These two negative effects on the labor demand are reflected by \( p_h \) that appears in the left hand side of equation (1.58). Finally, increasing market size due to a bigger \( p_h \) makes the aggregate income \( E(=A^\sigma P^{1-\sigma}) \) and the mass of firms \( M \) increase which pushes up the aggregate labor demand in the end. Therefore, whether or not the aggregate labor demand increases with \( p_h \) depends on whether or not the third (positive) effect dominates the first two negative effects. However, in either case, the aggregate labor demand is a monotonic function of \( p_h \) which assures the uniqueness of \( p_h \) that clears the labor market.

With Claim 3 and Claim 4 in hand, I only have to show that there is a non-empty set of parameter values within which the incentive compatible wage defined in equation (1.8) satisfies the constraint indicated in equation (1.30) in every labor submarket. In other words, I have to show that the minimum wage offered in the CES sector is weakly bigger than the wage offered in the homogeneous sector plus the disutility of exerting effort, or

\[
w_{\text{min}} \geq \psi(i) + p_h.
\]

First, note that labor endowment \( L \) does not affect wages and the minimum wage offered in the CES sector. This is because the solution of \((\bar{\theta}, A)\) in equilibrium does not depend on \( L \), and wages offered by firms in the CES sector only depend on \((\bar{\theta}, A, b)\).\footnote{Remember that the labor demand per firm in the CES sector is independent of \( p_h \) conditional on \((A, \bar{\theta})\).} Second, equation (1.60) indicates that \( p_h \) approaches zero when \( L \) approaches zero and \( \frac{\sigma-1}{\sigma} > \gamma \), and \( p_h \) approaches zero when \( L \) goes to infinity and \( \frac{\sigma-1}{\sigma} < \gamma \). Therefore, \( \frac{\sigma-1}{\sigma} > \gamma \). Therefore,
I conclude that there must exist a small enough $L$ such that

$$w_{\min} - \psi > p_h,$$

when $\frac{2-1}{\sigma} > \gamma$. Similarly, there must exist a big enough $L$ such that

$$w_{\min} - \psi > p_h,$$

when $\frac{2-1}{\sigma} < \gamma$.

In total, I show that with restrictions on parameter values, there must exist a unique equilibrium with unemployment in every labor submarket. The equilibrium is characterized by a unique quadruplet $(\bar{\theta}, M, p_h, E)$. QED.

1.7.11 Proof of Proposition 8

This proof consists of seven parts. I prove that the exit cutoff for the quality draw increases and all firms increase the number of layers first.

I make the following notations. Suppose $b$ decreases from $b_1$ to $b_2 (< b_1)$ due to an improvement in MT. Let $\bar{\theta}_1$ (or $\bar{\theta}_2$) be the demand threshold for exiting when $b = b_1$ (or $b = b_2$). Let $A_1$ (or $A_2$) be the adjusted market size when $b = b_1$ (or $b = b_2$).

First, I discuss how the adjusted market size $A$ changes when $b$ decreases by proving the following lemma.

**Lemma 7** When $b$ decrease from $b_1$ to $b_2$, the change in the adjusted market size must satisfy

$$1 > \frac{A_2}{A_1} > \frac{b_2}{b_1}.$$

Proof. First, note that if $A_2 \geq A_1$, the exit cutoff $\bar{\theta}$ must decrease as $b_2 < b_1$. However, a decreasing exit cutoff plus a weakly increasing adjusted market size violate the FE condition defined in Equation (1.23). Thus, it must be true that $A_2 < A_1$. Thus, it must be true that $A_2 < A_1$. 88
Second, if \( \frac{A_2}{A_1} \leq \frac{b_2}{b_1} \), the profit defined as the solution to Equation (1.9) must decrease for all firms. Thus, the exit cutoff must increase. However, the FE condition is violated again, as profit for all firms decreases, and the exit cutoff increases. In total, it must be true that

\[
1 > \frac{A_2}{A_1} > \frac{b_2}{b_1}.
\]

QED.

Second, I show that all firms increase the number of layers weakly. It is straightforward to observe that if \( \frac{A_2}{A_1} = \frac{b_2}{b_1} \), the optimal output, employment, and the number of layers would be unchanged. As I have proven that \( \frac{A_2}{A_1} > \frac{b_2}{b_1} \) in Lemma 7, all surviving firms weakly increase their number of layers. Furthermore, all surviving firms increase their output as well as employment after the management technology improves.

Third, I prove that the exit cutoff increases. I use \( T_0 + 1 \equiv T(\bar{\theta}_1, A_1, b_1) + 1 = T(\bar{\theta}_2, A_2, b_2) + 1 \) to denote the number of layers for firms on the exit cutoff and prove this result by contradiction. Suppose that the exit cutoff \( \bar{\theta} \) decreased weakly after MT improves (i.e., \( \bar{\theta}_2 \leq \bar{\theta}_1 \)). First, firms on the exit cutoff earn zero payoff due to the ZCP condition or

\[
\pi(\bar{\theta}_1, T(\bar{\theta}_1, A_1, b_1), A_1, b_1) = \pi(\bar{\theta}_2, T(\bar{\theta}_2, A_2, b_2), A_2, b_2) = f,
\]

as \( T_0 = T(\bar{\theta}_1, A_1, b_1) = T(\bar{\theta}_2, A_2, b_2) \). This leads to

\[
\frac{\pi(\bar{\theta}_2, T_0, A_2, b_2)}{\pi(\bar{\theta}_1, T_0, A_1, b_1)} = \left( \frac{\bar{\theta}_2}{\bar{\theta}_1} \right)^{2T_0 \sigma} \left( \frac{A_2}{A_1} \right)^{2T_0 \frac{\sigma}{2T_0 - 1}} \left( \frac{b_2}{b_1} \right)^{2T_0 \frac{\sigma - 1}{2T_0 - 1}} \equiv X(\bar{\theta}, T_0)Y(A, T_0)Z(b, T_0) = 1,
\]

89
where $\bar{\theta} = \bar{\theta}_2$, $A \equiv \frac{A_2}{A_1} < 1$, and $b \equiv \frac{b_1}{b_2} > 1$. As $\bar{\theta}_2 \leq \bar{\theta}_1$,

$$Y(A, T_0)Z(b, T_0) \geq 1.$$ 

Second, For a firm whose demand draw is higher than $\bar{\theta}_1$, its profit must increase if it does not change the number of layers. This is because

$$\frac{\pi(\theta, T(\theta, A_2, b_2), A_2, b_2)}{\pi(\theta, T(\theta, A_1, b_1), A_1, b_1)} = \left( \frac{A_2}{A_1} \right)^{\frac{2T(\theta)}{\sigma + (2T(\theta)-1)}} \left( \frac{b_1}{b_2} \right)^{\frac{(\sigma-1)2T(\theta)}{\sigma + (2T(\theta)-1)} \frac{(2T(\theta)-1)}{\sigma}} \geq 1,$$

where $T(\theta) \equiv T(\theta, A_1, b_1) = T(\theta, A_2, b_2)$ and $T(\theta) \geq T_0$ as $\theta \geq \bar{\theta}_1$. If the firm endogenously changes the number of layers, its profit must be bigger than the profit it earns when $b = b_1$ as well due to the revealed preference argument. In total, I have

$$\pi(\theta, T(\theta, A_2, b_2), A_2, b_2) \geq \pi(\theta, T(\theta, A_1, b_1), A_1, b_1) \forall \theta \geq \bar{\theta}_1$$

for $T(\theta, A_2, b_2) = T(\theta, A_1, b_1)$ and

$$\pi(\theta, T(\theta, A_2, b_2), A_2, b_2) > \pi(\theta, T(\theta, A_1, b_1), A_1, b_1) \forall \theta > \bar{\theta}_1$$

for $T(\theta, A_2, b_2) > T(\theta, A_1, b_1)$. Third, the ZCP condition in the new equilibrium implies that firms with the quality draws between $\bar{\theta}_2$ and $\bar{\theta}_1$ earn non-negative profit. In total, the expected profit from entry would exceed the entry cost $f_e$ if the exit cutoff decreased which violates the FE condition. Therefore, the exit cutoff must increase when $b$ decreases.

\[ ^{46}\text{Taking the log of } \frac{\pi(\theta, T(\theta, A_2, b_2), A_2, b_2)}{\pi(\theta, T(\theta, A_1, b_1), A_1, b_1)} \text{ leads to } B(T, A, b) = \frac{2T}{\sigma + 2T} \log(A) + \frac{(\sigma-1)2T}{\sigma + 2T-1} \log(b). \text{ As } B(T_0, A, b) > 0 \text{ and } \log(b) > 0, B(T, A, b) > B(T_0, A, b) > 1 \text{ for all } T \geq T_0. \]
Fourth, I prove that the distribution of the number of layers moves to the right in the FOSD sense when MT improves. I make the following simplifying notations. Let $\theta_{T,2}$ be the threshold for the quality draw at which the firm increases the number of layers from $T + 1$ to $T + 2$. Let $\text{Prob}(t > T, b)$ be the fraction of firms that have at least $T + 2$ layers when the quality of MT is $b$. Based on the above notations and the Pareto distribution on $\theta$, I have

$$\text{Prob}(t > T, b) = \left( \frac{\bar{\theta}}{\theta_{T,2}} \right)^k.$$ 

Therefore, the condition for $\text{Prob}(t > T, b_2) > \text{Prob}(t > T, b_1)$ to hold is

$$\frac{\bar{\theta}_1}{\theta_{T,2}} \Big|_{b=b_1} < \frac{\bar{\theta}_2}{\theta_{T,2}} \Big|_{b=b_2},$$

where $T \geq T_0$. I derive the expression for $\theta_{T,2}$ and prove the above inequality in what follows. First, conditional on $(b, A)$, the threshold for the firm to add a layer is

$$\frac{2^{T+1}}{\theta_{T,2}^{(2^T+1)-2}} - \frac{2^T}{\theta_{T,2}^{(2^T)-1}} = \frac{b^{(\sigma-1)(2^{T+1})-1}}{\sigma^{(2^T+1)-1}} \frac{(\sigma-1)(2^{T+1}-1)}{\sigma^{(2^T)-1}} (1 - \frac{\beta(2^{T+1})}{2^{T+1}}) \left( \psi \frac{2^{T+2}-2(T+1)}{2^{T+1}-1} / \beta \right) \frac{(\sigma-1)(2^{T+1})}{\sigma^{(2^T+1)-1}}$$

Thus, the ratio of $\frac{\theta_{T,2}^{b=b_1}}{\theta_{T,2}^{b=b_2}}$ can be written as

$$\left( \frac{\theta_{T,2}^{b=b_1}}{\theta_{T,2}^{b=b_2}} \right) = \frac{b^{(\sigma-1)(2^{T+1})-1}}{\sigma^{(2^T+1)-1}} \frac{(\sigma-1)(2^{T+1}-1)}{\sigma^{(2^T)-1}} \frac{\sigma^{2T+1}}{\sigma^{(2^T+1)-1}} \frac{\sigma^{2T}}{\sigma^{(2^T)-1}} \frac{A^{(2^T+1)-1}}{A^{(2^T)-1}},$$

where $A \equiv \frac{A_2}{A_1} < 1$, and $b \equiv \frac{b_1}{b_2} > 1$. This expression can be simplified further to

$$\frac{\theta_{T,2}^{b=b_1}}{\theta_{T,2}^{b=b_2}} = (bA)^\sigma.$$ 

(1.61)
Second, from the expression of firm’s profit function derived in Equation \((1.15)\), I have
\[
\frac{\bar{\theta}}{\bar{\theta}_2} = \bar{\theta}_1 b^{(\sigma-1)(1-\frac{1}{2T_0})}.
\] (1.62)

Finally, from equations \((1.61)\) and \((1.62)\), I conclude that
\[
\frac{\bar{\theta}_2}{\bar{\theta}_1} = \frac{\bar{\theta}_1 b^\sigma}{\bar{\theta}_2 b^{(\sigma-1)(1-\frac{1}{2T_0})}} > 1.
\]

Therefore, for all \(T \geq T_0\), \(\text{Prob}(t > T, b_2) > \text{Prob}(t > T, b_1)\) which is the condition for the result of the FOSD to hold.

Fifth, I prove that the firm size distribution in terms of revenue moves to the right in the FOSD sense when MT improves. I make the following simplifying notations. Let \(S(\bar{\theta}, A_i) \equiv S(\bar{\theta}, A_i, T(\bar{\theta}, A_i))\) be the revenue for firms with quality draw \(\bar{\theta}_i\) when they optimally choose the number of layers, and \(S(\bar{\theta}, A_i, T)\) be the revenue for firms with quality draw \(\bar{\theta}_i\) when they choose to have \(T+1\) number of layers. Similarly, let \(q(\bar{\theta}, A_i) \equiv q(\bar{\theta}, A_i, T(\bar{\theta}, A_i))\) be the output for firms with quality draw \(\bar{\theta}_i\) when they optimally choose the number of layers, and \(q(\bar{\theta}, A_i, T)\) be the output for firms with quality draw \(\bar{\theta}_i\) when they choose to have \(T+1\) number of layers.

As the distribution of \(\theta\) is Pareto, and the firm’s revenue increases with \(\theta\), what I have to show is that for any \(t > 1\),
\[
S(t\bar{\theta}_2, A_2) \geq S(t\bar{\theta}_1, A_1).
\]

As the distribution of the number of layers after an improvement in MT first order stochastically dominates the one before the management technology improves, I have the following two cases:
\[
T(t\bar{\theta}_2, A_2) = T(t\bar{\theta}_1, A_1)
\]
or

\[ T(t\bar{\theta}_2, A_2) > T(t\bar{\theta}_1, A_1). \]

I discuss these two cases one by one in what follows.

In the case of \( T(t\bar{\theta}_2, A_2) = T(t\bar{\theta}_1, A_1) \), if \( t \) is small enough such that \( T(t\bar{\theta}_2, A_2) = T(t\bar{\theta}_1, A_1) = T_0 \), then it is straightforward to see that

\[ S(t\bar{\theta}_2, A_2) = S(t\bar{\theta}_1, A_1). \]

For \( T(t\bar{\theta}_2, A_2) = T(t\bar{\theta}_1, A_1) = T_1 > T_0 \), I have

\[
S(t\bar{\theta}_2, A_2) = S(\bar{\theta}_2, A_2) V1(t, T_0, T_1) \left( \frac{\theta_2 A_2^{\bar{T}_1}}{\sigma + (\bar{T}_1 - 1)} - \frac{2T_0}{\sigma + (2T_0 - 1)} \right)_{\bar{b}_2}
\]

and

\[
S(t\bar{\theta}_1, A_1) = S(\bar{\theta}_1, A_1) V1(t, T_0, T_1) \left( \frac{\theta_1 A_1^{\bar{T}_1}}{\sigma + (\bar{T}_1 - 1)} - \frac{2T_0}{\sigma + (2T_0 - 1)} \right)_{\bar{b}_1},
\]

where \( V1(t, T_0, T_1) \) is a function of \((t, T_0, T_1)\). As \( S(\bar{\theta}_2, A_2) = S(\bar{\theta}_1, A_1) = \frac{f}{\beta^2 (2T_0 - 1)} \)

and

\[ \frac{\bar{\theta}_1}{\theta_2} = A \bar{b}^{(\sigma - 1)(1 - \bar{b})}. \]

Based on Equation (1.62), I conclude that

\[
\frac{S(t\bar{\theta}_2, A_2)}{S(t\bar{\theta}_1, A_1)} = \left[ \left( \frac{1}{b} \right)^{(\sigma - 1)(1 - \frac{1}{2T_0})} b^\sigma \right] \left( \frac{2T_1 - 2T_0}{\sigma + (2T_1 - 1)(\sigma + 2T_0 - 1)} \right) > 1. \tag{1.63}
\]

In the case of \( T(t\bar{\theta}_2, A_2) = T_2 > T(t\bar{\theta}_1, A_1) = T_1 \), I prove \( S(t\bar{\theta}_2, A_2) > S(t\bar{\theta}_1, A_1) \) using the result that when the firm optimally chooses to add a layer, output jumps
up discontinuously. Note that

\[
\frac{S(t\bar{\theta}_2, A_2, T_1)}{S(t\bar{\theta}_1, A_1, T_1)} \geq 1,
\]

and the equality holds only when \( T_1 = T_0 \) due to Equation (1.63). when the firm optimally chooses to add layers, it must be true that

\[
q(t\bar{\theta}_2, A_2, T_2) > q(t\bar{\theta}_2, A_2, T_1)
\]

and

\[
S(t\bar{\theta}_2, A_2, T_2) = A_2(t\bar{\theta}_2)^{\frac{1}{2}} q(t\bar{\theta}_2, A_2, T_2)^{\beta} \\
> S(t\bar{\theta}_2, A_2, T_1) = A_2(t\bar{\theta}_2)^{\frac{1}{2}} q(t\bar{\theta}_2, A_2, T_1)^{\beta} \\
> S(t\bar{\theta}_1, A_1, T_1).
\]

Thus, \( S(t\bar{\theta}_2, A_2) \) must be bigger than or equal to \( S(t\bar{\theta}_1, A_1) \) in all possible cases. Especially, \( S(t\bar{\theta}_2, A_2) = S(t\bar{\theta}_1, A_1) \) only when \( T(t\bar{\theta}_2, A_2) = T(t\bar{\theta}_1, A_1) = T_0 \). Therefore, the result of the FOSD for the distribution of firms’ revenue follows.

Sixth, I prove the result of FOSD for the distribution of the firms’ output and employment. Similar to what I have proven above, the goal is to show that for any \( t > 1 \),

\[
q(t\bar{\theta}_2, A_2) \geq q(t\bar{\theta}_1, A_1).
\]

for all \( t > 1 \). First, I prove that when \( T(\bar{\theta}_1, A_1) = T(\bar{\theta}_2, A_2) = T_0 \),

\[
q(\bar{\theta}_2, A_2, T_0) > q(\bar{\theta}_1, A_1, T_0).
\]
To see this, note that

\[
TVC(q(\bar{\theta}_2, A_2, T_0), b_2, T_0) = \frac{\beta(2T_0 - 1)}{2T_0} S(\bar{\theta}_2, A_2, T_0)
\]

\[
= TVC(q(\bar{\theta}_1, A_1, T_0), b_1, T_0) = \frac{\beta(2T_0 - 1)}{2T_0} S(\bar{\theta}_1, A_1, T_0),
\]

where

\[
TVC(q, T, b) = (2 - \frac{1}{2^{-T-1}})b\psi 2^{1 - \frac{T}{2^{T-1}}} q^{2^{T-1}},
\]

\[
S(\bar{\theta}_2, A_2, T_0) = S(\bar{\theta}_1, A_1, T_0) = \frac{f}{1 - \frac{\beta(2T_0-1)}{2^{T_0}}}
\]

and

\[
b_1 > b_2.
\]

Second, Based on the above result I derive that

\[
q(t\bar{\theta}_2, A_2, T_0) = q(\bar{\theta}_2, A_2, T_0) t^{\frac{2^{T_0-1}}{2^{T_0}+2^{T_0-1}}} > q(\bar{\theta}_1, A_1, T_0) t^{\frac{2^{T_0-1}}{2^{T_0}+2^{T_0-1}}} = q(t\bar{\theta}_1, A_1, T_0),
\]

if \(T(t\bar{\theta}_2, A_2) = T(t\bar{\theta}_1, A_1) = T_0\). Third, if \(T(t\bar{\theta}_2, A_2) = T(t\bar{\theta}_1, A_1) = T_1 > T_0\), I have

\[
q(t\bar{\theta}_2, A_2, T_1) = q(\bar{\theta}_2, A_2, T_0) V_2(t, T_0, T_1) \left(\frac{A_2^{\frac{1}{2}\bar{\theta}_2^2}}{b_2}\right) \frac{\sigma(2^{T_1-1})}{\sigma(2^{T_1-1})} - \frac{\beta(2T_0-1)}{\sigma(2^{T_0-1})}
\]

and

\[
q(t\bar{\theta}_1, A_1, T_1) = q(\bar{\theta}_1, A_1, T_0) V_2(t, T_0, T_1) \left(\frac{A_1^{\frac{1}{2}\bar{\theta}_1^2}}{b_1}\right) \frac{\sigma(2^{T_1-1})}{\sigma(2^{T_1-1})} - \frac{\beta(2T_0-1)}{\sigma(2^{T_0-1})},
\]

where \(V_2(t, T_0, T_1)\) is a function of \((t, T_0, T_1)\). Based on equation \((1.62)\), I conclude that

\[
\frac{q(t\bar{\theta}_2, A_2, T_1)}{q(t\bar{\theta}_2, A_2, T_0)} = \frac{q(t\bar{\theta}_2, A_2, T_0)}{q(t\bar{\theta}_1, A_1, T_0)} \left(\frac{\beta(2^{T_1-1})}{\beta(2^{T_0-1})}\right) \left(\frac{\sigma(2^{T_1-1})}{\sigma(2^{T_0-1})} - \frac{\beta(2T_0-1)}{\sigma(2^{T_0-1})}\right) > 1,
\]
as $T_1 > T_0$, $b > 1$, and $q(\bar{\theta}_2, A_2, T_0) > q(\bar{\theta}_1, A_1, T_0)$. Fourth, for $T(t\bar{\theta}_2, A_2) = T_2 > T(t\bar{\theta}_1, A_1) = T_1$, I have

$$q(t\bar{\theta}_2, A_2, T_1) > q(t\bar{\theta}_1, A_1, T_1)$$

and

$$q(t\tilde{\theta}_2, A_2, T_2) > q(t\tilde{\theta}_2, A_2, T_1),$$

where the second inequality comes from the result that when the firm optimally chooses to add layers output jumps up discontinuously. Therefore, it must be true that

$$q(t\tilde{\theta}_2, A_2, T_2) > q(t\tilde{\theta}_1, A_1, T_1)$$

for $T_2 > T_1$ as well. This completes the proof for the FOSD result on the distribution of the firms’ output. Finally, as firms with the same level of output (i.e., the same number of production workers) have the same employment, the result of the FOSD holds for the distribution of the firms’ employment as well.

Seventh, I prove that all firms increase the span of control given the number of layers when MT improves. First, the span of control is defined as

$$SC_i(T, q(\theta, b, A, T(\theta, b, A))) = \frac{m_{i+1}(T, q(\theta, b, A, T(\theta, b, A)))}{m_i(T, q(\theta, b, A, T(\theta, b, A)))}, \quad (1.64)$$

where $(T - 1) \geq i \geq 0$, and $q(\theta, b, A, T(\theta, b, A))$ is the number of production workers as well as output. Consider a firm with quality draw $\theta$ that does not adjust the number of layers after MT improves. This means

$$T(\theta, b_1, A_1) = T(\theta, b_2, A_2).$$
Its output and the number of production workers must increase as \[ \frac{A_2}{b_2} > \frac{A_1}{b_1}. \]

The span of control calculated in equation (1.13) increases with the number of production workers. Therefore, every surviving firm increases its span of control at all layers if it does not adjusted the number of layers after MT improves. QED.

1.7.12 Proof of Proposition \[9\]

I rewrite the firm’s optimization problem conditional on its staying in the market (i.e., equation (1.9)) as

\[
\max_{\{m_i\}_{i=1}^T} b \left[ \frac{A}{b} \theta^{\frac{1}{2}} m_i \frac{\sigma - 1}{\sigma - 1} - \sum_{i=1}^T \psi m_i x_i \right]
\]

s.t. \[ x_i = \frac{m_i}{m_{i-1}}, \]
\[ m_0 = 1. \]

It is evident from the above optimization problem that \( b \) does not affect the firm’s choices of employment and output given the number of layers and the value of \( \frac{A}{b} \). Furthermore, \( b \) does not affect the optimal number of layers conditional on \( \frac{A}{b} \), as a change in \( b \) changes profits of the hierarchy with different number of layers proportionately given the value of \( \frac{A}{b} \). Therefore, I only need to discuss how \( \frac{A}{b} \) changes when the quality of MT improves for a single firm and for a group of firms.

When MT improves for a single firm, \( b \) decreases for that firm and the adjusted market size \( A \) does not change, as each firm is “atomic”. Thus, \( \frac{A}{b} \) increases. An increase in \( \frac{A}{b} \) has qualitatively the same effect on the firm’s choices (i.e., employment, output, revenue, and the number of layers) as an increase in \( \theta \). Therefore, the firm’s

\[ 47 \text{For detailed proof of this result, see Appendix 1.7.11.} \]
output, profit, and revenue increase. Furthermore, the number of layers also increases 
*weakly*. And the span of control increases if the number of layers does not change, since the output increases.

However, if MT improves for all other firms except for one firm, the adjusted market size $A$ decreases as proven in Lemma 7. Thus, $\frac{A}{b}$ decreases for that single firm. As a result, its output, profit, and revenue decrease. Furthermore, the number of layers also decreases *weakly*. And the span of control decreases if the number of layers does not change, since the output decreases. QED.

### 1.7.13 The Continuous Number of Layers

In this section, I treat the number of layers as a continuous variable à la Keren and Levhari (1979) and Qian (1994). First, I show that the main results established in the paper (i.e., Proposition 8) hold in *all* possible cases when the number of layers is treated as a continuous variable. Second, I derive a condition under which the type of equilibrium I investigate exists, and the condition only consists of exogenous parameters. Finally, I obtain qualitative results on the welfare effect of an improvement in MT.

The only change in the specifications of the model from the paper is the treatment of the number of layers. First, the effort choice is still treated as a binary variable. As a result, every worker is incentivized to exert effort in equilibrium. Second, the economy still consists of two sectors: a traditional sector and a CES sector. Third, workers can still choose which type of jobs to apply for. Based on these specifications,

\[48\] Again, remember that each firm is “atomic”.

98
I can express the firm’s optimization problem as

\[
\begin{align*}
\max_{s_t, T} & \quad A \theta^T x_T^3 - \int_{t=0}^{T} g s_t x_t dt \\
\text{s.t.} & \quad \dot{x}_t = x_t \log(s_t) \\
& \quad x_0 = 1,
\end{align*}
\]

where \( g \equiv b \psi \). Now, the firm chooses the span of control at each layer as well as the number of layers in equilibrium.\(^{49}\) Different from the optimization problem in Qian (1994), the problem now becomes an “open-final-time” (i.e., \( T \)) and “open-end-point” (i.e., \( x_T \)) optimal control problem with one state variable \( x_t \) and one control variable \( s_t \).

I solve this optimization problem in the following several steps. First, the corresponding Hamiltonian is

\[
H(t) = -g s_t x_t + p_t x_t \log(s_t).
\]

The corresponding optimality conditions are

\[
\dot{p}_t = gs_t - p_t \log(s_t), \tag{1.66}
\]

\[
-g x_t + p_t x_t/s_t = 0 \tag{1.67}
\]

and

\[
H(t) = -g s_t x_t + p_t x_t \log(s_t) = 0. \tag{1.68}
\]

Above three conditions are exactly the same as in Qian (1994), as these three conditions are sufficient and necessary conditions for an “open-final-time” and “fixed-

\(^{49}\)For more details, see Qian (1994).
end-point” optimal control problem. Moreover, there is an additional transversality condition for this “open-final-time” and “open-end-point” problem as follows:

\[ p_T = \frac{\partial A \theta^{1/2} x_T^\sigma}{\partial x_T} = A \beta \left( \frac{\theta}{x_T} \right)^{1/\sigma} = g s_T. \] (1.69)

Next, based on equations (1.66) to (1.69), I derive closed-form solutions for this optimal control problem as follows:

\[ s_t = e, \ w_t = ge, \ x_t = e^t, \ x_T(\theta, A) = \theta \left( \frac{A \beta}{ge} \right)^\sigma, \ T = \log \theta + \sigma \log \left( \frac{A \beta}{ge} \right), \] (1.70)

where \( e \) is Euler’s number and equals 2.71828... Total employment is

\[ l(\theta, A) = \int_{t=0}^{T} x_t dt = (x_T - 1). \] (1.71)

The resulting operating profit and revenue are

\[ \pi(\theta, A) = (1 - \beta) A^\sigma \theta \left( \frac{\beta}{ge} \right)^{\sigma-1} + ge \] (1.72)

and

\[ S(\theta, A) = A^\sigma \theta \left( \frac{\beta}{ge} \right)^{\sigma-1} \] (1.73)

respectively. The operating profit function in equation (1.72) is similar to the standard operating profit function in models with the CES preference except that it includes an additional term (i.e., \( ge \)). This implies that the Free Entry (FE) condition of the current model is the same as the in Melitz (2003) except that the implied fixed cost now becomes \( f - ge \).\textsuperscript{51}

\textsuperscript{50}See Meagher (2003) for more details.

\textsuperscript{51}Throughout the this section, I assume that \( f > ge \).
Third, I prove the main result of this subsection. As in the paper, I still assume that the quality draw follows a Pareto distribution with parameter $k$.

\[ G(\theta) = 1 - \left( \frac{\theta_{\text{min}}}{\theta} \right)^k, \]

where $G(\theta)$ is the Cumulative Distribution Function (CDF) of $\theta$.

**Proposition 10** When the quality of MT improves in an economy (i.e., $b$ decreases), the exit cutoff for the quality draw increases. On top of that, the firm size distribution (i.e., sales or output or employment) and the distribution of the number of layers move to the right in the FOSD sense.

Proof: First, the key observation is that the ratio of operating profit of a firm with a higher quality draw to operating profit of a firm with a lower quality draw increases when the quality of MT improves. More specifically, for $\theta_2 > \theta_1$ I have

\[
\frac{\pi(\theta_2, A)}{\pi(\theta_1, A)} = \frac{(1 - \beta)A^\sigma \theta_2 \left( \frac{\sigma}{g\psi} \right)^{\sigma-1} + ge}{(1 - \beta)A^\sigma \theta_1 \left( \frac{\sigma}{g\psi} \right)^{\sigma-1} + ge}
\]

decreases in $g$. As $g = b\psi$ decreases when the quality of MT improves, this ratio must increase when the quality of MT increases. Therefore, the exit cutoff must increase when $b$ decreases.

Next, I discuss what happens to firms at the exit cutoff. I make the the following simplifying notations before the discussion. Let $\bar{\theta}_1$ and $\bar{\theta}_2 (> \bar{\theta}_1)$ be the exit cutoffs before and after MT improves respectively. Let $A_1$ and $A_2$ be the corresponding adjusted market size. Let $b_1$ and $b_2 (< b_1)$ be the inefficiency of MT before and after MT improves respectively. I prove that firms on the exit cutoff increase their revenue, employment, output and the number of layers after MT improves. First, the revenue

\[^{52}\text{The Pareto assumption is needed only for the proof of the change in distributions. The result on the change of the exit cutoff is independent of this distributional assumption.}\]
must increase for firms at the exit cutoff as

\[ S(\bar{\theta}_1, A_1) = \sigma(f - g_1e) < \sigma(f - g_2e) = S(\bar{\theta}_2, A_2). \]

Second, output that equals \( x_T^* \) also increases for firms at the exit cutoff. This is because

\[ A_1 > A_2. \]

Third, as

\[ T = \log(x_T) \]

and employment equals \( x_T - 1 \), both the number of layers and employment increase for firms at the exit cutoff.

Third, I prove the result that the firm size distribution and the distribution of the number of layers move to the right in the FOSD sense when MT improves. Similar to the proof in the paper, the key thing to show is that for any \( c > 1 \)

\[ S(c\bar{\theta}_2, A_2) > S(c\bar{\theta}_1, A_1), \]

\[ x_T(c\bar{\theta}_2, A_2) > x_T(c\bar{\theta}_1, A_1), \]

\[ l(c\bar{\theta}_2, A_2) > l(c\bar{\theta}_1, A_1) \]

and

\[ T(c\bar{\theta}_2, A_2) > T(c\bar{\theta}_1, A_1). \]

First, note that

\[ \frac{S(c\bar{\theta}, A)}{S(\bar{\theta}, A)} = c \]

and

\[ \frac{x_T(c\bar{\theta}, A)}{x_T(\bar{\theta}, A)} = c. \]
which do not depend on $A$ and $\bar{\theta}$. As it has already been shown that $S(\bar{\theta}_2, A_2) > S(\bar{\theta}_1, A_1)$ and $x_T(\bar{\theta}_2, A_2) > x_T(\bar{\theta}_1, A_1)$, I conclude that

$$S(c\bar{\theta}_2, A_2) > S(c\bar{\theta}_1, A_1)$$

and

$$x_T(c\bar{\theta}_2, A_2) > x_T(c\bar{\theta}_1, A_1),$$

which mean that the distributions of revenue and output move to the right in the FOSD sense when MT improves. Second, as $l(\theta, A) = x_T(\theta, A) - 1$ and $T(\theta, A) = \log(x_T(\theta, A))$ for any $(\theta, A)$, I also have

$$l(c\bar{\theta}_2, A_2) > l(c\bar{\theta}_1, A_1)$$

and

$$T(c\bar{\theta}_2, A_2) > T(c\bar{\theta}_1, A_1)$$

which mean that the distributions of employment and the number of layers move to the right in the FOSD sense as well when MT improves. QED.

The intuitions behind this proposition are the same as the intuitions in the paper. The management shock common across all firms benefits firms with more layers disproportionately more. As a result, the least efficient firms exit the market and the most efficient firms thrive. Therefore, the firm size distribution and the distribution of the number of layers move to the right in the FOSD sense after the management shock.

Fourth, I discuss the welfare implication of an improvement in MT. In order to derive results on welfare, I solve for the exit cutoff and the adjusted market size first. Based on the discussion in the previous section, I derive the ZCP condition and the
FE condition using equation (1.72). More specifically, the ZCP condition states that

\[(1 - \beta)A^\sigma \bar{\theta} \left( \frac{\beta}{ge} \right)^{\sigma - 1} = f - ge. \] (1.74)

The FE condition says that

\[
\int_{\theta = \bar{\theta}}^{\infty} (\pi(\theta, A) - f)g(\theta)d\theta = f_e, 
\]

which can be further reduced to

\[(f - ge) \frac{(\theta_{\text{min}}/\bar{\theta})^k}{(k - 1)} = f_e. \]

Therefore, the exit cutoff can be solved as

\[
\bar{\theta} = \theta_{\text{min}} \left( \frac{f - ge}{f_e(k - 1)} \right)^{\frac{1}{k}}. \] (1.75)

Substituting equation (1.31) into equation (1.28) leads to the solution of the adjusted market size \(A\) as

\[
A^\sigma = \frac{(f - ge)(ge)^{\sigma - 1}}{(1 - \beta)\beta^{\sigma - 1}\bar{\theta}} = \frac{(f - ge)^{1 - \frac{1}{k}}(ge)^{\sigma - 1}((k - 1)f_e)^{\frac{1}{k}}}{(1 - \beta)\beta^{\sigma - 1}\theta_{\text{min}}}. \] (1.76)

Now, I derive an explicit expression of welfare in the closed economy. Since welfare equals \(p_h\) (i.e., wage offered in the homogeneous sector), I calculate the welfare using the following equation:

\[
\left[ WC_{\text{per}}(A, \bar{\theta}) - \psi LD_{\text{per}}(A, \bar{\theta}) \right] P_1(\bar{\theta}, A)^{\sigma - 1} + \frac{(1 - \gamma)A^\sigma}{\gamma} = \frac{1 - (1 - \gamma)(\sigma - 1)}{\gamma} L. \] (1.77)
Based on equations (1.70), (1.71), (1.75) and (1.76), I obtain that

\[ P_1(\bar{\theta}, A) = \left( \frac{ge}{\beta} \right)^{\sigma - 1} \left( \frac{k - 1}{k \bar{\theta}} \right), \]  

(1.78)

\[ LD_{per}(A, \bar{\theta}) = \frac{k}{k - 1} x_T(\bar{\theta}) - 1, \]  

(1.79)

and

\[ WC_{per}(A, \bar{\theta}) = LD_{per}(A, \bar{\theta}) ge = ge \left[ \frac{k}{k - 1} x_T(\bar{\theta}) - 1 \right], \]  

(1.80)

where \( x_T(\bar{\theta}) = \frac{(\sigma - 1)(f - ge)}{ge} \) is the output level of the smallest firms in equilibrium in terms of the final composite good. Note that both \( LD_{per}(A, \bar{\theta}) \) and \( WC_{per}(A, \bar{\theta}) \) must be greater than or equal to zero. Thus, it must be true that

\[ x_T(\bar{\theta}) \geq \frac{k - 1}{k}, \]

which implies that\(^53\)

\[ f > \frac{\sigma - 1}{\sigma - 1} \frac{k - 1}{k} \psi e. \]  

(1.81)

After having substituting equations (1.78) to (1.80) into the L.H.S. of equation (1.77), I end up with

\[ \psi \left( be - 1 \right) \left[ \frac{k(\sigma - 1)(f - ge)}{(k - 1)ge} - 1 \right] \left( \frac{ge}{\beta} \right)^{\sigma - 1} \left( \frac{k - 1}{k \theta_{min}} \right) \left( \frac{e(k - 1)}{(f - ge)} \right)^{\frac{1}{2}} \]

\[ + \frac{(1 - \gamma) (f - ge)^{1 - \frac{1}{2}} (ge)^{\sigma - 1} ((k - 1)f_e)^{\frac{1}{2}}}{\gamma (1 - \beta) \beta^{\sigma - 1} \theta_{min}}. \]

\(^53\)Note that price charged by a firm with a efficiency level \( \theta \) is \( A^{1/\sigma} x_T(\theta)^{-1/\sigma} = \frac{ge}{\beta} \). This means that all firms charge the same price, as the quality of their goods differs.\(^54\)This condition must be true, since the entrepreneur can always produce one unit of good by herself. This implies that output level of any active firm must be greater than or equal to one, if all firms use management hierarchies to produce in equilibrium. As in the main context of the paper, I focus on the case in which there are no self-employed entrepreneurs.
Denote that
\[
B(b) \equiv \frac{(ge)^{\sigma-1}(ge - \psi)}{(f - ge)^{\frac{1}{k}}} \left[ \frac{(\sigma - 1)(f - ge)}{ge} - \frac{k - 1}{k} \right] + \frac{(1 - \gamma)\sigma}{\gamma} (f - ge)^{1 - \frac{1}{k}} (ge)^{\sigma-1}. \tag{1.82}
\]

Therefore, welfare in the closed economy is
\[
p^*_b = \left[ \frac{(f_e(k - 1))^\frac{1}{k} B(b)}{\beta^{\sigma-1}\theta_{min}L} \right]^{\frac{1}{\gamma-\frac{1}{k}(\sigma-1)}}, \tag{1.83}
\]
and there are welfare gains from an improvement in MT if and only if one of the following conditions holds.

\[
\frac{d \ln B(b)}{db} > 0, \quad \frac{\sigma - 1}{\sigma} > \gamma
\]
or
\[
\frac{d \ln B(b)}{db} < 0, \quad \frac{\sigma - 1}{\sigma} < \gamma.
\]

The above condition is not intuitive, since it involves many parameters. Now, I derive a necessary and sufficient condition for the existence of welfare gains from better MT. From equation (1.82), I conclude that \( B'(b) > 0 \) if both

\[
(f - b\psi e)^{1 - \frac{1}{k}} (b\psi e)^{\sigma-1}
\]
and
\[
(b\psi e)^{\sigma-2} (f - b\psi e)^{1 - \frac{1}{k}} (b\psi e - \psi)
\]
increase in \( t \). Calculation shows that the first expression increases with \( b \), since
\[
\frac{\sigma - 1}{b\psi e} > \frac{1 - \frac{1}{k}}{f - b\psi e}
\]
implied by condition (1.81). Moreover, the second expression also increases with \( b \), since
\[
\frac{\sigma - 2}{b \psi e} + \frac{1}{b \psi e - \psi} > \frac{\sigma - 1}{b \psi e} > \frac{1 - \frac{1}{k}}{f - b \psi e}.
\]
Therefore, \( B(b) \) increases with \( b \). As a result, welfare increases after an improvement in MT if and only if
\[
\frac{\sigma - 1}{\sigma} > \gamma.
\]

Finally, I discuss the condition under which the type of equilibrium I investigate exists. Namely, when does every labor submarket have unemployment? Every employed worker receives \( b \psi e \) as wage compensation in the CES sector and every worker receive \( p_h \) as wage compensation in the homogeneous sector. Therefore, a necessary and sufficient condition for the equilibrium to exist is that
\[
p_h^{c} \leq \psi (be - 1).
\]

Note that the above condition contains only exogenous parameters, as \( p_h^{c} \) only contains exogenous parameters. Furthermore, this condition puts an upper bound on \( L \) when \( \frac{\sigma - 1}{\sigma} > \gamma \) and a lower bound on \( L \) when \( \frac{\sigma - 1}{\sigma} < \gamma \).

### 1.7.14 Another Type of Equilibrium

The paper considers one type of equilibrium in which there is unemployment in every labor submarket. This type of equilibrium must uniquely exist under restrictions on parameter values. However, when the outside option of workers is high enough in equilibrium, firms in the CES sector that pay the lowest wages are forced to raise their wages up to the level of the sum of the workers’ outside option and the cost to exert effort. As a result, job applicants who enter the labor submarkets of these firms are fully employed. In other words, these labor submarkets are perfectly competitive.
markets. Therefore, this type of equilibrium contains both labor submarkets that have no unemployment and labor submarkets that have unemployment. This is the third type of equilibrium which I have not considered in the paper.

I change the setup of the model slightly in order to simplify the analysis. Different from the paper, I assume that there is only one sector (i.e., the CES sector) in the economy (i.e., $\gamma = 1$), and there are infinitely many workers who have a fixed outside option $p_h$ (the net payoff they receive from working elsewhere) and can choose whether or not to enter the CES sector. Admittedly, I cannot discuss welfare implications under this alternative specification of the model, since the welfare of workers is fixed by an exogenous parameter (i.e., $p_h$). However, I show that qualitative results of the model except for the results on welfare hold in this type of equilibrium as well using a numerical example.

There are four equilibrium conditions in total under the above alternative specifications. First, the two equilibrium conditions for the product market (i.e., the CES sector) are still the ZCP condition and the FE condition (i.e., equations (1.22) and (1.23)). Second, the outside option of workers (i.e., $p_h$) pins down the number of job applicants in each labor submarket (e.g., equation (1.25)) and the total number of job applicants in the CES sector (e.g., equation (1.27)). Finally, the aggregate income $E$ is still determined by the market-clearing condition of the final composite good that is equation (1.29). Note that as the pool of potential entrepreneurs is large enough, the FE condition holds with equality. The indifference condition of workers holds with equality as well, since there are infinitely many workers who are willing to enter the CES sector, as long as the expected payoff obtained from entering the CES sector is bigger than or equal to $p_h$. In total, I end up with four equations (i.e., equations (1.22), (1.23), (1.25), and (1.27)).

The first type of equilibrium is an equilibrium in which every labor submarket is a competitive market that has a uniform wage and no unemployment. This equilibrium exists when the cost of exerting effort were zero (i.e., no incentive problems). The second type of equilibrium is an equilibrium in which every labor submarket has unemployment. This type of equilibrium exists when the cost of exerting effort is high enough.
(1.22), (1.23), (1.27) and (1.29)) and four endogenous variables: the exit cutoff $\bar{\theta}$, the mass of firms $M$, the aggregate labor demand of the CES sector $L_c$ and the aggregate income $E$. I eliminate one condition (i.e., equation (1.29)) and normalize the ideal price index of the CES goods to one.

I show that qualitative results derived in the paper hold in the current case as well. As it is hard to prove results analytically, I simulate the model to show the existence of the pro-competitive effect of improvements in MT. The approach to solve the model in the current case is similar to the simulation algorithm I have used in the paper. More specifically, first I assume that the outside option of workers (i.e., $p_h$) does no bind when firms choose their optimal wage schedules (i.e., assume that $w_i(\theta) - \psi \geq p_h \forall (i, \theta)$), and then solve for the value of $(\bar{\theta}, A)$. Next, I check whether or not $w_i(\theta) - \psi \geq p_h$ for all $(i, \theta)$, and I consider the case in which $w_i(\theta) - \psi < p_h$ for some $(i, \theta)$. In the simulation example, I find that among firms that have two layers (i.e., $T = 1$), the smallest ones would have offered wages lower than $p_h + \psi$, if they ignored the outside option of workers. Thus, the optimization problem of these firms has to be resolved, and the outside option of workers has to be taken into account when the firm designs the optimal wage schedule.

I solve the optimization problem of firms that are constrained to offer wage schedules in two steps. First, I derive the cutoff for $\theta$ below which firms have to raise their wages up to $p_h + \psi$. As only firms that have two layers may be constrained to offer the wage schedules in the simulated example, I only consider firms that have two layers now. Without being constrained by the outside option of workers, a firm that has two layers would offer wage $w_i(\theta)$ at

$$w_i(\theta) = \psi b\psi \left( \frac{A/s}{2b\psi} \right)^{\frac{s}{s+1}}.$$  

\footnote{As workers’ wage at layer $i$ does not change monotonically with firm size, it is not necessarily true that $w_i(\theta) - \psi < p_h$ for the smallest firms.}
Therefore, the cutoff $\theta_1$ is

$$
\theta_1 = \left(\frac{2}{A\beta}\right) \frac{(p_h + \psi)^{\sigma+1}}{b\psi}.
$$

(1.84)

Next, I discuss the optimization problem for a firm that has the demand draw below $\theta_1$. Its objective function can be stated as

$$
\max_{w,N} \quad A\theta^{\frac{1}{2}} N^\beta - wN
$$

s.t.

$$
w \geq p_h + \psi, \\
\frac{w}{bN} \geq \psi,
$$

where the first constraint is the lower bound on wage that the firm can offer, and the second constraint is the incentive compatibility (IC) constraint for the production workers to exert effort.\footnote{Note that the span of control is $\frac{1}{N}$ for the entrepreneur.} There are two types of solutions to the above optimization problem depending on the value of $\bar{\theta}$. First, if $\bar{\theta} \geq \frac{\theta_1}{2\sigma}$, the solution is

$$
w^* = p_h + \psi, \quad N^* = \frac{p_h + \psi}{b\psi}
$$

(1.85)

for $\theta \in [\bar{\theta}, \theta_1]$. Second, if $\bar{\theta} < \frac{\theta_1}{2\sigma}$, the solution for firms with $\theta \in \left[\frac{\theta_1}{2\sigma}, \theta_1\right]$ is still characterized by equation (1.85). The solution for firms with $\theta \in [\bar{\theta}, \frac{\theta_1}{2\sigma}]$ is characterized by

$$
w^* = p_h + \psi, \quad N^*(\theta) = \left(\frac{A\beta \theta^{\frac{1}{2}}}{p_h + \psi}\right)^\sigma.
$$

(1.86)

The average variable cost (AVC) and the marginal cost (MC) are slightly different in the current case compared with the case discussed in paper. First, the AVC curve given $T = 1$ becomes a horizontal line at $w_0 = p_h + \psi$ for output less than $N_0 = \frac{p_h + \psi}{b\psi}$. Firms cannot reduce wages paid to production workers below $p_h + \psi$, even if output and the span of control fall below $\frac{p_h + \psi}{b\psi}$. Second, the MC curve also becomes a
horizontal line at $w_0 = p_h + \psi$ for output less than $N_0$, as firms pay $p_h + \psi$ to every production worker for any output level smaller than $\frac{p_h + \psi}{\psi}$. This implies that there is a discontinuous increase in the MC when output exceeds $N_0$. When output exceeds $N_0$, a firm has to increase wage payment to all existing workers in order to produce more, while wage paid to every worker is a constant (i.e., $w_0$) for any output level less than $N_0$. Figure 1.11 shows the AVC and MC curves of the firm in the current case.

Due to these changes in the cost functions, there are three cases to consider when the firm produces output using a management hierarchy with two layers. First, if the demand draw is bigger than $\theta_1$, which implies that the optimal output exceeds $N_0$, output and wage are the same as the ones I have derived in the paper (i.e., equation (14) in the paper and the equation for $w_i(\theta)$ on page 48 of the paper). This case is illustrated by the marginal revenue (MR) curve for $\theta = 4.5$ in Figure 1.11. Second, if the demand draw of a firm is between $\frac{\theta_1}{2\sigma}$ and $\theta_1$, the firm produces output at $N_0$, as there is a discontinuous increase in the MC at $N_0$. This case is represented by the MR curve for $\theta = 1$ in Figure 1.11. Finally, if the demand draw of a firm is smaller than $\frac{\theta_1}{2\sigma}$, the firm faces a constant MC. The optimal wage and output are solved in equation
Note that whether or not the third case appears in equilibrium depends on whether or not $\bar{\theta}$ is smaller than $\frac{\theta_1}{2}$. In the numerical example I am going to show, only the first two cases appear in equilibrium.

Simulation results show that qualitative results of the paper still hold in the current case. First, the exit cutoff for demand draw $\theta$ increases when MT improves, which is shown in Table 1.7. Second, Figure 1.12 shows that the firm size distribution moves to the right in the FOSD sense when the quality of MT improves. Third, when MT improves, aggregate productivity of active firms increases due to both the within-firm effect and the between-firm effect. Finally, all surviving firms increase the number of layers weakly when the quality of MT improves. The last two results are shown in Figure 1.13.

Table 1.7: The Cutoffs

<table>
<thead>
<tr>
<th></th>
<th>$\theta$ (the exit cutoff)</th>
<th>$\theta_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>b=1.45</td>
<td>4.4498</td>
<td>4.5998</td>
</tr>
<tr>
<td>b=1.38</td>
<td>4.6991</td>
<td>5.4608</td>
</tr>
</tbody>
</table>

$\psi = 0.44$, $p_h = 0.56$, $\sigma = 3.8$, $k = 1.1$, $\theta_{min} = 1$

There are two effects that make the exit cutoff increase when MT improves. First, an improvement in MT benefits firms having more layers disproportionately more when they are not constrained to offer the wage schedules (i.e., firms having demand draw $\theta(>\theta_1)$). This effect is due to the cost structure of the firm that have been discussed in the paper. Second, firms with the lowest demand draws among firms that have two layers (i.e., firms with demand draw $\theta \in [4.6991, 5.4608]$ in the simulation) cannot fully realize the benefit of improved MT. They offer the lowest wages to production workers before MT improves. When MT improves, they are forced to set the wage at $p_h + \psi$ which is higher than what they would choose without the constraint of the workers’ outside option (i.e., $p_h$). The second effect reinforces the first one, as it also favors firms that have better demand draws. In total, these two effects together make the exit cutoff increase and intensify market competition.
Changes in firm-level and aggregate-level outcomes are results of the intensified competition triggered by an improvement in MT. First, surviving firms become bigger on average when MT improves, since the smallest firms exit the market and the biggest firm expand. Second, aggregate productivity increases, as the least productive firms exit the market and the productivity of surviving firms increases. Finally, surviving firms weakly increase the number of layer, as better MT reduces their labor costs and incentives firms to produce more. In sum, the qualitative results of an improvement in MT discussed in the paper hold in the current case as well.

The takeaway from this subsection is that the key economic force of an improvement in MT discussed in the paper (i.e., firms with better demand draws benefit disproportionately more from such an improvement) also exists in the case in which
some labor submarkets don’t have unemployment. Moreover, there is another channel through which this uneven effect works as what I have discussed above.

1.7.15 Two-Dimensional Heterogeneity in the Quality of MT

I introduce heterogeneity in MT both across firms and across countries in this section. The purpose of this section is to show that the main results of the paper (i.e., Proposition 8) still hold under some assumptions in this alternative environment.

In order to proceed the analysis and obtain analytical results, I make the following simplifying assumptions. First, I assume that the draw of the inefficiency of MT (i.e., $b$) and the quality draw are independent. Second, I assume that the draw of the inefficiency of MT for firms in the advanced country is either $b_1$ or $b_2 (> b_1)$, while this becomes $cb_1$ and $cb_2 (c > 1)$ for firms in the developing country. The probability
of receiving a draw of \( b_1 \) in the advanced country is the same as the probability of obtaining a draw of \( cb_1 \) in the developing country.

Different from the case considered in the paper, I have that the exit cutoff on the quality draw now depends on the draw of the efficiency of MT as well. In other words, the expression of the cutoff now becomes \( \theta(A, b) \). Like before, I only consider the case in which firms on the exit cutoffs for each realization of \( b \) and \( cb \) have the same number of layers in the two economies.

I show that the main results of the paper (i.e., Proposition 8) still hold in this alternative environment by proving the following proposition.

**Proposition 11** A common improvement in MT across all firms increases all the exit cutoffs for the quality draw. Furthermore, both the firm size distribution and the distribution of the number of layers move to the right in the FOSD sense when MT improves. At the firm level, all surviving firms increase the number of layers (weakly) and the span of control given the number of layers.

**Proof.** I use the same trick that is used in the paper to prove the first part of this proposition. More specifically, I postulate that all exit cutoffs either decrease on stay the same first (i.e., \( \bar{\theta}(b) \leq \bar{\theta}(cb) \) for \( b \in \{b_1, b_2\} \)). Then, the expected profit after entry must increase for every realization of \( b \in \{b_1, b_2\} \). Thus, the expected profit from entering the product market must exceed the entry cost \( f_e \) after the draw of the inefficiency of MT improves. This violates the FE condition. Therefore, at least one exit cutoff must increase. Finally, as at least one cutoff increases, the adjusted marker size (i.e., \( A \)) must decrease due to the ZCP condition. This implies all the exit cutoff increase when MT improves.

Next, I prove the FOSD results. I make the following notations for simplification first. Let \( A_1 \) and \( A_2 \) be the adjusted market size for the economy with \( b \in \{b_1, b_2\} \) and \( b \in \{cb_1, cb_2\} \) respectively. Denote \( T0(b) \equiv T(\bar{\theta}(b)) = T(\bar{\theta}(cb)) \) for each \( b \in \{b_1, b_2\} \).

I prove the FOSD result for the distribution of the number of layers first. The goal
to achieve is that for each $b \in \{b_1, b_2\}$

$$\frac{\bar{\theta}(b)}{\theta_T(b)} > \frac{\bar{\theta}(cb)}{\theta_T(cb)},$$

where $\theta_T$ is the switching point on the quality draw where the firm switches from having $T + 1$ layers to having $T + 2$ layers. Simple calculation shows that for all $b \in \{b_1, b_2\}$:

$$\frac{\theta_T(b)}{\theta_T(cb)} = \left(\frac{A}{c}\right)^\sigma$$

and

$$\frac{\bar{\theta}(b)}{\bar{\theta}(cb)} = A^\sigma\left(\frac{1}{c}\right)^{(\sigma-1)(1-\frac{1}{2T_0(\sigma)})},$$

where $A \equiv \frac{A_2}{A_1}$. As $A > 1$ and $c > 1$, I conclude that $58$

$$\frac{\bar{\theta}(b)}{\theta_T(b)} > 1$$

for all $b \in \{b_1, b_2\}$ which completes the proof.

Third, I prove the FOSD result for the firm size distribution. Similar to the case discussed in the paper, the thing to prove is that for all $b \in \{b_1, b_2\}$

$$S(t\bar{\theta}(b), A_1) \geq S(t\bar{\theta}(cb), A_2)$$

and

$$q(t\bar{\theta}(b), A_1) \geq q(t\bar{\theta}(cb), A_2),$$

where $t > 1$. The rest of the proof follows exactly the proof of Proposition $^8$ in the paper.

$^58$This is because I have already shown that all the exit cutoffs increase when MT improves.
The proof of the final part of this proposition follows the proof of Proposition 8 in the paper as well. Similar to the proof in the paper, I can show that

\[ c > \frac{A_2}{A_1}, \]

which means that all firms increase the output. Moreover, the number of layers also increases, as the output increases. Finally, as the span of control at each layer increases with the output given the number of layers, the span of control is bigger at all layers when MT improves and the firm does not adjust the number of layers. QED.

1.7.16 Supplementary Materials for Section 1.5

I discuss the details of my empirical work in this subsection. First, I describe the data sets that are used in the empirical work. Next, I explain how various variables were constructed and used. Third, I discuss the theoretical foundation for implementing the Olley-Pakes type productivity estimation in the current model. I close this subsection by showing how I implement various robustness checks for the results reported in Tables 1.3 and 1.4.

Two main data sets I use were obtained from the World Management Survey. They were originally used in Bloom and Van Reenen (2010). The first data set is a cross-sectional data set that contains roughly 5700 firms across 16 countries. Each firm was interviewed once between 2004 and 2008, and the score on each management practice (i.e., MT in hits paper) was given after the firm had been interviewed. Most interviews happened after 2006, and 2006 and 2008 are the two years that most interviews were done. The second data set is a panel data set that contains accounting information (e.g., employment, assets and sales etc.) for firms whose management scores were increased.

---

\[59\text{Remember that if } \frac{c}{\tau} = \frac{A_2}{A_1}, \text{ output is unchanged when MT improves.}\]
given in the first data set, and the time span is from 2003 to 2008. I obtain country-level data from website of either the World Bank or the Penn World Table. The Doing Business index (i.e., the distance from frontier), GDP and GDP per capita are obtained from the World Bank. The GDP deflator and the asset formation deflator are obtained from the Penn World Table. The absolute price level of countries is obtained from the Penn World Table as well. The labor market rigidity index is included in the first data set I obtained from the World Management Survey already.

I constructed several variables for the productivity estimation. First, sales are deflated using the GDP deflator. Second, as there is no information on the use of intermediate materials, I used the Olley-Pakes approach to estimate firms’ TFP. More specifically, I constructed a firm’s investment by using the following equation.

\[
inv_t = \frac{assets_t - 0.9 \times assets_{t-1}}{\text{the asset formation deflator}},
\]

where I assume that the value of assets depreciates by 10% annually. Finally, as the sample size is small, I grouped firms in all countries into four “big” sectors based on their SIC codes to implement the productivity estimation.

The data I used to run the regressions are observations in 2006 or 2007 mainly. The main reason is that more than 99% of the interviews were done after 2006, and firms’ management quality might vary from one year to another. Thus, I chose to use accounting information after 2006 only. The second reason is that the Doing Business index (i.e., the distance from frontier) is only available from 2006. Finally, as accounting information is missing for most firms in 2008, only about thirty observations in 2008 can be used to run the regressions.

Now, I discuss the theoretical foundation for implementing the Olley-Pakes type productivity estimation in the current setup. The key to prove is that investment is a monotonically increasing function of the measured TFP conditional on the value of
assets and the number of layers. I consider two types of investment in what follow. The first one is the type of investment that raises up firm’s productivity. This includes R&D investment and expenditure on office automation and softwares etc. Suppose such an investment increases firm’s profitability and incurs a cost. More specifically, let me assume that firm’s operating profit is

\[ \pi_I(A, \theta, T) = \max_I \pi(A, \theta, T) - I - \left( \frac{I}{K} \right)^2, \]

where \( \pi(A, \theta, T) \) is defined in equation (1.15), and \( I \) and \( K \) are investment and value of assets respectively. Conditional on \( K \) and \( T \), optimal investment increases with the measured TFP that is an increasing function of \( A \) and \( \theta \). This is exactly what I want to prove. The second type of investment I consider is the expenditure on capital. Namely, machines and equipments are “consumed” in the production process (i.e., depreciation). The simplest way to establish a positive relationship between this type of investment and firm’s measured TFP is to assume that each unit of output “consumes” a fixed amount of capital (say, \( k \)), and all firms face the same price for capital (say, \( r \)). Under these assumptions, the firms’ MC now becomes the sum of the MC derived in the paper and \( rk \). Therefore, investment in capital is an increasing function of firm’s measured productivity conditional on the number of layers and value of assets. For other production technologies such as the Cobb-Douglas production technology and the CES production technology, the logic to prove the monotonicity is similar. In total, the investment has been shown to be an increasing function of the measured TFP conditional on the number of layers and value of assets which validates the Olley-Pakes type productivity estimation in my model.

Finally, I have done several robustness checks for results reported in Tables 1.3 and 1.4. First, I re-estimated the TFP by grouping firms in each sector into three bins and implementing the Olley-Pakes productivity estimation. After having obtained
the new TFP estimates, I rerun the regressions specified in equations (1.28) and (1.29) and found that regression results are similar to what are reported in Tables 1.3 and 1.4. Namely, the firm-level management score positively affects firm size, ceteris paribus. Furthermore, the average management score negatively affects firm size, ceteris paribus, and the labor market rigidity index positively affects firm size, ceteris paribus. Second, I run the regressions specified in equations (1.28) and (1.29) for firms in each bin separately, as the effect of management quality on firm size might be heterogeneous across firms that have different numbers of layers. The estimation results for both groups of firms are consistent with Proposition 9 as well. Finally, it is possible that the quality of MT might differ across time for a given firm. Thus, I adjusted the data set by only keeping observations whose survey year and account year are the same. When I rerun the regressions for these remaining observations, the estimation results are again consistent with the prediction from Proposition 9. In sum, the above robustness checks confirmed the predictions established in Proposition 9.
References:


Chapter 2

Management Quality, Firm Organization and International Trade

2.1 Introduction

Recent empirical research using firm-level survey data has found that the quality of management technologies differs substantially across countries and has considerable impact on firm performance and organization (e.g. Bloom and Van Reenen (2007, 2010)). Since the 1990s, trade liberalization has accelerated between countries with different levels of management quality. The enactment of the North American Free Trade Agreement (NAFTA) and the creation of World Trade Organization (WTO) are perfect examples of multilateral trade liberalization for economies with different levels of management quality. Despite of the importance of management quality for firm-level outcomes and the increasing speed of globalization, we still know little about how differences in management quality across countries affect the responses of an economy to trade liberalization. More specifically, how country-level management...
quality affects trade volume as well as the welfare gains from trade (henceforth, WGT) needs to be explored in a rigorous general equilibrium environment.

To study these questions, I extend the model developed in Chapter 1 into international context à la Melitz (2003) in this chapter. The key economic insight from the extended model is that trade liberalization (a trade shock) and an improvement in management quality (a management shock) complement each other due to the selection effects provided by both of them. Trade liberalization favors bigger firms, since they export. An improvement in management technology (MT) favors bigger firms as well, as they choose to adopt management hierarchies with more layers. When both shocks are present, two interesting results emerge. First, the increase in trade share (i.e., imports divided by total income) from autarky to a costly trade regime is bigger for economies with better MT. Due to the selection effect of an improvement in MT, firms are bigger and more productive on average in autarky, if the country-level management quality is higher. For a given level of reduction in the variable trade cost, this results in a disproportionate increase in the number of exporting firms relative to the total number of firms. This larger increase in the fraction of exporting firms leads to a larger trade share when the economy opens up to trade. In short, a bigger fraction of firms trade in an economy with better MT. Second, economies with superior MT benefit disproportionately more from opening up to trade under certain conditions\(^1\). In other words, better MT amplifies the WGT under certain conditions\(^2\). The second result is related to the first one, since the bigger increase in the trade share brings more foreign varieties to domestic consumers (the variety effect) and reduces the ideal price index of the CES goods more due to lower prices charged by foreign firms (the productivity effect). In total, the interaction between

\(^1\)In my model, WGT are not guaranteed. Whether or not there are WGT depends on parameters values such as the elasticity of substitution and the quality of MT.

\(^2\)In a different setting, Kambourov (2009) found that institutional features of labor markets affect the responses of an economy to trade liberalization. In particular, higher costs to fire workers dampen WGT.
the trade shock and the management shock leads to systematic changes in aggregate trade variables.

This essay provides evidence to support the firm-level predictions of the model. The internal organization of firms changes after bilateral trade liberalization. More specifically, non-exporting firms flatten their hierarchies by reducing the number of layers and increasing the span of control. Furthermore, non-exporters increase the amount of incentive-based pay after import competition has increased. Employees in non-exporting firms face lower monitoring intensities due to the increasing span of control. Thus, higher incentive-compatible wages are needed to illicit work effort. The above results are consistent with the finding from Guadalupe and Wulf (2010) that American firms facing increasing import competition from Canada flattened their hierarchies and increased the use of incentive-based pay after the enactment of NAFTA.

The current paper also quantitatively evaluates how an improvement in MT affects firm size, aggregate productivity and welfare. In order to estimate model parameters, I calibrate the model to match six moments obtained from the data of the U.S. economic census. Then, I implement two counterfactual experiments by improving the management quality for all firms in the economy. Calibration results show that an improvement in MT has quantitatively important effects on aggregate productivity, firm size and welfare. More specifically, a 22% improvement in MT results in a 22.2% increase in the weighted average of firm productivity, a 72.2% increase in average employment, and a 47.1% increase in welfare. Furthermore, its impact on the WGT is quantitatively sizable as well. The WGT increase by about 1.21% after the improvement in MT. In short, the calibrated model is able to generate quantitatively important effects of an improvement in MT on aggregate-level economic outcomes.

This paper is closely related to the literature that applies efficiency wage models into the international context. Early contributions include Copeland (1989) and Ma-
tusz (1996). More recently, Chen (2011) investigated how the consideration of the efficiency wage affects multinational firms’ organizational choices (i.e., FDI or outsourcing). Davis and Harrigan (2011) used the standard efficiency wage model (i.e., Shapiro and Stiglitz (1984)) to discuss how trade liberalization affects wages of various jobs (i.e., jobs in non-exporting firms and exporting firms etc.) differently. The key departure of this paper from the existing literature is the endogenous determination of the monitoring intensity (i.e., management quality). Firms endogenously determine their monitoring intensities, since they make organizational choices. The endogenous formation of internal firm organization is the key to generating the results on the changes in aggregate trade variables.

The literature on knowledge-based hierarchies has been applied to the international context in the past ten years. It provided us many insights on how international trade affects matching patterns between agents from different countries and how internal firm organization affects the WGT (e.g., Antràs, Garicano and Rossi-Hansberg (2006) and Caliendo and Rossi-Hansberg (2012)). A paper in this literature that is closely related to mine is Caliendo and Rossi-Hansberg (2012). There are two major differences between my paper and Caliendo and Rossi-Hansberg (2012). First, the moral hazard problem that is absent in Caliendo and Rossi-Hansberg (2012) is the key ingredient of this paper. I focus on a different mechanism (effort and incentives), and it yields some different predictions such as what matters for aggregate trade statistics (e.g., the trade share and the WGT) is not the information and communication technology (henceforth, the ICT) but institutions that affect the ability of firms to monitor and incentivize employees. Second, this paper yields some important micro-level predictions that Caliendo and Rossi-Hansberg (2012) did not have such as changes in the pay structure after trade liberalization.

This article is related to the literature on heterogeneous firms and international trade (e.g., Bernard, Eaton, Jensen and Kortum (2003), Melitz (2003), Yeaple (2005)
and Melitz and Ottaviano (2007)). The current paper complements this literature by showing how the organizational structure of the firm and the quality of MT affect the responses of firms and economies to trade liberalization. In particular, the WGT are impacted by the quality of MT which is affected by institutional quality.

This paper is also related a recent literature discussing the WGT. An influential paper by Arkolakis, Costinot and Rodriguez-Clare (henceforth, ACR) argued that the WGT are completely pinned down by two aggregate trade statistics in a number of canonical trade models (e.g., Krugman (1980), Eaton and Kortum (2002), and Melitz (2003)). They are the share of expenditure on domestic goods and the elasticity of trade (i.e., imports divided by domestic consumption) with respect to variable trade costs. They also derived a formula (henceforth, the ACR formula) common across different models to quantitatively evaluate the WGT. Based on these findings, they claimed that although a number of canonical models pointed out different channels through which trade liberalization affects welfare, sufficient statistics for evaluating the WGT are the same across these models. Evaluating the WGT in my framework, however, requires more variables than the two statistics appearing in the ACR formula. This is because one macro-level restriction in the ACR formula is violated in my model. Namely, the share of aggregate profit in total revenue is not a constant. Importantly, I show that information on micro-level variables which are related to the quality of MT is needed for the evaluation of the WGT. This finding shows that explicitly modelling micro-level activities inside the firm matters not only for firm-level outcomes, but for aggregate economic outcomes as well.

The remainder of the paper is organized as follows. Section two extends the model developed in Chapter 1 into the international context à la Melitz (2003) to investigate how the internal organization of the firm responds to bilateral trade liberalization. Section three studies how the quality of MT affects the WGT as well as the trade share by treating the number of layers as a continuous variable. Section four calibrates
the model to quantitatively evaluate how an improvement in MT affects various aggregate economic outcomes including aggregate productivity and the WGT. Section five concludes.

2.2 Trade Liberalization and Firm Organization

In this section, I extend the model developed in Chapter 1 into the international context by considering two symmetric countries. My analysis of opening up to trade in the symmetric two-country case follows Melitz (2003). I make the standard assumption that there is a fixed trade cost denoted by \( f_x \) and a variable trade cost denoted by \( \tau (\geq 1) \) for firms in the CES sector to export. Similar to the fixed production cost, the fixed trade cost is also paid in the form of the final composite good defined in Chapter one. The variable trade cost implies that if \( \tau \) units of output are shipped to the foreign country, only one unit arrives. Furthermore, it is assumed that the fixed trade cost is big enough such that there is selection into exporting in the CES sector. The homogeneous good is not traded regardless of the trade costs, because the two countries are symmetric.

This section focuses on how firms respond to trade liberalization. The analysis is motivated in part by recent empirical evidence on how the internal organization of firms evolves after bilateral trade liberalization. Guadalupe and Wulf (2010) showed that after the enactment of the NAFTA, American firms in sectors with larger reductions in import tariffs flattened their hierarchies. They did so by reducing the number of layers between the chief executive officer (CEO) and division managers and increasing the span of control of the CEO. Furthermore, division managers received more incentive-based pay after the CEO increased the span of control. The extended model presented in this section rationalizes these findings.
In a world of two countries, the firm in the CES sector allocates output between the two markets to equalize its marginal revenues. The optimal allocation of output in the domestic market is

$$q_d = \frac{q(A_H^{\tau \beta})^\sigma}{1 + (A_H^{\tau \beta})^\sigma}, \quad (2.1)$$

where $q$ is the total output, and $A_H$ and $A_F = A_H$ are the adjusted market sizes of the domestic market and the foreign market respectively. For non-exporters, the adjusted market size is $A_H$. For exporters, the adjusted market size is

$$\left(1 + \frac{1}{\tau^{\sigma-1}}\right)^{1/\sigma} A, \quad (2.2)$$

where $A \equiv A_H = A_F$.

The equilibrium conditions in the open economy are similar to those derived in the closed economy. First, the product-market-equilibrium conditions now involve four equations. The equilibrium condition that pins down the cutoff for exporting (i.e., $\bar{\theta}_x$) is

$$\Pi(\bar{\theta}_x, (1 + \frac{1}{\tau^{\sigma-1}})^{1/\sigma} A) - \Pi(\bar{\theta}_x, A) = f_x. \quad (2.3)$$

Note that the cutoff for exporting cannot be solved analytically, since the average cost is endogenously determined and depends on the number of layers the firm has. Next, the FE condition now becomes

$$\int_{\tilde{\theta}_x}^{\bar{\theta}_x} \Pi(\theta, A) g(\theta) d\theta + \int_{\tilde{\theta}_x}^{\infty} \Pi(\theta, (1 + \frac{1}{\tau^{\sigma-1}})^{1/\sigma} A) g(\theta) d\theta = f_e, \quad (2.4)$$

since non-exporters and exporters face different adjusted market sizes. And, the ZCP condition is the same as in the closed economy, or

$$\Pi(\bar{\theta}, A) = 0. \quad (2.5)$$

131
Finally, the market-clearing condition for the homogeneous sector is still given by

$$p_h L_h = (1 - \gamma) E. \quad (2.6)$$

Second, the equilibrium conditions for the labor markets are similar to those derived in the closed economy except that labor demand now contains two parts: one from non-exporting firms and the other one from exporting firms. More specifically, the number of workers who choose to enter the CES sector (i.e., $L_c$) can be derived from the workers’ indifference condition\footnote{For details, see Chapter [1].} or

$$L_c = \int_{\bar{\theta}}^{\infty} T(\theta, A) \sum_{i=1}^{Q(\theta, i)} \frac{M g(\theta)}{1 - G(\theta)} d\theta = \frac{W P(\bar{\theta}, A, M) - \psi L D(\bar{\theta}, A, M)}{p_h}, \quad (2.7)$$

where $W P(\bar{\theta}, A, M)$ is the total wage payment in the CES sector, and $L D(\bar{\theta}, A, M)$ is the number of workers employed in the CES sector, or

$$L D = \int_{\bar{\theta}}^{\infty} T(\theta, A) \sum_{i=1}^{m(\theta, i)} \frac{M g(\theta)}{1 - G(\theta)} d\theta.$$

Note that $L_c$ and $L D$ now consist of two parts and are affected by the trade costs (i.e., $\tau$ and $f_x$). Next, the labor-market-clearing condition indicates that the number of workers employed in the homogeneous sector is the difference between the labor endowment and the number of workers who choose to enter the CES sector, or

$$L_h = L - L_c. \quad (2.8)$$
Finally, the market-clearing condition of the final composite good is modified to

\[
E = \int_{\tilde{\theta}}^{\bar{\theta}} LC(\theta, A) \frac{Mg(\theta)}{1 - G(\theta)} d\theta + \int_{\tilde{\theta}}^{\theta_x} LC(\theta, (1 + \frac{1}{\tau^{\sigma - 1}})^{\frac{1}{\sigma}} A) \frac{Mg(\theta)}{1 - G(\theta)} d\theta
\]

\[
\left[ f_0 + f_x \left( \frac{\tilde{\theta}}{\theta_x} \right)^k + f_e \left( \frac{\tilde{\theta}}{\theta_{min}} \right)^k \right] M + \psi M, 
\]

since exporting firms use the final composite good to pay the fixed trade cost and face a different market size than exporters. The equilibrium in the open economy is characterized by seven equations (i.e., equations (2.3) to (2.9)) and seven endogenous variables (i.e., $\bar{\theta}, \bar{\theta}_x, M, p_h, L_c, L_h$ and $E$). It is easy to prove that there exists a unique equilibrium under restrictions on parameter values using the same approach as the one used in the closed economy. I omit the proof to save space.

I analyze how the opening up of trade affects the internal organization of firms. The key to understanding why the the opening up of trade brings about a differential impact on non-exporters and exporters is the difference in the change of the adjusted market size. The following lemma shows that the adjusted market size shrinks for non-exporters and increases for exporters.

**Lemma 8** When the economy opens up to trade, the adjusted market size faced by non-exporters shrinks, while the adjusted market size faced by exporters increases. Furthermore, the exit cutoff for the quality draw increases.

Proof: See Appendix 2.6.1.

With Lemma 8 in hand, I can analyze how the internal organization of firms and firm productivity change when the economy moves from autarky to the open economy. The main result is that non-exporters flatten their hierarchies by reducing the number of layers and increasing the span of control at existing layers, while exporting firms increase the number of layers and reduce the span of control at all existing layers. The following proposition summarizes these results.
Proposition 12: When the economy opens up to trade, non-exporting firms reduce firm size, while exporting firms increase firm size. Non-exporting firms de-layer weakly and increase the span of control at existing layers when the number of layers is reduced. Exporting firms increase the number of layers weakly and reduce the span of control at existing layers when a new layer is added. Non-exporters increase the amount of incentive-based pay when they de-layer.

Proof: See Appendix 2.6.2.

The effect of bilateral trade liberalization on the internal organization of firms is heterogeneous and depends on the situation firms face. In the theory, non-exporting firms reduce the number of layers and increase the span of control at existing layers due to the shrinking market size after bilateral trade liberalization. This is what Guadalupe and Wulf (2010) found for American firms in industries with increasing import competition. Furthermore, according to the theory, firms increase the incentive-based pay owning to the increasing span of control. This is another finding from Guadalupe and Wulf (2010). Of course, the theory also predicts that firms with increasing opportunities to export (weakly) increase their number of layers and reduce the span of control after bilateral trade liberalization. Guadalupe and Wulf (2010) did find that American firms in sectors with larger reductions in Canadian import tariffs increased the number of layers and reduced the span of control, although these results are not statistically significant. In sum, firms facing different changes in the adjusted market size change their internal organization differently after bilateral trade liberalization.
2.3 Management Quality and Aggregate Trade Variables

In this section, I investigate how management quality affects the trade share and the WGT. I treat the number of layers as a continuous variable à la Keren and Levhari (1979) and Qian (1994) in order to derive analytical results in this section.\footnote{The analysis in all other sections treats the number of layers as a discrete variable.} As I have shown in Chapter 1, the qualitative results of the model at the aggregate-level are unchanged, if the number of layers is treated as a continuous variable instead of a discrete variable. I derive three main results. First, the WGT are not guaranteed due to the existence of multiple frictions in the economy. Second, the trade share is shown to be bigger between economies with better MT. Third, I show that under certain conditions, WGT do exist, and economies with superior MT benefit disproportionately from the opening of trade. Finally, I relate my result to the ACR formula and argue that information on micro-level variables which are related to the management quality is needed for the evaluation of the WGT. In particular, knowing the two aggregate trade statistics that appear in the ACR formula are not enough for us to calculate the WGT.

2.3.1 Equilibrium Outcomes

I calculate the WGT using following several steps. First, optimal allocation of output between two symmetric markets implies that

\[
q_d(\theta) = \frac{q^{\sigma-1}}{1 + \tau^{\sigma-1}} \tag{2.10}
\]

and

\[
p_x(\theta) = \tau p_d(\theta) = \frac{\tau g e}{\beta}, \tag{2.11}
\]
where
\[ ge \equiv b\psi e, \]
and \( p_x(\theta) \) and \( p_d(\theta) \) are prices charged in the foreign market and in the domestic market respectively, if the firm exports.

Second, I solve for the product market equilibrium. First, the indifference condition between exporting or not is
\[
\tau^{\sigma-1}f_x = \frac{A^\sigma}{\sigma} \bar{\theta}_x \left( \frac{\beta_{ge}}{ge} \right)^{\sigma-1}.
\]
(2.12)

Second, the FE condition in the open economy becomes
\[
\int_{\theta=\tilde{\theta}}^{\bar{\theta}} (f - ge) \frac{\theta}{\tau} g(\theta) d\theta + \int_{\theta=\bar{\theta}}^{\infty} (f - ge) \frac{\bar{\theta}}{\theta} \left( 1 + \frac{1}{\tau^{\sigma-1}} \right) \frac{\theta}{\bar{\theta}} g(\theta) d\theta
- (f - ge) \left( \frac{\theta_{\min}}{\theta} \right)^k - f_x \left( \frac{\theta_{\min}}{\theta_x} \right)^k = f_e,
\]
(2.13)

Using equations (2.12) to (2.13) and the ZCP condition in equation (2.5), I solve for the exit cutoff which equals
\[
\bar{\theta} = \theta_{\min} \left( \frac{f - ge}{f_e(k - 1)} \right)^{\frac{1}{k}} \left[ 1 + \left( \frac{f_x \tau^{\sigma-1}}{f - ge} \right)^{\frac{1}{k}} \left( \frac{1}{\tau^{\sigma-1}} \right) \right]^{-\frac{1}{k}}.
\]
(2.14)

And, the exporting cutoff is
\[
\bar{\theta}_x = \bar{\theta} \frac{f_x \tau^{\sigma-1}}{(f - ge)}.
\]
(2.15)

Substituting equation (2.15) into equation (2.12) yields the solution for the adjusted market size which is equal to
\[
A^\sigma = \frac{(f - ge)(ge)^{\sigma-1}}{(1 - \beta)\beta^{\sigma-1} \theta} = \frac{(f - ge)^{1-\frac{1}{k}}(ge)^{\sigma-1}((k - 1)f_e)^{\frac{k}{2}}}{(1 - \beta)\beta^{\sigma-1} \theta_{\min}} \left[ 1 + \left( \frac{f_x \tau^{\sigma-1}}{f - ge} \right)^{\frac{1}{k}} \left( \frac{1}{\tau^{\sigma-1}} \right) \right]^{-\frac{1}{k}}.
\]
(2.16)
Third, I solve for the welfare (i.e., the workers’ expected payoff). Using the labor-market-clearing conditions in equations (2.7) and (2.8), I restate the labor-market-clearing condition as

\[
\psi(b - 1) \left[ \frac{k}{k - 1} \frac{(\sigma - 1)(f - ge)}{ge} \left[ 1 + \tau^{1-\sigma} \left( \frac{f_x \tau^{\sigma-1}}{(f - ge)} \right)^{-k+1} \right] - 1 \right] (ge)^{-\sigma-1} \\
\frac{(k - 1)}{k \theta_{\text{min}}} \left( \frac{f_x (k - 1)}{f - ge} \right) \frac{1}{k} \left[ 1 + \left( \frac{f_x \tau^{\sigma-1}}{f - ge} \right)^{-k+1} \right]^{-\frac{1}{k}} \left[ 1 + \tau^{1-\sigma} \left( \frac{f_x \tau^{\sigma-1}}{(f - ge)} \right)^{-k+1} \right]^{-1} \\
+ \frac{(1 - \gamma)}{\gamma} \left( f - ge \right)^{1-\frac{1}{k}} (ge)^{\sigma-1} \left( \frac{1}{1 - \beta} \right) \beta^{\sigma-1} \theta_{\text{min}} \left[ 1 + \left( \frac{f_x \tau^{\sigma-1}}{f - ge} \right)^{-k+1} \right]^{-\frac{1}{k}} \\
= p_h \frac{1}{\gamma \beta^{\sigma-1}} L, \quad (2.17)
\]

where the first part of the left hand side (LHS) of the above equation is related to the number of job applicants in the CES sector, while the second part of the LHS of the above equation is related to the number of workers in the homogeneous sector. In total, it states that the number of workers who seek jobs is equal to the fixed labor endowment, \( L \). Solving equation (2.17), I obtain the welfare in the open economy as

\[
p_h^0 = \left[ \frac{f_x (k - 1)}{\beta^{\sigma-1} \theta_{\text{min}} L} \right] \frac{\gamma}{1 - \gamma} (ge)^{\sigma-1}, \quad (2.18)
\]

where

\[
A(b, \tau) \equiv \psi(b - 1) \left[ \frac{k}{k - 1} \frac{(\sigma - 1)(f - ge)}{ge} \left[ 1 + \tau^{1-\sigma} \left( \frac{f_x \tau^{\sigma-1}}{(f - ge)} \right)^{-k+1} \right] - 1 \right] (ge)^{-\sigma-1} \\
\frac{(k - 1)}{k} \left( \frac{1}{f - ge} \right) \frac{1}{k} \left[ 1 + \left( \frac{f_x \tau^{\sigma-1}}{f - ge} \right)^{-k+1} \right]^{-\frac{1}{k}} \left[ 1 + \tau^{1-\sigma} \left( \frac{f_x \tau^{\sigma-1}}{(f - ge)} \right)^{-k+1} \right]^{-1} \\
+ \frac{(1 - \gamma)}{\gamma} (f - ge)^{1-\frac{1}{k}} (ge)^{\sigma-1} \left[ 1 + \left( \frac{f_x \tau^{\sigma-1}}{f - ge} \right)^{-k+1} \right]^{-\frac{1}{k}}. \quad (2.19)
\]
Therefore, the change in welfare is
\[ WGT(b, \tau) \equiv \frac{p_h^g}{p_h^g} - 1 = \left[ \frac{A(b, \tau)}{B(b)} \right]^{\frac{\gamma}{(1-\gamma)(\sigma-1)}} - 1. \tag{2.20} \]

Note that WGT are not guaranteed. However, I am going to derive some sufficient conditions assuring WGT in next subsection.

Before proceeding to the discussion of WGT, I discuss the condition under which the type of equilibrium I investigate exists. The equilibrium exists and is unique if
\[ p_h^g \leq \psi(be - 1), \]
where \( p_h^g \) is defined in equation \( (2.17) \). Note that the above condition only contains exogenous parameters (i.e., primitives of the model). Furthermore, this condition puts an upper bound on \( L \) when \( \frac{\sigma-1}{\sigma} > \gamma \) and a lower bound on \( L \) when \( \frac{\sigma-1}{\sigma} < \gamma \).

### 2.3.2 Management Quality, the Trade Share, and the Welfare Gains from Trade

In this subsection, I show how management quality affects the trade share and discuss how it affects the WGT. I discuss the relationship between management quality and the trade share first. Let \( \lambda \) be the share of domestic consumption in total expenditure on the CES goods. Using equations \( (2.10)-(2.11) \) and \( (2.14)-(2.16) \), I derive the domestic consumption share as
\[ \lambda(\tau, b) = \frac{\tau^{\sigma-1} \left( \frac{\theta}{\theta} \right)^{k-1}}{1 + \tau^{\sigma-1} \left( \frac{\theta}{\sigma} \right)^{k-1}}. \tag{2.21} \]

\(^5\)I use \( WGT(b, \tau) \) to denote the welfare change from opening up to trade to save notations. For some parameter values, \( WGT(b, \tau) \) might be negative, which implies welfare losses from trade.

138
and the import share as

\[
1 - \lambda(\tau, b) = \frac{1}{1 + \tau^{\sigma-1}\left(\frac{\theta}{\sigma}\right)^{k-1}.}
\]  

(2.22)

Since \(\frac{t}{b}\) increases in \(b\) from equation (2.15), the better the MT, the bigger the share of exporting firms. Furthermore, equation (2.22) shows that the better the MT, the bigger the trade share. In economies with better MT, the exit cutoff for the demand draw is higher when these economies are in autarky as a result of the selection effect. This leads to the result that active firms in these economies are bigger and more productive when these economies are in autarky. Since it is the bigger firms that start to export when the economies move from autarky to open economies, disproportionately more firms export in economies with better MT.

Next, I derive sufficient and necessary conditions under which there are WGT. Before moving on, I make the following simplifying notations.

\[
t \equiv ge = b\psi e;
\]

\[
C(t, \tau) \equiv \frac{f_x^{-k+1}T^{-k(\sigma-1)}}{(f-ge)^{k+1}}.
\]

Note that \(C(t, \tau, f_x)\) decreases with \(t, \tau\) and \(f_x\). Calculation shows that

\[
\frac{dA(b, \tau)}{d\tau} < 0,
\]

if and only if

\[
Q(\sigma, \gamma, f_x, \psi, \tau) = \frac{(\sigma - 1)(f - t)}{t} + \frac{(1 - \gamma)\sigma f - t}{\gamma t - \psi} - \frac{k^2 - 1}{k(1 + C(t, \tau))} > 0. \quad (2.23)
\]

\footnote{Note that this result still holds if I define trade share as imports divided by total income, since expenditure on the CES goods is a constant fraction of total income.}
Therefore, there are WGT (i.e., $WGT(b, \tau) > 0$), if and only if one of the following two conditions holds:

$$Q(\sigma, \gamma, f_x, f, \tau, \psi) > 0, \quad \frac{\sigma - 1}{\sigma} > \gamma$$

or

$$Q(\sigma, \gamma, f_x, f, \tau, \psi) < 0, \quad \frac{\sigma - 1}{\sigma} < \gamma.$$  

I will discuss the economic interpretation of these conditions in what follows.

Since the above conditions are not intuitive, I derive sufficient conditions that guarantees WGT. I focus on the case in which $\frac{\sigma - 1}{\sigma} > \gamma$ first. In this case, the inequality in equation (2.23) holds for all possible values of $k$ and $\tau$ if

$$f > \left[1 + \frac{(\sigma - 1)k}{\sigma^2 - \sigma + 1}\right]t = \left[1 + \frac{(\sigma - 1)k}{\sigma^2 - \sigma + 1}\right]\psi eb. \quad (2.24)$$

The above condition indicates that when the elasticity of substitution is big, there are WGT if the management quality is high. When the iceberg trade cost goes down, it induces resource reallocation between sectors. When the elasticity of substitution (i.e., $\sigma$) is high, resources are reallocated from the homogeneous sector to the CES sector. There are two countervailing forces that cause this result. First, average firm size in the CES sector increases due to a reduction in the iceberg trade cost. Second, the mass of active firms in the CES sector decreases when the iceberg trade cost goes down. The relative strength of these two forces depends on the elasticity of substitution (and the relative size of the CES sector). When the elasticity is high (i.e., smaller market power in the CES sector), the increase in average firm size is bigger, and the decrease in the mass of active firms is smaller. Therefore, resources are reallocated from the homogeneous sector to the CES sector after a reduction in the iceberg trade cost. I.e., the monopolistically competitive sector expands as a result
of the trade liberalization. When the management quality of firms in the CES sector is high, resources that are reallocated into the CES sector are used efficiently, which ensures WGT. Note that the condition in equation (2.24) is a sufficient condition for the existence of WGT for all possible levels of the iceberg trade cost. In other words, there might be welfare losses for certain reductions in the iceberg trade cost, if the condition in equation (2.24) is not satisfied.

In the case in which \( \frac{\sigma - 1}{\sigma} \leq \gamma \), a sufficient condition for the existence of the WGT is

\[
\psi_{eb} \left[ 1 + \frac{k^2 - 1}{2k^2} \frac{\sigma - 1}{\sigma^2 - \sigma + 1} \right] > \frac{k^2 - 1}{2k^2} \frac{\sigma - 1}{\sigma^2 - \sigma + 1} \psi + f. \tag{2.25}
\]

The above condition says that when the elasticity of substitution is small, there are WGT if the management quality of firms in the CES sector is low. The economic reasoning follows the same logic that I have discussed above. Namely, the increase in average firm size is dominated by the decrease in the mass of active firms in the CES sector, when the iceberg trade cost falls and the elasticity of substitution is small. As a result, resources are reallocated from the CES sector to the homogeneous sector after a reduction in the iceberg trade cost. When the management quality of firms in the CES sector is low, resources that are reallocated away from the CES sector are used inefficiently before the trade liberalization. Therefore, there are WGT when these resources are reallocated to the homogeneous sector after the trade shock. In summary, WGT are not guaranteed in a world with multiple frictions. In particular, whether or not there are WGT crucially depends on whether or not resources that are reallocated after the trade liberalization are used inefficiently before the trade liberalization.

Now, I show that under one of the above sufficient conditions (i.e., equation (2.24) and \( \frac{\sigma - 1}{\sigma} > \gamma \)), the complementarity result holds. Namely, the better is the MT, the larger are the WGT. I focus on this condition, since the parameters of the calibrated

\footnote{Remember that there is no incentive problem for firms in the homogeneous sector.}
model presented in the next section satisfy this condition. Recall that the WGT are

\[ WGT(b, \tau) = \frac{p_h^b}{p_h^b} - 1 = \left[ \frac{B(b)}{A(b, \tau)} \right]^{\gamma} - 1. \]

Denote

\[ D(b, \tau) \equiv \frac{B(b)}{A(b, \tau)}. \]

As long as \( \frac{d \ln D(b)}{db} < 0 \), the complementarity result holds when \( \frac{\sigma - 1}{\sigma} > \gamma \). Calculation shows that

\[ \frac{d \ln D(b)}{db} < 0, \]

since \( t^2 + \psi(f - 2t) > 0, C(t, \tau) > 0 \) and

\[ f > \left[ 1 + \frac{(\sigma - 1)k}{\sigma^2 - \sigma + 1} \right] t. \]

Therefore, there are the WGT, and better MT makes the WGT larger if

\[ f > \left[ 1 + \frac{(\sigma - 1)k}{\sigma^2 - \sigma + 1} \right] t, \quad \frac{\sigma - 1}{\sigma} > \gamma. \]

### 2.3.3 A Comparison to the ACR Formula

In an influential paper, Arkolakis, Costinot and Rodriguez-Clare showed that the WGT are completely determined by two aggregate statistics in a number of canonical trade model (e.g., Krugman (1980), Eaton and Kortum (2002) and Melitz (2003)). They are the share of expenditure on domestic goods and the elasticity of trade with respect to variable trade costs. In this subsection, I show that the model presented in this paper does not fall into the set of models whose predictions for the WGT are completely characterized by the above two statistics. Furthermore, I show that micro-level information which is related to the management quality is needed for the evaluation of the WGT.
Among the two aggregate statistics appearing in the ACR formula, one statistic (i.e., the domestic consumption share) is derived in equation (2.21). The other one which is the elasticity of the trade share with respect to the variable trade cost is defined as
\[
\epsilon = \frac{\partial \ln (1 - \lambda(\tau,b))/\lambda(\tau,b)}{\partial \ln \tau},
\]
and its value is
\[
\epsilon = -(\sigma - 1)k. \tag{2.26}
\]
Note that this elasticity is constant in the case of the continuous number of layers and does not depend on the management quality. However, this specific result is valid, only when I treat the number of layers as a continuous variable.

Now, I evaluate the WGT. Calculation shows that the WGT are
\[
WGT(b,\tau) = \lambda(\tau,b)\frac{\gamma}{\kappa(\tau-(\gamma-1)(\sigma-1))} \left[ \frac{1 + \frac{(1-\gamma)\sigma}{\gamma(\sigma-1)(\omega-1)}}{1 + \frac{(1-\gamma)\sigma}{\gamma(\sigma-1)(\omega-1)}} x_T(\bar{\theta}) - \frac{k-1}{k} \lambda(\tau,b) \right] \left[ \frac{1 + \frac{(1-\gamma)\sigma}{\gamma(\sigma-1)(\omega-1)}}{1 + \frac{(1-\gamma)\sigma}{\gamma(\sigma-1)(\omega-1)}} x_T(\bar{\theta}) - \frac{k-1}{k} \right], \tag{2.27}
\]
where
\[
x_T(\bar{\theta}) = \frac{(\sigma - 1)(f - t)}{t} \tag{2.28}
\]
is the output level of the smallest firms in equilibrium. Note that this variable is unit-free, since all variables in the model are denominated in terms of the final composite good. Equation (2.28) is a function of ratios as well, since \(f/t\) is the ratio of the fixed production cost to the equilibrium wage offered in the CES sector. Thus, the value of equation (2.28) does not depend on the units that are used to measure the output in equilibrium.

Again, I need information on the expenditure share on the CES goods (i.e., \(\gamma\)), the management quality (i.e., \(\frac{1}{b}\)) and the output level of the smallest firms (i.e., \(x_T(\bar{\theta})\)) in order to evaluate the welfare change due to opening up to trade. In an extreme
In a world with information frictions which validate the use of management hierarchies, evaluating the WGT requires more information than that contained in the ACR formula. Of course, this does not mean that the ACR formula is incorrect, since one restriction of the ACR formula that aggregate profit is a constant fraction of total income is violated in my framework. The above result only shows that micro-level factors such as the management quality should be taken into account, even if we only care about changes in aggregate economic outcomes.

2.4 Calibration

In this section, I calibrate the model in order to implement counterfactual experiments. The counterfactual experiments show that an improvement in MT has quantitatively important effects on aggregate productivity, firm size and welfare. Furthermore, its impact on the WGT is quantitatively sizable.

I choose to match six moments from the data in order to estimate six parameters: $b\psi$, $k$, $f_e$, $f$, $f_x$, and $\gamma$. The first four moments are moments that are not related to international trade. They are the average employment of firms, the fraction of firms that have less than ten employees, the fraction of firms that have more than five hundred employees, and the Pareto shape parameter obtained from the regression of
The Pareto shape for the firm size distribution is obtained from the regression of $\ln(Pr(employment > x))$ on $\ln(employment) = x$, where $Pr(employment > x)$ is the share of firms that have employment more than $x$.

No information is available for this moment in 2007.

In other words, I use the identity matrix as the weighting matrix.
identification check, I obtained values for two additional moments that are closely related to the internal organization of firms to check whether the calibrated model does a good job at matching these two moments as well. First, the U.S. economic census data shows that wage share of production workers was 57.7% in 2007. As Table 2.3 shows, the calibrated model predicts that this value is 52.1% which is close to the data. \(^{11}\) Second, Guadalupe and Wulf (2010) reported that the average number of managerial layers was 3.1 – 3.2 for various industries in 1999 in a sample containing 230 large U.S. manufacturing firms. \(^{12}\) Although I use data on the entire population of manufacturing firms in 2007, it is interesting to compare the characteristics of firms in their sample relative to what my model yields. The calibrated model predicts that the average number of managerial layers is 3.44 which is close to what Guadalupe and Wulf (2010) have found. In summary, the difference between the moments obtained from the data and the moments generated by the calibrated model is small.

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pareto Shape Parameter</td>
<td>1.095</td>
<td>1.097</td>
</tr>
<tr>
<td>Employment per firm</td>
<td>46.57</td>
<td>46.57</td>
</tr>
<tr>
<td>Fraction of small firms (less than 10 employees)</td>
<td>55.04%</td>
<td>55.14%</td>
</tr>
<tr>
<td>Fraction of big firms (more than 500 employees)</td>
<td>1.03%</td>
<td>0.76%</td>
</tr>
<tr>
<td>Fraction of exporting firms</td>
<td>18%</td>
<td>17.84%</td>
</tr>
<tr>
<td>Trade Share (relative to GDP)</td>
<td>14%</td>
<td>14.00%</td>
</tr>
</tbody>
</table>

The Pareto shape parameter is obtained from the regression of log rank on log firm size.

Table 2.1: Moments from the Data and the Model

I implement two counterfactual experiments to investigate the quantitative impact of an improvement in MT on firm size, aggregate productivity and the WGT. The

\(^{11}\) Although non-production workers only monitor in my model, it is straightforward to extend the model into the case in which non-production workers both monitor and exert effort to increase firm productivity under certain assumptions. If I assume that monitoring requires only time and no effort as in Qian (1994), the extend model yields exactly the same firm-level outcomes and resource allocation as the current model. Therefore, the value of this moment in the data is comparable to the value of this moment generated by the model.

\(^{12}\) The number of managerial layers is the number of layers ranging from the CEO to division managers who supervise the lowest production units. In the model, it is measured as the number of layers ranging from layer \(T - 1\) to layer 0.
Table 2.2: Parameter Values

<table>
<thead>
<tr>
<th>Value</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$</td>
<td>4</td>
</tr>
<tr>
<td>$\tau$</td>
<td>1.3</td>
</tr>
<tr>
<td>$f$</td>
<td>4.98</td>
</tr>
<tr>
<td>$f_e$</td>
<td>3.00</td>
</tr>
<tr>
<td>$f_x$</td>
<td>4.84</td>
</tr>
<tr>
<td>$k$</td>
<td>1.188</td>
</tr>
<tr>
<td>$b\psi$</td>
<td>0.649</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.545</td>
</tr>
</tbody>
</table>

Table 2.3: Moments Related to Internal Firm Organization

<table>
<thead>
<tr>
<th>The number of managerial layers</th>
<th>Data</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.1-3.2</td>
<td>3.44</td>
</tr>
<tr>
<td>Wage share of production workers</td>
<td>57.7%</td>
<td>52.1%</td>
</tr>
</tbody>
</table>

average score of Indian firms on monitoring and incentive practices is about 22% lower than American firms in the World Management Survey (WMS). I consider a similar percentage change in the management quality in the model. More specifically, I consider a comparative statistics in which the value of $b$ decreases from $2.141 (=1.755 \times 1.22)$ (India) to $1.755$ (the U.S.). Remember that the management quality is denoted by $\frac{1}{b}$. Thus, such an improvement implies a 22% improvement in MT. The first quantitative exercise is to evaluate how such an improvement affects firm size, productivity and welfare. Table 2.4 shows that average employment and sales increase by 72.2% and 36.9% as a result of this improvement. Furthermore, weighted average firm productivity and welfare increase by 22.2% and 47.1% respectively. The second quantitative exercise is to evaluate how the deterioration in MT affects the trade share and the WGT. Table 2.5 reports the result. The improvement in MT yields a 1.21% increase in the WGT which is about 7.4% of the WGT before the improvement. This improvement also generates a 5.31% increase in the fraction of
exporting firms. In total, the calibrated model is able to generate a quantitatively sizable impact of an improvement in MT on firm size, productivity and the WGT.

### 2.5 Concluding Remarks

In this chapter, I extend the baseline presented in Chapter 1 to an international context to study how an improvement in MT and trade liberalization jointly affect firm-level as well as aggregate-level economic outcomes. At the firm level, trade liberalization brings changes to the internal organization of firms as well as pay structures. At the aggregate level, an improvement in MT amplifies the response of economies to trade liberalization by increasing the trade share and the WGT. In total, the interplay between trade liberalization and the improvement in MT yields systematic changes in the firm-level outcomes and the aggregate trade variables.

Undoubtedly, much more research remains to be done. First, it is important to consider two asymmetric countries that differ in MT and investigate how such a difference affects trade patterns, trade volume and the WGT. Second, although MT considered in this chapter is exogenous and invariant across firms, recent evidence suggests that MT does differ across firms and is affected by shocks such as trade liberalization (Bloom and Van Reenen (2010) and Bloom, Sadun, and Van Reenen (2012b)). Therefore, it is worth exploring how trade liberalization endogenously affects MT which in turn impacts aggregate economic variables.

### Table 2.4: Firm Size, Aggregate Productivity and Welfare

<table>
<thead>
<tr>
<th></th>
<th>Prod&lt;sub&gt;ω&lt;/sub&gt;</th>
<th>Employment</th>
<th>Sales</th>
<th>Welfare</th>
</tr>
</thead>
<tbody>
<tr>
<td>M&amp;I ↑ 22%</td>
<td>22.2%</td>
<td>72.2%</td>
<td>36.9%</td>
<td>47.1%</td>
</tr>
</tbody>
</table>

*Prod<sub>ω</sub>: percentage change in weighed average of firm productivity

*Employment*: percentage change in average employment

*Sales*: percentage change in average sales

*Welfare*: percentage change in welfare
Table 2.5: Trade Share and the Welfare Gains from Trade

<table>
<thead>
<tr>
<th></th>
<th>b=2.141</th>
<th>b=1.755</th>
<th>Complementarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>WGT</td>
<td>16.42%</td>
<td>17.63%</td>
<td>1.21%</td>
</tr>
<tr>
<td>Share of exporting firms</td>
<td>12.53%</td>
<td>17.84%</td>
<td>5.31%</td>
</tr>
</tbody>
</table>

2.6 Appendix

2.6.1 Proof of Lemma 8

I make several notations before the proof. Let $A_0$ be the adjusted market size for all firms in the closed economy. Denote the adjusted market size faced by non-exporters and exporters in the open economy by $A_1$ and $A_2 (> A_1)$ respectively. As countries are symmetric, I only need to discuss what happens to domestic firms. First, note that what matters for the firm’s profit is the adjusted market size $A$. Second, suppose that the exit cutoff stayed the same or decreased after the economy opens up to trade. This would immediately imply that $A_2 > A_1 ≥ A_0$ due to the ZCP condition and the result that firms on the exit cutoff are non-exporting firm. In other words, all surviving firms are facing the bigger adjusted market size than before. However, this result together with the FE condition would imply that the exit cutoff goes up which contradicts the assumption that the exit cutoff either stayed the same or decreased. Therefore, the exit cutoff must increase after the economy opens up to trade. Third, due to this result, the adjusted market size faced by non-exporters must go down when the economy moves from autarky to trade (i.e., $A_0 > A_1$). Finally, suppose that the adjusted market size faced by exporters also went down weakly (i.e., $(A_1 <)A_2 ≤ A_0$). This would imply that both exporters and non-exporters lose in the open economy. This result and the result that the exit cutoff for the quality draw increases when the economy opens up to trade would violate the FE condition. Therefore, it must be the case that $A_2 > A_0 > A_1$ in equilibrium. This means that exporters gain and non-exporters lose when the economy moves from autarky to trade. QED
2.6.2 Proof of Proposition 12

The firm’s optimization problem defined implies that an increase (or a decrease) in $A$ has the same effect on firm-level outcomes (i.e., revenue, employment, and output) as an increase (or a decrease) in $\theta^{\frac{1}{\sigma}}$. It is straightforward to observe that firm’s revenue, employment, and output increase in $\theta$, and Lemma 8 shows that the adjusted market size increases for exporters and decreases for non-exporters after the economy opens up to trade. Therefore, non-exporters’ revenue, employment, and output go down, and exporters’ revenue, employment, and output go up after the economy opens up to trade. As Propositions proved in Chapter 1 show that the number of layers is an increasing function of output, non-exporters de-layers and exporters increase their number of layers after the economy opens up to trade. Propositions proved in Chapter 1 show that the firm increases the span of control when it adds a layer and decreases the span of control when it deletes a layer. Therefore, non-exporters that de-layer increase the span of control at the same time, while exporters that add a layer decrease the span of control. Finally, as wage payment in my paper is incentive-based and increases with the span of control, non-exporters increase the use of incentive-based pay when they de-layer. QED.
References:


Chapter 3

Trade Liberalization and
Within-Firm Productivity Gains

3.1 Introduction

Recent empirical research studying the impact of trade liberalization on firm productivity found substantial productivity gains coming from within-firm improvements following trade liberalization. For instance, Pavnick (2002) documents that one-third of the overall productivity gains after trade liberalization in the case of Chile came from within-firm productivity gains. Moreover, she showed that these within-firm productivity gains were larger for import-competing sectors due to improved managerial quality and reduction in “X” inefficiency. Another example is Schmitz (2005), who documents large within-firm productivity gains for the U.S. iron ore industry after the enactment of a free-trade agreement between the U.S. and Brazil in the 1980s. These findings do not seem to square well with a recent literature that emphasizes the importance of improved market access for productivity improvements at the firm level (e.g., Lileeva and Trefler (2010) and Bustos (2011)), since firms do not seem to have better market access after import competition intensifies. On the contrary,
firms seem to face shrinking market size when foreign competitors arrive. Therefore, a question to ask is why firm productivity increases when firms face tougher competition. The purpose of this paper is to point out a channel through which being in adversity incentivizes managers to exert more effort to increase firm productivity.

Following the tradition dating back to Berle and Means (1932), I open the black box of the firm and treat the separation of ownership and control as the fundamental problem within the firm. An investor has enough resources to form a firm and a rough idea to start a business. However, she needs to be matched with a manager who has knowledge and experience to make this rough idea implementable. The overall quality of an implementable idea depends on two components. First, after the investor meets and discusses the rough idea with the manager, an initial quality of the implementable idea is randomly realized. Second, the manager has to exert effort to develop this implementable idea after the initial quality is realized, which affects the overall quality of the implementable idea as well. In the end, the overall quality of the implementable idea pins down the efficiency (i.e., the cost) of producing the product generated by the implementable idea, which eventually determines firm productivity.

In this paper, I propose a general equilibrium model consisting of one industry. The industry is populated by firms that produce differentiated products with a constant elasticity of substitution (CES) under conditions of monopolistic competition à la Dixit and Stiglitz (1977). There is a large pool of investors who can enter this industry by paying a fixed cost. There is also a large pool of managers who can be matched with the investors to form firms. The timing of the game is as follows. First, an investor can enter the industry by paying a fixed entry cost, and then she is randomly matched with a manager. After the match, a firm is set up, and the initial quality of an implementable idea generated by these two agents is randomly realized. Second, the manager makes his occupational choice. He can quit the firm...
and become a worker, in which case the investor receives zero profit afterwards. Or he can choose to work for the investor and exert effort to develop the implementable idea, which leads to a blueprint for a product. 第三，如果经理选择为投资者工作并付出努力，投资者决定是否支付固定生产成本以启动生产。第四，如果生产开始，经理决定输出水平和雇用的工人数。然后，公司在市场上竞争，公司收入和经营利润得到，变量成本得到支付。最终，投资者和经理就经营利润进行讨价还价以获得收入。为了简单起见，我假设他们玩一个一般化的纳什讨价还价游戏。

根据不完全合同的文献（例如，格罗斯曼和哈特（1986）和哈特和摩尔（1990）），我认为经理的努力是可观测但不可验证的，而且经理和投资者通过事后讨价还价获取收入。这种假设意味着投资者观察到可实施想法的整体质量，尽管她无法向法庭展示经理付出的努力的数量。因此，根据努力制定合同是不可能的。此外，两个代理获得的经营利润的份额是通过事后讨价还价解决的。我用 $\alpha$（对于经理）和 $1 - \alpha$（对于投资者）来表示这两个份额。

在不完全合同的文献（例如，格罗斯曼和哈特（1986）和哈特和摩尔（1990）），我假设努力由经理 observable 但 not verifiable，而且经理和投资者通过事后讨价还价获得收入。这种假设意味着投资者观察到可实施想法的整体质量，尽管她无法向法庭展示经理付出的努力的数量。因此，根据努力制定合同是不可能的。此外，两个代理获得的经营利润的份额是通过事后讨价还价解决的。我用 $\alpha$（对于经理）和 $1 - \alpha$（对于投资者）来表示这两个份额。

在不完全合同的文献（例如，格罗斯曼和哈特（1986）和哈特和摩尔（1990）），我假设努力由经理 observable 但 not verifiable，而且经理和投资者通过事后讨价还价获得收入。这种假设意味着投资者观察到可实施想法的整体质量，尽管她无法向法庭展示经理付出的努力的数量。因此，根据努力制定合同是不可能的。此外，两个代理获得的经营利润的份额是通过事后讨价还价解决的。我用 $\alpha$（对于经理）和 $1 - \alpha$（对于投资者）来表示这两个份额。

How do the manager and the investor make their decisions in equilibrium? I use backward induction to solve the equilibrium. At the fourth stage, the choice of output is to maximize the operating profit, since the manager and the investor bargain over the operating profit at the final stage. At the third stage, the investor is willing to

---

1. In order to save notations, I refer to managers who choose to become workers at stage two as “managers” in subsequent analysis.
2. The operating profit equals revenue minus the variable cost.
3. The two shares are determined in the following way. Assume that the manager and the investor receive $\alpha_1$ and $\alpha_2$ ($\alpha_1 + \alpha_2 \leq 1$) fractions of the operating profit if the bargaining breaks down. The bargaining power of the manager (and the investor) is assumed to be $\alpha_3$ (and $1 - \alpha_3$). Therefore, the shares of the operating profit received by the manager and the investor are $\alpha = \alpha_1 + \alpha_3(1 - \alpha_1 - \alpha_2)$ and $1 - \alpha = \alpha_2 + (1 - \alpha_3)(1 - \alpha_1 - \alpha_2)$, respectively.
start production, if and only if the fraction of operating profit she receives from the ex post bargaining at least covers the fixed production cost. The manager’s decision at the second stage is more involved. I discuss it case by case. First, if the initial quality of the implementable idea is low, the optimal choice of the manager is to quit the firm and become a worker. In this case, the manager cannot find an effort level that leads to a high enough operating profit to both compensate the investor (for the fixed production cost) and give the manager himself a high enough payoff. Thus, the manager chooses to become a worker and receives a wage, which is his outside option. Second, if the initial quality of the implementable idea is high, the manager chooses an effort level to maximize the fraction of operating profit he receives from the ex post bargaining minus the effort cost. Obviously, there is under-provision of the effort in this case, since the manager takes into account only a fraction of the operating profit when making his effort choice. However, the investor is still willing to start production under the second-best level of effort since the high initial quality of the implementable idea leads to a high enough operating profit. Third, if the initial quality of the implementable idea is in the middle range, the manager chooses an effort level that is higher than the second-best level to induce his investor to start production. Since the initial quality is not high, the investor would not start production if the manager exerted effort at the second-best level. However, there is room for both agents to achieve a Pareto improvement. In equilibrium, the manager exerts effort at a level higher than the second-best level, and the profit the investor receives under this higher level of effort exactly equals the fixed production cost. As a result, the investor is willing to produce, and the manager earns a payoff higher than his outside option. At the first stage, the investor enters the monopolistically

---

4The second-best level of effort is defined as the one that maximizes the profit the manager receives from the ex post bargaining minus the cost of exerting effort given that the production is carried out.

5Note that the manager has no incentive to exert effort at a level under which the profit the investor receives is larger than the fixed production cost in this case. This is because the purpose of
competitive industry if and only if the expected profit from entering is non-negative. Since there is a large pool of investors who can enter the industry, the expected profit from entering is zero in equilibrium.

The relationship between the initial quality of the implementable idea and the manager’s optimal effort choice is non-monotonic. First, there is no effort exerted by the manager when the initial quality is low, since the manager chooses to become a worker in this case. Second, when the initial quality of the implementable idea is high, the optimal effort increases with the initial quality. This is because a higher initial quality increases the marginal return to exerting effort. Finally, when the initial quality of the implementable idea is in the middle range, the optimal effort decreases with the initial quality. This is because a higher initial quality coupled with a lower effort level can induce the investor to produce. In total, the relationship between the initial quality of the implementable idea and the optimal effort level is “U” shaped.

In this paper, firm productivity consists of two parts: the initial quality of the implementable idea and the manager’s effort to develop the idea. Two points are worth noting about firm productivity. First, firm productivity increases weakly with the initial quality. When the initial quality of the implementable idea is in the middle range, firm productivity is constant. This is because the level of firm productivity under which the investor breaks even is independent of the initial quality. When the initial quality is high, firm productivity increases with the initial quality of the implementable idea, since the optimal effort increases with it. Second, since the initial quality of the implementable idea is exogenous from the perspective of the firm and fixed throughout the paper, changes in the manager’s effort are directly translated into changes in firm productivity.

I extend the model described above into an international context à la Melitz (2003) to study how opening up to trade affects the manager’s effort choice as well as deviating from the second-best level of effort is to induce the investor to produce. Thus, any further division from the effort level under which the investor breaks even is suboptimal for the manager.
firm productivity. Opening up to trade triggers within-industry resource reallocation and generates productivity gains for two types of firms. First, the least productive non-exporting firms exit the market. This is due to the selection effect (see Melitz (2003)). Second, productivities of the least productive surviving firms (i.e., non-exporters) increase after opening up to trade, even though market size faced by these firms shrinks. After opening up to trade, managers of the least productive surviving non-exporters are incentivized to exert more effort to improve their firms’ productivity levels. The minimum level of firm productivity (i.e., the overall quality of the idea) needed to induce the investor to produce goes up after opening up to trade. The manager has to exert more effort, if he still wants to incentivize the investor to produce. This disciplining effect on the least productive surviving non-exporting firms is a new channel through which opening up to trade affects firm productivity, and it is the main result of this paper. Finally, productivity levels of the smallest exporters are also higher in the open economy than in the closed economy, for two reasons. First, enlarged market size faced by exporters increases the marginal return to exerting effort. As a result, the second-best level of effort is larger in the open economy than in the closed economy. Second, managers of these firms exert effort higher than the second-best level in order to induce their investors to export, while they exert effort at the second-best level in autarky. This again is due to the disciplining effect. In total, managers whose firms are close to either the exit cutoff or the exporting cutoff exert more effort to improve productivity levels of their firms after opening up to trade.

This article aims to speak to the empirical literature on the response of firm productivity to trade liberalization. Lileeva and Trefler (2010) document that new Canadian exporting firms experienced productivity gains after the enactment of the Canada-U.S. Free Trade Agreement. Bustos (2011) finds that Argentinean firms hose

\footnote{Remember that market size faced by non-exporters shrinks after opening up to trade.}
size is in the third quartile of the size distribution received productivity gains after MERCOSUR went into effect, and these firms were most likely to be the smallest exporters. These two findings are consistent with the prediction of my model. Trefler (2004), Biesebroeck (2006), De Loecker (2007), and Fernandes (2007) also document similar evidence to show the positive effect of exporting on firm productivity. Furthermore, Schor (2004) presents evidence that the least productive Brazilian firms improved their productivity levels after import tariffs in Brazil were reduced. This finding is consistent with the main prediction of my model. That is, the least productive firms that have survived trade liberalization improve their productivity. In total, theoretical predictions yielded by the model gain support from the evidence.

Although this paper considers firms whose ownership is separate from control, there are firms that do not have such an agency problem in reality. It should be noted that the main result of this paper (i.e., small firms that have survived trade liberalization increase their productivity) applies only to firms that have the agency problem. Indeed, this distinctive prediction of the current paper gains support from the data. For instance, using panel data on small UK establishments, Griffith (2001) finds that only establishments whose managerial control and ownership were separated gained in productivity after the introduction of the European Union Single Market Program. She also conjectures that this was because intensified competition reduced agency costs for those small establishments. Her finding and conjecture echo the main insight and economic implication of this paper.

The relationship between market competition and firm productivity is an old question in economics. A Schumpeterian view suggests that intensified competition

\[\text{For example, proprietary firms probably do not have the agency problem, since the owner controls and runs the firm.}\]

\[\text{If the investor and the manager are the same person, what matters for the effort choice is the marginal return to exerting effort, which is affected by the market size faced by the firm. After opening up to trade, the market size does not increase for non-exporting firms that are small. Therefore, an increase in productivity should not be observed for small firms that do not have the agency problem.}\]
destroys firms’ profit and, accordingly, their incentive to improve productivity.

However, this seems to stand at odds with a vast set of empirical findings and case studies showing that competitive pressure does make firms produce more efficiently and managers work harder. Therefore, economists have constructed various models in order to explain these findings. However, none of them takes firm heterogeneity into account. Furthermore, most of these papers derive results from partial equilibrium analysis without worrying about endogenous changes in market competition. This paper bridges the gap between the partial equilibrium analysis of the manager’s effort choice and the general equilibrium analysis of market competition under firm heterogeneity.

The rest of this paper is organized as follows. Section two analyzes the model for a closed economy. Section three analyzes the model for an open economy. Section four concludes. Proofs of the main results are relegated to the appendices.

### 3.2 The Closed Economy

In this section, I characterize the equilibrium in the closed economy. The key feature of the model is that the equilibrium effort exerted by the manager is a non-monotonic function of the initial quality of the implementable idea.

---

9Seminal papers in this literature include Grossman and Helpman (1991), Aghion and Howitt (1992), and others. Aghion and Howitt (1998) provides an excellent survey on this topic.

10Holmes and Schmitz (2010) offer an excellent survey on this topic.

11Seminal papers include Hermelin (1982), Schmidt (1997), Raith (2003), and Vives (2008). Aghion et al. (2005) show that the relationship between competition and innovation is non-monotonic.

12Wu (2001) is one exception. In that paper, he considers the role of manager in production explicitly and derives interesting results on changes in managerial remuneration schemes after trade liberalization. However, his paper does not focus on the impact of trade liberalization on firm productivity.
3.2.1 Environment

There are three types of agents in the economy: workers, managers who can choose to be either managers or workers, and investors. Their endowments are \( L, M, \) and \( I, \) respectively, and they are fixed throughout the paper. I assume that the measure of investors (i.e., \( M \)) is big enough such that the free-entry condition discussed below holds as an equality. Workers are homogeneous and used as inputs to production and receive a uniform wage from employment. Investors are homogeneous and have rough business ideas and enough resources to form firms. Managers are also homogeneous, and some of them are matched with the investors after the investors enter the industry.

There is one industry populated by firms that produce differentiated products under conditions of monopolistic competition à la Dixit and Stiglitz (1977). Each variety is indexed by \( \omega, \) and \( \Omega \) is the set of all varieties. Consumers derive utility from consuming these differentiated goods according to

\[
U = \left[ \int_{\omega \in \Omega} q(\omega)^{\frac{\sigma+1}{\sigma}} d\omega \right]^{\frac{\sigma}{\sigma-1}}, \tag{3.1}
\]

where \( q(\omega) \) is the consumption of variety \( \omega, \) and \( \sigma \) is the constant elasticity of substitution (CES) between differentiated goods.

The timing of the game is illustrated in Figure 3.1 and can be described as follows. First, an investor can enter the industry by paying a fixed entry cost, denoted by \( f_e, \) and then she is randomly matched with a manager. After the match, a firm is set up, and the manager and the investor discuss an implementable idea whose initial quality, \( \rho, \) is randomly realized. Second, the manager makes his occupational choice. He can quit the firm and become a worker, in which case the investor receives zero profit afterwards. Alternatively, he can choose to work for the investor and exert effort, denoted by \( \psi, \) to develop the implementable idea that leads to a blueprint for a product (i.e., variety \( \omega \)). Third, if the manager chooses to work for the investor, the
investor needs to decide whether or not to pay a fixed production cost, $f$, to start production. I assume that the investor observes the overall quality of the implementable idea, which equals $\rho \psi$, when deciding whether or not to start production. The overall quality of the implementable idea determines the labor productivity of the firm in the subsequent production.\(^{13}\)

Fourth, if the production starts, the manager decides the output level and the number of workers employed. At that point, firms compete in the market, revenue and the operating profit are received, and the variable cost is paid. Finally, the investor and the manager bargain over the operating profit to receive income. For simplicity, I assume that they play a generalized Nash bargaining game. As a result, the manager and the investor receive fractions $\alpha$ and $1 - \alpha$ of the operating profit, respectively, as discussed in the introduction.\(^{14}\)

Workers and managers are inputs to production. In order to produce $q$ units output, a firm must employ $\frac{q}{\rho \psi(p)}$ units of workers. One point worth mentioning is

\(^{13}\)Alternatively, I can assume that the overall quality of the implementable idea pins down the quality of the product. Qualitative results of the model are unchanged under this alternative specification.

\(^{14}\)The partnership firm is a perfect example of the type of firm I consider in this paper. An investor who has enough money and a rough idea to produce a certain product has to find a partner (i.e., the manager) who has knowledge and experience to make this rough idea implementable. After agreeing to the match, they talk about the idea and then come up with an implementable idea. The implementable idea might be a bad business idea, since it might be technologically impossible to produce the product the investor has in mind, or the product the investor has in mind might be out of date. On the contrary, if the implementable idea is a good idea, the manager exerts effort to further develop the idea, which eventually leads to a blueprint for the product. After the blueprint comes out, the investor decides whether or not to start production by comparing the profit she is going to receive with the fixed production cost. If the production is carried out, the investor and the manager split the ex post profit through bargaining. I assume that it is impossible to make an ex ante contract (i.e., make a contract before the realization of the profit) due to the existence of unforeseen contingencies, which is assumed in the literature on incomplete contracts.
that the manager’s effort considered here does not literally mean the amount of time he works. It represents the amount of time the manager works and how hard he works in the interest of the firm\(^{15}\). In order to exert effort, the manager must incur a cost (i.e., disutility) in terms of the numeraire of \( \gamma \psi^{\theta_0} \). The parameters \( \gamma (> 0) \) and \( \theta_0 (> \sigma - 1) \) measure the cost of exerting effort.

### 3.2.2 Effort Provision and the Decision to Produce

I use backward induction to solve the equilibrium and highlight the interaction between the manager’s effort choice and the investor’s decision to start production. Based on the utility function defined in equation (3.1), the demand function for a firm charging price \( p \) is derived as

\[
q(p) = \left( \frac{p}{P} \right)^{-\sigma} \frac{Y}{P},
\]

where \( P \) is the ideal price index of the CES goods and defined as

\[
P \equiv \left[ \int_{\omega(\rho) \in \Omega} p^{1-\sigma}(\rho) EdF(\rho) \right]^{\frac{1}{1-\sigma}},
\]

where \( F(\rho) \) is the cumulative density function (c.d.f.) of the random draw, \( \rho \), and \( E \) is the measure of varieties (or the measure of entering firms).

Since the manager’s effort choice does not affect the fraction of operating profit he receives, the optimal price determined at the fourth stage is to maximize the operating profit. As a result, the optimal pricing rule is the same across firms and can be written as

\[
p(\rho) = \frac{w}{\rho \psi(\rho) \lambda},
\]

\(^{15}\)Bandiera et al. (2011) show that the amount of time a manager spends inside the firm is highly positively correlated with firm profitability.
where \( w \) is the worker’s wage and \( \lambda \equiv (\sigma - 1)/\sigma \) is the inverse of the markup. I choose the worker’s wage \( w \) to be the numeraire. From equations (3.2) and (3.3), I derive the operating profit as

\[
\pi(\rho, \psi) = \frac{1}{\sigma} R(\rho, \psi) = \frac{1}{\sigma} (\rho \psi \lambda P)^{\sigma - 1} Y,
\]

where \( R(\rho, \psi) \) is the revenue.

At stage three, the investor is willing to start production, if and only if the fraction of operating profit he receives is larger than or equal to the fixed production cost. Formally, the participation constraint of the investor is

\[
(1 - \alpha)\pi(\rho, \psi) - f = \frac{(1 - \alpha)}{\sigma} (\rho \psi \lambda P)^{\sigma - 1} Y - f \geq 0.
\]

The manager’s effort choice at stage two is more involved. I discuss it case by case. If the investor is willing to produce, the objective function of the manager is

\[
\max_{\psi} \frac{\alpha}{\sigma} (\rho \psi \lambda P)^{\sigma - 1} Y - \gamma \psi^\theta
\]

s.t. \( \frac{\alpha}{\sigma} (\rho \psi \lambda P)^{\sigma - 1} Y - \gamma \psi^\theta \geq 1, \)

which can be transformed into

\[
\max_{\beta} \alpha \eta(P, Y) \varphi \beta - \gamma \beta^\theta
\]

s.t. \( \alpha \eta(P, Y) \varphi \beta - \gamma \beta^\theta \geq 1, \)

where \( \varphi \equiv \rho^{\sigma - 1}, \beta \equiv \psi^{\sigma - 1}, \theta \equiv \theta^{\sigma - 1}, \) and \( \eta(P, Y) \equiv \frac{1}{\sigma} (\lambda P)^{\sigma - 1} Y. \) Note that the inequality inside the above optimization problem is the manager’s participation constraint. Namely, the payoff the manager receives has to be higher than or equal
to his outside option. The solution to this optimization problem is

$$
\beta_a(\varphi) = \beta_2(\varphi) \equiv \left( \frac{\alpha \eta(P,Y) \varphi}{\gamma \theta} \right)^{\frac{1}{\theta-1}},
$$

(3.6)

which is defined as the second-best level of effort.

Two things are worth noting here. First, when $\varphi$ is sufficiently small, the profit the investor receives from the ex post bargaining must be smaller than $f$ under the second-best level of effort. Therefore, there is a cutoff, $\varphi_a'$, such that the manager cannot compensate his investor by exerting effort at the second-best level, if the initial quality, $\varphi$, is below this cutoff. Formally, the cutoff $\varphi_a'$ is defined as

$$(1 - \alpha) \pi(\varphi_a' \beta_2(\varphi_a')) = f.$$  

(3.7)

Second, the manager with an implementable idea whose initial quality is $\varphi_a'$ chooses to be a worker if his payoff from running the firm is less than his outside option, $\alpha$. Specifically,

$$\frac{\theta - 1}{\theta} \alpha \pi(\varphi_a' \beta_2(\varphi_a')) < 1.$$

If the above inequality is satisfied, we are in an uninteresting case in which all managers exert effort at the second-best level, and all investors earn strictly positive profit. In reality, it is probably true that when firm owners (i.e., investors) barely make profits, their managers still obtain high compensation and stick to their jobs. Thus, it is more likely that we are in the case in which managers in firms earning zero profit obtain payoffs that are strictly larger than their outside option. The following assumption guarantees the existence of such a case, and I adopt this assumption in the subsequent analysis.

---

16 The mathematical expression for this statement is when $(1 - \alpha) \pi(\varphi_a' \beta_a(\varphi_a')) = f$, $\frac{\theta - 1}{\theta} \alpha \pi(\varphi_a' \beta_a(\varphi_a')) < 1$. This expression can be further simplified to $\frac{\theta - 1}{\theta} \alpha \pi(\varphi_a') < 1$. When the fixed investment cost is big, this condition cannot be satisfied.
**Assumption 2**

\[
\alpha > \frac{1}{1 + f[1 - \frac{1}{\theta}].}
\]

How does the manager with an implementable idea whose initial quality is below \( \phi_a' \) make the effort choice? First, choosing an effort level lower than \( \frac{\phi_a' \beta_2(\phi_a')}{\phi} \) is suboptimal for him, since the investor would not start production at the third stage. Second, choosing an effort level higher than \( \frac{\phi_a' \beta_2(\phi_a')}{\phi} \) is suboptimal for the manager as well. The investor is induced to start production if the effort level equals \( \frac{\phi_a' \beta_2(\phi_a')}{\phi} \). Any further upward deviation from this effort level reduces the manager’s payoff, since this effort level is already above the second-best level of effort. Finally, if the initial quality of the idea is low, exerting effort at the level of \( \frac{\phi_a' \beta_2(\phi_a')}{\phi} \) gives the manager a payoff lower than his outside option. Thus, there is another cutoff denoted by \( \phi_a^* \) such that if the initial quality is below this cutoff, the manager chooses to become a worker. In total, I have three cases for the manager’s optimal effort choice. If the initial quality of the implementable idea is high, the manager exerts effort at the second-best level. If the initial quality is low, the manager chooses to become a worker. If it is in the middle range, the manager exerts effort at the level that generates zero payoff to the investor. The following lemma summarizes these results.

**Lemma 9** Suppose Assumption 2 is satisfied. Define two cutoffs as follows:

\[
\phi_a' \equiv \left( \frac{f}{1 - \alpha} \right) \left( \frac{\gamma \theta}{\alpha \eta(P,Y)^\theta} \right)^{\frac{1}{\theta}}
\]

and

\[
\phi_a^* \equiv \frac{\phi_a'(\alpha f)^{\frac{1}{\theta}}}{\left( \theta[\alpha f - (1 - \alpha)] \right)^{\frac{1}{\theta}}} < \phi_a'. \tag{3.8}
\]

If the initial quality of the implementable idea is larger than \( \phi_a' \), the optimal effort level is

\[
\beta_a(\varphi) = \beta_{a2}(\varphi) \equiv \left( \frac{\alpha \eta(P,Y)\varphi}{\gamma \theta} \right)^{\frac{1}{\theta - 1}}.
\]

168
If the initial quality is between $\varphi_a^*$ and $\varphi'_a$, the optimal effort level is

$$\beta_a(\varphi) = \beta_{a0}(\varphi) \equiv \frac{\beta_2(\varphi'_a)\varphi'_a}{\varphi}. \quad (3.9)$$

If the initial quality is lower than $\varphi_a^*$, the manager chooses to become a worker.

Proof: See Appendix 3.5.1.

Figure 3.2 shows how the optimal effort level varies with the initial quality draw for a particular set of parameter values. It is “U” shaped in the sense that the optimal effort decreases first and increases afterwards with the initial quality draw. For the downward-sloping part, the fixed production cost acts as a disciplining device. Managers in the smallest firms exert effort higher than the second-best level, since they want to induce their investors to produce. The lower the initial quality, the more effort needed to incentivize the investor to start production. As a result, the optimal effort decreases with the initial quality of the implementable idea, when the initial quality is low. For the upward-sloping part, a complementarity plays a role. Namely, the higher the initial quality, the larger the marginal return to exerting effort, when the investor’s participation constraint is not a concern for the manager. In short, the initial quality and the optimal effort are substitutes when the initial quality is low and they are complements when it is high. The relative incentive, which is the difference between the equilibrium effort level and the first-best level of effort, decreases weakly with the initial quality. The marginal return to exerting effort rises with firm size (i.e., the initial quality). These results are consistent with findings from the literature on managerial payment (e.g., Baker and Hall (2004)).

I discuss the relationship between the initial quality of the implementable idea and firm productivity. In this paper, I define firm productivity as the product of the initial quality and the manager’s effort level, since it pins down the marginal

\footnote{The first-best level of effort is defined as the one that maximizes the total operating profit minus the effort cost.}
cost of production. Figure 3.3 shows two properties of this relationship. First, for big firms whose products have high initial quality, firm productivity increases with the initial quality (and firm size). Second, for small firms whose products have low initial quality, firm productivity is a constant. For this type of manager, the optimal effort decreases with the initial quality. Moreover, the decrease in the optimal effort perfectly offsets the increase in the initial quality. As a result, firm productivity is a constant for small firms. In total, there is a non-decreasing relationship between the initial quality of the implementable idea and firm productivity.

For future use, I derive the manager’s payoff \( V_m(\varphi) \) from equations (3.6) and (3.9) as follows:

\[
V_m(\varphi) = \frac{\alpha f}{1 - \alpha} - \left( \frac{\varphi'_{\alpha}}{\varphi} \right)^{\theta - 1} \frac{\alpha f}{\theta (1 - \alpha)},
\]

(3.10)
when $\varphi \in [\varphi^*_a, \varphi'_a)$, and

$$V_m(\varphi) = \frac{\alpha(\theta - 1)}{\theta} \varphi \beta_2(\varphi) \eta(P, Y),$$  \hspace{1cm} (3.11)$$

when $\varphi \geq \varphi'_a$.

### 3.2.3 Aggregation in the Closed Economy

In this subsection, I analyze the general equilibrium of the closed economy. In order to obtain analytical results, I assume that the initial quality of the implementable idea is drawn from a Pareto distribution:

$$G(\varphi) = 1 - \left(\frac{\varphi_{\text{min}}}{\varphi}\right)^k,$$
where the shape parameter $k$ is negatively related to the variance of the distribution.\textsuperscript{18}

There are three sets of equilibrium conditions. The first set is related to the cutoffs. The zero cutoff profit condition (ZCP) indicates that firms whose products’ initial quality is $\varphi'_{a}$ break even in equilibrium, or

$$
(1 - \alpha)\varphi'_{a}^2(\varphi'_{a})\eta(P, Y) = f. \tag{3.12}
$$

The free entry (FE) condition indicates that the investors make zero expected profit upon entry, or

$$
f \int_{\varphi'_{a}}^{\infty} \left[ \left( \frac{\varphi}{\varphi'_{a}} \right)^{\frac{\theta}{\vartheta-1}} - 1 \right] g(\varphi) d\varphi = f_{e}. \tag{3.13}
$$

Since there are a large group of investors and a large group of managers (i.e., $I >> E$ and $M >> E$), every investor who enters the industry is matched with a manager for sure, and the FE condition holds as an equality. Note that firms whose products’ initial quality is between $\varphi^*_a$ and $\varphi'_{a}$ make zero profit in equilibrium. I do not include the expected profit of these firms in the FE condition. Finally, the exit cutoff for the manager (i.e., $\varphi^*_a$) is pinned down by equation (3.8).

The second set of equilibrium conditions is related to the effort choice. As Lemma \textsuperscript{9} states, the optimal effort is determined by equation (3.6) when $\varphi \geq \varphi'$ and by equation (3.9) when $\varphi \in [\varphi^*, \varphi')$.

The final set of equilibrium conditions is about market clearing. The demand for labor contains three parts: labor used for firm entry, for the fixed production cost, and for the variable cost. The supply of labor is the sum of workers and managers who are not matched with the investors or who choose to be workers. Therefore, the labor-market-clearing condition is

$$
Ef_{e} + Ef[1 - G(\varphi^*_a)] + \int_{\varphi^*_a}^{\infty} \lambda R(\varphi) d\varphi = L + M - E[1 - G(\varphi^*_a)]. \tag{3.14}
$$

\textsuperscript{18}In order to have a finite expected profit from entry, $k$ has to be bigger than $\frac{\theta}{\vartheta-1}$.
For the product market, the FE condition implies that the value of entry equals the sunk entry cost. Therefore, the total income of the economy equals total revenue of firms, or

\[
Y \equiv L + M - E[1 - G(\varphi_a^*)] + \int_{\varphi_a^*}^{\infty} \left[ V_m(\varphi) + \gamma(\beta_a(\varphi))\theta \right] Eg(\varphi) d\varphi \\
+ \int_{\varphi_a^*}^{\infty} V_f(\varphi) Eg(\varphi) d\varphi - Ef_e \\
= \int_{\varphi_a^*}^{\infty} R(\varphi \beta_a(\varphi)) Eg(\varphi) d\varphi, \tag{3.15}
\]

where \( V_f(\varphi) \equiv (1 - \alpha) \varphi \beta_a(\varphi) \eta(P, Y) - f \) is the owner’s profit after entry, and \( V_m(\varphi) \) is defined in equations (3.10) and (3.11).

The general equilibrium of the economy is characterized by the zero profit cutoff, \( \varphi'_a \), the mass of entrants, \( E \), the exit cutoff, \( \varphi_a^* \), the effort choice, \( \beta_a(\varphi) \), the worker’s wage, \( w \), and the total income, \( Y \). These variables are obtained by solving equations (3.6), (3.8), (3.9), and (3.12) to (3.15). One equilibrium condition is redundant due to Walras’ law, and I normalize the worker’s wage to one. It is straightforward to use the method that is employed in Melitz (2003) to show that a unique equilibrium exists. I omit the discussion here to save space.

### 3.3 The Open Economy

In this section, I analyze the properties of managerial effort and firm productivity in the open economy. The focus of my analysis is to explore the differential impact of trade liberalization (or the opening up to trade) on the equilibrium effort choice and firm productivity.
3.3.1 Trade Costs and Selection into Exporting

Similar to Melitz (2003), I assume there are two symmetric countries in the world: \( \tau > 1 \) is the iceberg (or variable) trade cost, and \( f_x \) is the fixed trade cost. The iceberg trade cost means that if \( \tau \) units of output are shipped to the foreign market, only one unit of it arrives. The fixed trade cost means that the firm (i.e., the investor) must incur an additional fixed cost in order to export.\(^{19}\)

The analysis for the behavior of the manager and the investor is similar to before. First, the optimal price decided by the manager at the fourth stage is still designed to maximize the expected profit. Second, the investor’s participation constraint (i.e., the decision to start production at the third stage) is still governed by equation (3.5). This participation constraint yields a cutoff on firm productivity such that the investor is willing to produce, if and only if firm productivity is above this threshold. This cutoff is derived as

\[
(\beta \varphi)_f = \frac{f}{(1 - \alpha)\eta(P,Y)}. \tag{3.16}
\]

In addition to this decision, the investor now has to choose whether or not to export at stage three as well. Since the two countries are symmetric, the investor whose product’s initial quality is \( \varphi \) is willing to export, if and only if

\[
(1 - \alpha)\varphi \beta(\varphi)\eta(P,Y) \left(1 + \frac{1}{\tau^{\sigma-1}}\right) - f_x - f > (1 - \alpha)\varphi \beta(\varphi)\eta(P,Y) - f, \tag{3.17}
\]

where \( \beta(\varphi) \) is the effort exerted by the manager at stage two, and \( P \) and \( Y \) are the ideal price index and total income of the two countries. Note that the ratio of the operating profit obtained from the foreign market to that obtained from the domestic market is \( \frac{1}{\tau^{\sigma-1}} \). This participation constraint for exporting yields a cutoff on firm productivity, such that the investor is willing to export if firm productivity is above

\(^{19}\)Similar to the timing assumed in the closed economy, I assume that the investor decides whether or not to export at stage three in the open economy.
this cutoff. The cutoff on firm productivity is derived as

$$(\beta \varphi)_x = \frac{f_x^{\sigma-1}}{(1 - \alpha) \eta(P, Y)}. \quad (3.18)$$

### 3.3.2 The Optimal Effort Choice in the Open Economy

A given level of effort brings more profits to the firm if the initial quality of its product is higher. Thus, there is an exporting cutoff $\varphi_x^*$, meaning that if the initial quality of the implementable idea is higher than this cutoff, the investor chooses to export. I consider the case in which there is selection into exporting among firms making a positive profit\(^{20}\) (i.e., the exporting cutoff, $\varphi_x^*$, is bigger than the zero profit cutoff, $\varphi'_f$), and a sufficiently large fixed trade cost ensures it is the case. Empirical evidence motivates this choice.\(^{21}\)

I analyze how the manager makes his effort choice at the second stage, case by case. Since firms whose products’ initial quality is low can’t afford to export in the open economy, the analysis in the closed economy applies to non-exporters in the open economy as well. More specifically, I derive two cutoffs similar to those derived in Lemma 9, as follows:

$$\varphi'_f \equiv \left( \frac{f}{(1 - \alpha)} \right)^{\theta - 1} \left( \frac{\gamma P Y}{\alpha \eta(P, Y) \theta} \right)^{\frac{1}{\theta}} \quad (3.19)$$

and

$$\varphi^*_f \equiv \frac{\varphi'_f (\alpha f)^{\frac{1}{\theta}}}{\left( \theta [\alpha f - (1 - \alpha)] \right)^{\frac{1}{\theta}}} < \varphi'_f. \quad (3.20)$$

The only difference here is that $P$ and $Y$ are the ideal price index and the total income in the open economy. If the initial quality of the implementable idea is above $\varphi'_f$ and

\(^{20}\)Subscript “f” is for firms serving only the domestic market in the open economy, subscript “x” is for firms serving both markets in the open economy, and subscript “a” is for firms in autarky.

\(^{21}\)Data show that only a small fraction of firms export, and exporting firms receive higher profits and revenue than non-exporting firms. For instance, only 18% of U.S. manufacturing firms exported in 2002.
below some threshold which I will discuss soon, the manager’s optimal effort choice is

\[ \beta_f(\varphi) = \beta_{f2}(\varphi) \equiv \left( \frac{\alpha\eta(P,Y)\varphi}{\gamma\theta} \right)^{\frac{1}{\theta-1}}, \quad (3.21) \]

and his investor starts production but does not export. If the initial quality is between \( \varphi_f^* \) and \( \varphi_f' \), the optimal effort level is

\[ \beta_f(\varphi) = \beta_{f0}(\varphi) \equiv \frac{\beta_{f2}(\varphi')\varphi'}{\varphi}, \quad (3.22) \]

and his investor starts production but, again, does not export.

The analysis for a firm whose product’s initial quality is much higher than \( \varphi_f' \) is more involved, since its manager realizes that he can exert effort at a level higher than the one specified in equation (3.19) to induce his investor to not only produce but also export. I use the following proposition to summarize the manager’s optimal effort choice in the open economy.

**Proposition 13 Assumption 3** Assume that

\[ \frac{f_x^\sigma \tau^{\sigma-1}}{f} \geq (1 + \frac{1}{\tau^{\sigma-1}})^{\frac{\theta}{\sigma-1}} \left[ \theta - \frac{\theta - 1}{(1 + \frac{1}{\tau^{\sigma-1}})^{\frac{\theta}{\sigma-1}}} \right]^{\frac{1}{\sigma-1}}. \]

When \( \varphi \geq \varphi_x' \), the optimal effort choice is given by

\[ \beta_x(\varphi) = \beta_{x2}(\varphi) \equiv \left( \frac{\alpha\eta(P,Y)\varphi}{\gamma\theta} (1 + \frac{1}{\tau^{\sigma-1}}) \right)^{\frac{1}{\theta-1}}, \quad (3.23) \]

where

\[ \varphi_x' \equiv \left( \frac{(f_x^\sigma \tau^{\sigma-1})^{\theta-1}\gamma\theta}{\alpha(1 - \alpha)^{\theta-1}\eta(P,Y)^{\theta}(1 + \frac{1}{\tau^{\sigma-1}})} \right)^{\frac{1}{\sigma}}. \quad (3.24) \]

When \( \varphi_x' > \varphi \geq \varphi_x^* \), the optimal effort level is

\[ \beta_x(\varphi) = \beta_{x0}(\varphi) \equiv \frac{\beta_{x2}(\varphi_x')\varphi_x'}{\varphi}, \quad (3.25) \]
where
\[
\varphi_x^* \equiv \frac{\varphi' \chi}{\left[ \theta - \frac{\theta - 1}{(1 + \frac{1}{\tau})^{\frac{1}{\tau}}} \right]^{\frac{1}{\theta}}}
\]
(3.26)
is the exporting cutoff such that the investor decides to export, if her product’s initial quality is higher than this threshold. When \( \varphi_x^* > \varphi \geq \varphi_f' \), the optimal effort level is
\[
\beta_{f2}(\varphi) \equiv (\frac{\alpha \eta(P,Y)\varphi}{\gamma \theta})^{\frac{1}{\theta - 1}},
\]
and his investor produces but does not export. When \( \varphi_f' > \varphi > \varphi_x^* \), the optimal effort is
\[
\beta_{f0}(\varphi) \equiv \frac{\beta_{f2}(\varphi')\varphi_f'}{\varphi},
\]
and his investor produces but, again, does not export. When \( \varphi \leq \varphi_x^* \), the manager chooses to become a worker.

Proof: See Appendix 3.5.2.

Figure 3.4 illustrates how the optimal effort level varies with the initial quality of the implementable idea in the open economy. It contains two “U”-shaped curves under Assumption 3. For firms’ whose products’ initial quality is between \( \varphi_f' \) and \( \varphi_x^* \), their managers choose effort levels higher than the second-best level, since they want to induce their investors to produce and obtain payoffs higher than their outside option. Similarly, for firms whose products’ initial quality is between \( \varphi_x' \) and \( \varphi_x^* \), their managers choose effort levels higher than the second-best level as well. They want to induce their investors to export and obtain payoffs higher than what they would obtain if they exerted effort at the second-best level (i.e., \( \beta_x(\varphi) \)). In total, the existence of the fixed costs (i.e., \( f \) and \( f_x \)) act as a disciplining device for managers in the least productive non-exporting and exporting firms. Assumption 3 is needed to ensure that the exporting cutoff, \( \varphi_x^* \), is bigger than the zero profit cutoff, \( \varphi_f' \). Vast
3.3.3 Aggregation in the Open Economy

Similar to the case of the closed economy, there are again three sets of equilibrium conditions in the open economy. The first set is still related to the cutoffs. First, the ZCP condition that pins down the zero profit cutoff \((\varphi_f')\) is still given by equation (3.19). Second, the exit cutoff, \(\varphi^*_f\), is determined by equation (3.20). Third, the exporting cutoff, \(\varphi^*_x\), is given by equation (3.26). Fourth, the zero exporting cutoff\(^{22}\)

\(^{22}\)The zero exporting cutoff is defined as the one under which the firm makes zero profit from exporting given the second-best level of managerial effort.
$\psi_x$, is determined by equation (3.24). Finally, the FE condition now becomes

$$f \int_{\psi_f}^{\psi_x} \left[ \left( \frac{\psi}{\psi_f} \right)^{\sigma - 1} - 1 \right] g(\psi) d\psi + f \int_{\psi_f}^{\psi_x} \left[ \left( \frac{\psi}{\psi_f} \right)^{\sigma - 1} (1 + \frac{1}{\tau^{\sigma-1}}) \frac{1}{\psi_f} - 1 \right] g(\psi) d\psi$$

$$+ f \int_{\psi_f}^{\infty} \left[ \left( \frac{\psi}{\psi_f} \right)^{\sigma - 1} - 1 \right] g(\psi) d\psi + \int_{\psi_f}^{\psi_x} \left[ f_x \tau^{\sigma-1} - f \right] g(\psi) d\psi = f_e,$$

(3.27)

where the first term of the above equation is the expected profit the investor receives when his product’s initial quality is between $\psi_f$ and $\psi_x$. The second term of the above equation is the expected profit the investor receives when his product’s initial quality is between $\psi_x$ and $\psi_f$, and the third term is the expected profit the investor receives when his product’s initial quality is above $\psi_f$. After recalling the definition of the cutoffs, I simplify the above equation to

$$f \int_{\psi_f}^{\psi_x} \left[ \left( \frac{\psi}{\psi_f} \right)^{\sigma - 1} - 1 \right] g(\psi) d\psi + f \int_{\psi_f}^{\psi_x} \left[ \left( \frac{\psi}{\psi_f} \right)^{\sigma - 1} (1 + \frac{1}{\tau^{\sigma-1}}) \frac{1}{\psi_f} - 1 \right] g(\psi) d\psi$$

$$+ f \int_{\psi_f}^{\infty} \left[ \left( \frac{\psi}{\psi_f} \right)^{\sigma - 1} - 1 \right] g(\psi) d\psi + \int_{\psi_f}^{\psi_x} \left[ f_x \tau^{\sigma-1} - f \right] g(\psi) d\psi = f_e.$$

(3.28)

The second set of equilibrium conditions is related to the manager’s effort choice. Equations (3.21), (3.22), (3.23), and (3.25) pin down the manager’s equilibrium effort choice. The third set is related to market clearing. First, the labor-market-clearing condition indicates that

$$Ef_e + Ef[1 - G(\psi_f)] + Ef_x[1 - G(\psi_x)] + \int_{\psi_f}^{\psi_x} \lambda R(\psi \beta_f(\psi)) Eg(\psi) d\psi$$

$$+ \int_{\psi_x}^{\infty} \lambda R(\psi \beta_x(\psi)) Eg(\psi) d\psi = L + M - E[1 - G(\psi_f)].$$

There is also a product-market-clearing condition similar to the one derived in the closed economy (i.e., equation (3.15)), which I omit here.
The general equilibrium of the open economy is characterized by the zero profit cutoff, $\varphi_f'$, the mass of entrants, $E$, the exit cutoff, $\varphi_f^*$, the exporting cutoffs, $\varphi_x^*$, the cutoff, $\varphi_x'$, the effort choices, $\beta_f(\varphi)$ and $\beta_x(\varphi)$, the worker’s wage (normalized to one), and the total income, $Y$. These variables are obtained by solving equations (3.19) to (3.29).

### 3.3.4 Opening Up to Trade and Firm Productivity

In this subsection, I discuss how opening up to trade affects the optimal effort choice as well as firm productivity. The key economic insight is that intensified competition due to the introduction of international trade acts as a disciplining device for managers in the smallest surviving non-exporting and exporting firms. The following proposition summarizes the main result of the paper.

**Proposition 14** After opening up to trade, the exit cutoff ($\varphi_f^*$) and the zero profit cutoff ($\varphi_f'$) both increase. If the trade costs are not too small in the open economy, there exists a cutoff on the initial quality draw, $\varphi''_f \in (\varphi_f^*, \varphi_f')$, such that, for surviving non-exporters with $\varphi \leq \varphi''_f$, the equilibrium effort level and firm productivity are higher in the open economy than in the closed economy. Productivities of the smallest exporters and the equilibrium effort level of managers working in these firms are also higher in the open economy than in the closed economy.

Proof: See Appendix 3.5.3.

Figure 3.5 shows how the optimal effort level changes after the economy opens up to trade. Managers of the smallest surviving non-exporters and exporters exert more effort in the open economy than in the closed economy. This is mainly due to the disciplining effect. In order to make the investors start production and receive higher

---

23 Although simulation results suggest that managers in all exporting firms increase their effort provision after the economy opens up to trade, I can prove only that it is true at least for the smallest exporters.
payoffs than their outside option, managers in the smallest surviving (non-exporting) firms have to exert more effort. In other words, the introduction of international trade reduces rents earned by managers of the smallest firms and mitigates the agency problem inside these firms. Managers of the smallest exporting firms exert more effort for two reasons. First, enlarged market size increases the marginal return to exerting effort (i.e., the market size effect) and the second-best level of effort. Second, the disciplining effect works for them as well. Managers of the smallest exporters exert effort higher than the second-best level to induce their investors to export. The two effects together incentivize managers of the smallest exporting firms to exert more effort in order to induce their investors to export. Since changes in the manager’s effort level directly translate into changes in firm productivity, productivities of the smallest surviving non-exporters and exporters increase after opening up to trade. In total, the disciplining effect, which is the key idea of this paper, induces managers of the smallest surviving non-exporters and exporters to exert more effort and improve firm productivity when the economy opens up to trade.

3.3.5 Discussion

In this subsection, I discuss the empirical relevance of the main result of this paper (i.e., Proposition 14). The first set of empirical evidence related to my main result comes from the literature on the response of firm productivity to trade liberalization. Bustos (2011) finds that Argentinean firms whose size is in the third quartile of the size distribution received productivity gains after MERCOSUR went into effect, and these firms were most likely to be the smallest exporters. De Loecker (2007), Fernandes (2007), and Lileeva and Trefler (2010) also document similar evidence. This evidence supports one of the main results of the paper: that at least the smallest exporters experience improvements in productivity after the economy opens up to trade. Furthermore, Schor (2004) presents evidence that the least productive Brazil-
ian firms (most likely non-exporting firms) improved their productivity levels after import tariffs were reduced in Brazil. This finding is consistent with the other main prediction of my paper. That is, the least productive (non-exporting) firms that have survived trade liberalization improve productivity levels. In total, theoretical predictions yielded by the model gain support from existing evidence.

Although this paper considers firms whose ownership is separate from control, there are firms that do not have such an agency problem in reality. It should be noted that one main result of this paper (i.e., small firms that have survived trade liberalization increase their productivity) applies only to firms that have the agency problem. Indeed, this prediction, distinct to this paper, gains support from the data.

---

24 For example, proprietary firms probably do not have the agency problem, since the owner controls and runs the firm.

25 If the investor and the manager are the same person, what matters for the effort choice is the marginal return to exerting effort, which is affected by the market size the firm faces. After trade liberalization, the market size does not increase for non-exporting firms, which are mainly small firms. Therefore, an increase in productivity should not be observed for small firms that do not have the agency problem.
ta. For instance, using panel data on small UK establishments, Griffith (2001) finds that only establishments whose managerial control and ownership were separate saw productivity gains after the European Union Single Market Program was introduced. She also conjectured that this was because intensified competition reduced agency costs for those small establishments. Her finding and conjecture echo the main insight and economic implication of this paper. Additionally, using a survey data set of Australian workplaces, Roger (2004) finds similar results. He documented that intensified competition increased labor productivity of small managerial firms (i.e., firms whose ownership is separate from control), but not that of small non-managerial firms. In total, the main prediction of my paper gains support from several findings in the literature.

Finally, I should point out that the main insight of this paper applies to other types of economic reforms, such as industry deregulation, as well. It is straightforward to observe that market competition becomes tougher, and the exit cutoff on the initial quality draw increases, if the government reduces the entry cost $f_e$. Then the same logic argued before applies to the smallest surviving firms as well: namely, managers of these firms exert more effort in order to induce their investors to produce and continue to receive rents after industry deregulation. It would be interesting to see if this prediction is supported by the data, but that is left for future research.

### 3.4 Concluding Remarks

This paper presents a model that captures the agency problem inside the firm in order to explain why some firms improve their productivity after trade liberalization. The main economic insight is that the least productive non-exporters increase their productivity levels when the economy opens up to trade. This is because their man-
agers exert more effort in order to incentivize the investors to produce and continue to receive rents by running the firms after opening up to trade.

Nevertheless, much remains to be done. From the theoretical point of view, there are at least two issues that can be investigated further. First, this model has the potential to explain changes in managerial effort in the context of gradual trade liberalization. It is clear that, although the least productive firms exit the market eventually, they improve their productivities before exiting in the process of gradual trade liberalization. In order to implement such an analysis, I probably need to construct a dynamic model. Second, using the current model to discuss how other types of economic reforms affect firm productivity is also an interesting topic for future research. From the empirical point of view, information on managerial effort is needed if I want to test the model’s predictions directly.

3.5 Appendix

3.5.1 Proof of Lemma 9

Proof. The proof proceeds in the following way. First, \( \beta_{a2}(\varphi) \) is the optimal effort choice when \( \varphi > \varphi'_a \), since the operating profit the investor receives is bigger than the fixed cost. Assumption 2 assures that the manager receives a payoff higher than his outside option when \( \varphi > \varphi'_a \) and \( \beta = \beta_{a2}(\varphi) \). Second, when \( \varphi < \varphi'_a \), the effort level of \( \beta_{a0}(\varphi) \) is the minimum effort level under which the investor breaks even. Furthermore, this is also the optimal effort level for the manager, if he wants to induce the investor to produce.\(^{26}\) Now, the question becomes whether or not this effort provision yields a higher payoff to the manager than his outside option, or

\[
\frac{\alpha f}{(1 - \alpha)} - \frac{\alpha f}{\theta(1 - \alpha)} \left( \frac{\varphi'_a}{\varphi} \right)^\theta \geq 1,
\]

\(^{26}\)Remember that \( \beta_{a0}(\varphi) > \beta_{a2}(\varphi) \) and the payoff function of the manager is concave in \( \varphi \).
where the first part of the above inequality is the manager’s profit when \( \varphi = \varphi'_a \), and the second term is his effort cost. Solving this inequality, I obtain the result that when \( \varphi'_a > \varphi \geq \varphi^*_a \), the manager chooses to run the firm and exerts effort at the level of \( \beta_{a0}(\varphi) \). He chooses to become a worker, if \( \varphi < \varphi^*_a \). QED.

### 3.5.2 Proof of Proposition 13

Proof. First, I consider the manager whose product’s initial quality is very high in the sense that his investor is willing to export (and start production), even if the manager exerts effort at the second-best level. More specifically, the manager’s objective function in this case is

\[
\max_{\beta} \alpha \eta(P,Y) \varphi \beta \left( 1 + \frac{1}{\tau^{\sigma-1}} \right) - \gamma \beta^\theta,
\]

which yields the solution that

\[
\beta_{x2}(\varphi) = \left( \frac{\alpha \eta(P,Y) \varphi}{\gamma \theta} \left( 1 + \frac{1}{\tau^{\sigma-1}} \right) \right)^{\theta(1-\sigma)}.
\]

The term \( \left( 1 + \frac{1}{\tau^{\sigma-1}} \right) \) shows the complementarity between exporting and the manager’s effort choice. The resulting firm productivity for this type of firm is

\[
\varphi \beta_x(\varphi) = \varphi \beta_{x2}(\varphi) = \left( \frac{\alpha \eta(P,Y) \varphi^\theta}{\gamma \theta} \left( 1 + \frac{1}{\tau^{\sigma-1}} \right) \right)^{\frac{1}{\theta(1-\sigma)}}.
\]

Based on equations (3.18) and (3.31), I derive a cutoff on the initial quality as

\[
\varphi'_x = \frac{(\beta \varphi)_{x}}{\beta_{x2}(\varphi'_x)} = \left( \frac{f_x}{\alpha(1-\alpha)^{\theta-1} \eta(P,Y)^\theta \left( 1 + \frac{1}{\tau^{\sigma-1}} \right)} \right)^{\frac{1}{\theta}} = \left( \frac{f_x}{\tau^{\sigma-1}} \right)^{\frac{\theta}{\theta-1}} \frac{\tau^{\sigma-1}}{(1 + \tau^{\sigma-1})^{\frac{1}{\theta}}} \varphi'_x.
\]
In total, if the initial quality of the implementable idea is bigger than $\varphi'_x$, the manager exerts effort at the second-best level denoted by $\beta_{x2}(\varphi)$, and the investor chooses to both produce and export.

Second, if the initial quality is below $\varphi'_x$, the manager realizes that if he exerts effort at the level of $\beta_{x2}(\varphi)$, his investor will not start to export. However, there is room for achieving a potential Pareto improvement. That is, the manager can exert effort at the level under which productivity equals $(\beta \varphi)_x$. Under this level of effort, the investor is willing to export which generates more operating profit.\(^{27}\) Alternatively, the manager can exert effort at the level specified in equation (3.21), and his investor will not export since $\beta_{f2}(\varphi) < \beta_{x2}(\varphi)$, which validates the manager’s effort choice. Now, the question becomes which option yields the highest payoff to the manager. First, if the manager chooses to exert effort at the level of $\beta_{f2}(\varphi)$, his payoff is

$$\frac{\theta - 1}{\theta} \frac{\alpha f}{1 - \alpha} \left( \frac{\varphi}{\varphi_f} \right) \frac{\varphi}{\varphi_f} = \frac{\theta - 1}{\theta} \left( \frac{\alpha f}{1 - \alpha} \right) \left( \frac{\varphi}{\varphi_f} \right) \frac{\varphi}{\varphi_f} \left( 1 + \frac{1}{1 - \tau} \right) \frac{\varphi}{\varphi_f},$$

where $\frac{\theta - 1}{\theta} \frac{\alpha f}{1 - \alpha}$ is the payoff received by the manager whose product’s initial quality is $\varphi'_f$. Second, if the manager wants to induce his investor to export, he has to exert effort at the level of

$$\beta_{x0}(\varphi) = \frac{\beta_{x2}(\varphi'_x) \varphi_x}{\varphi} = \frac{\alpha f \eta (P, Y) \varphi'_x}{\gamma \theta \varphi^{\theta - 1}} \left( 1 + \frac{1}{1 - \tau} \right) \frac{1}{\varphi_f}.$$

When $\varphi = \varphi'_x$ and the effort level equals $\beta_{x2}(\varphi)$, the manager’s payoff is

$$\frac{\theta - 1}{\theta} \frac{\alpha f}{1 - \alpha} \left( \frac{\varphi}{\varphi_f} \right) \frac{\varphi}{\varphi_f} \left( 1 + \frac{1}{1 - \tau} \right) \frac{1}{\varphi_f}. \quad (3.34)$$

\(^{27}\)Similar to the reasoning used in the closed economy, the manager has no incentives to exert effort at a level under which the investor strictly prefers exporting over not exporting.
the second choice yields the payoff for the manager as follows:

$$\frac{\theta - (\frac{\varphi'}{\varphi})^\theta}{\theta} \frac{\alpha f}{(1 - \alpha)} \left( \frac{\varphi}{\varphi_f} \right)^{\frac{\theta}{1+1}} \left( 1 + \frac{1}{\tau^\sigma - 1} \right)^{\frac{\theta}{\tau^\sigma - 1}}. \quad (3.35)$$

By comparing equation (3.33) with equation (3.35), I conclude that the manager with \( \varphi < \varphi' \) chooses the effort level of \( \beta_{x0}(\varphi) \), if and only if

$$\varphi \geq \varphi^*_x \equiv \left[ \frac{\varphi'}{\theta - \frac{\theta - 1}{(1 + \frac{1}{\tau^\sigma - 1})^{\frac{\theta}{\tau^\sigma - 1}}} \right]^{\frac{1}{\tau^\sigma - 1}} \quad (3.36)$$

and \( \varphi \leq \varphi' \). Since exporters are rare in the data and most of them make positive profits, I adopt an assumption to assure that all exporters make positive profits in what follows. This implies that \( \varphi^*_x \geq \varphi' \). By comparing equation (3.36) with equation (3.19), I obtain the following condition:

$$\frac{f_{x^\tau^\sigma - 1}}{f} \geq (1 + \frac{1}{\tau^\sigma - 1})^{\frac{1}{\tau^\sigma - 1}} \left[ \theta - \frac{\theta - 1}{(1 + \frac{1}{\tau^\sigma - 1})^{\frac{\theta}{\tau^\sigma - 1}}} \right]^{\frac{1}{\tau^\sigma - 1}}. \quad (3.37)$$

Note that the above inequality holds, if either the variable trade cost or the fixed trade cost is big enough.

Third, when \( \varphi < \varphi^*_x \), the analysis is exactly the same as the one done for the closed economy. If the initial quality of the implementable idea is between \( \varphi'_f \) and \( \varphi^*_x \), the manager’s optimal effort choice is

$$\beta_f(\varphi) = \beta_{f2}(\varphi) \equiv \left( \frac{\alpha \eta(P,Y) \varphi}{\gamma \theta} \right)^{\frac{1}{\tau^\sigma - 1}},$$
and his investor starts production but does not export. If the initial quality is between
\( \varphi_f^* \) and \( \varphi_f' \), the optimal effort level is

\[
\beta_f(\varphi) = \beta_{f0}(\varphi) \equiv \frac{\beta_{f2}(\varphi_f')\varphi_f'}{\psi},
\]

and his investor start production but does not export as well. If the initial quality \( \varphi \)
is smaller than \( \varphi_f^* \), the manager quits the firm and becomes a worker. QED.

### 3.5.3 Proof of Proposition 14

Proof. I prove this proposition by contradiction. Suppose the exit cutoff decreases
when the economy opens up to trade (i.e., \( \varphi'_a > \varphi'_f \)). This immediately implies that

\[
\eta(P_a, Y_a) < \eta(P_f, Y_f).
\]

However, I am going to prove that the value from entry must be bigger than the fixed
entry cost, if the above inequality holds. I show it in three steps.

First, for \( \varphi \geq \varphi'_x \), I must have

\[
(1 - \alpha)\eta(P_f, Y_f)\varphi \beta_{x2}(\varphi)(1 + \frac{1}{\tau_{\sigma-1}}) - f - f_x \\
\geq (1 - \alpha)\eta(P_f, Y_f)\varphi \beta_{x2}(\varphi) - f \\
> (1 - \alpha)\eta(P_a, Y_a)\varphi \beta_{a2}(\varphi) - f,
\]

where the first inequality comes from the result that exporters prefer exporting over
not exporting, and the second one comes from the result that \( \eta(P_a, Y_a) < \eta(P_f, Y_f) \).
Second, for $\varphi \in [\varphi^*_x, \varphi'_x]$ (if $\varphi^*_x > \varphi'_a$), I also have

$$(1 - \alpha)\eta(P_f, Y_f)\varphi\beta_{x0}(\varphi)(1 + \frac{1}{\tau\sigma - 1}) - f - f_x$$

$$= (1 - \alpha)\eta(P_f, Y_f)\varphi\beta_{x0}(\varphi) - f$$

$$> (1 - \alpha)\eta(P_a, Y_a)\varphi\beta_{a2}(\varphi) - f,$$

where the second inequality hold since $\eta(P_f, Y_f) > \eta(P_a, Y_a)$ and $\beta_{x0}(\varphi) > \beta_{x2}(\varphi) > \beta_{a2}(\varphi)$ for $\varphi \in [\varphi^*_x, \varphi'_x]$. The above two results together imply that for firms with $\varphi \geq \varphi^*_x$, their investors must earn high payoffs in the open economy than in the close economy.

Third, for firms with $\varphi \in [\varphi'_a, \varphi^*_x]$, I have

$$(1 - \alpha)\eta(P_f, Y_f)\varphi\beta_{f2}(\varphi) - f > (1 - \alpha)\eta(P_a, Y_a)\varphi\beta_{a2}(\varphi) - f,$$

since $\eta(P_a, Y_a) < \eta(P_f, Y_f)$, and for firms with $\varphi \in [\varphi'_f, \varphi'_a]$,

$$(1 - \alpha)\eta(P_f, Y_f)\varphi\beta_{f2}(\varphi) - f > 0$$

The above two inequality together imply that non-exporters also earn high payoffs in the open economy than in the close economy if $\varphi'_a > \varphi'_f$. In total, the FE condition cannot hold, if $\varphi'_a > \varphi'_f$. Therefore, I must have $\varphi'_a < \varphi'_f$, which implies that $\eta(P_a, Y_a) > \eta(P_f, Y_f)$. Namely, the zero profit cutoff increases after the economy opens up to trade. Furthermore, since the relationship between the exit cutoff and the zero profit cutoff is unchanged when the economy opens up to trade, the exit cutoff increases as well. Namely, I have $\varphi^*_a < \varphi^*_f$.

Next, I prove that when trade costs are not sufficiently small in the open economy, managers of the smallest surviving non-exporters must exert more effort in the open economy than in the closed economy. First, simple calculation shows that the manager
on the zero profit cutoff exerts the same level of effort in the open economy as in the closed economy, or

$$\beta_a(\varphi_a') = \left[ \frac{\alpha f}{\theta \gamma (1 - \alpha)} \right]^{\frac{1}{\theta}} = \beta_f(\varphi_f'),$$

which implies that

$$\beta_f(\varphi_f') = \beta_a(\varphi_a') < \beta_a(\varphi_f').$$

Namely, managers with the random draw of $\varphi_f'$ exert less effort in the open economy than in the closed economy. Next, when trade costs are not sufficiently small in the open economy, the increase in the zero profit cutoff is not too big. Since the relationship between the exit cutoff and the zero profit cutoff is unaffected by trade costs, it must be true that $\varphi_f^* < \varphi_a'$ when the trade costs are not sufficiently small. Finally, for firms with the random draw of $\varphi_a'$, I must have

$$\beta_f(\varphi_a') > \beta_f(\varphi_f') = \beta_a(\varphi_a').$$

Since $\beta_f(\varphi)$ decreases continuously with $\varphi$ when $\varphi \in [\varphi_a', \varphi_f']$, and $\beta_a(\varphi)$ increases continuously with $\varphi$ when $\varphi \in [\varphi_a', \varphi_f']$, it must be true that there exists a cutoff $\varphi_f'' \in (\varphi_f', \varphi_f)$ such that the effort level of managers whose products’ initial quality is between $\varphi_f^*$ and $\varphi_f''$ is higher in the open economy than in the closed economy.

Finally, I prove that at least the effort level of managers in the smallest exporting firms is higher in the open economy than in the closed economy. Suppose it is not. This would imply

$$\beta_{x0}(\varphi) < \beta_{a2}(\varphi)$$

for $\varphi \in [\varphi_x^*, \varphi_x']$ and

$$\beta_{x2}(\varphi) < \beta_{a2}(\varphi)$$

190
for $\varphi > \varphi'$. Then, for firms with $\varphi \in [\varphi^*_x, \varphi'_x]$, I have

$$
(1 - \alpha)\eta(P_f, Y_f)\varphi\beta_{x0}(\varphi)(1 + \frac{1}{\tau_{\sigma-1}}) - f - f_x
$$

$$
= (1 - \alpha)\eta(P_f, Y_f)\varphi\beta_{x0}(\varphi) - f
$$

$$
< (1 - \alpha)\eta(P_a, Y_a)\varphi\beta_{a2}(\varphi) - f,
$$

where the equality comes from the result that these firms are indifferent between exporting or not, and the inequality comes from the assumption that $\beta_{x0}(\varphi) < \beta_{a2}(\varphi)$.

Next, for firms with $\varphi > \varphi'_x$, I have

$$
(1 - \alpha)\eta(P_f, Y_f)\varphi\beta_{x2}(\varphi)(1 + \frac{1}{\tau_{\sigma-1}}) - f - f_x
$$

$$
= \frac{(1 - \alpha)\theta}{\alpha}\gamma\beta_{x2}(\varphi)^\theta - f - f_x
$$

$$
< \frac{(1 - \alpha)\theta}{\alpha}\gamma\beta_{a2}(\varphi)^\theta - f
$$

$$
= (1 - \alpha)\eta(P_a, Y_a)\varphi\beta_{a2}(\varphi) - f,
$$

due to the result that $\beta_{x2}(\varphi) < \beta_{a2}(\varphi)$. The above two inequalities together imply that exporters earn less profit in the open economy than in the closed economy. Furthermore, since $\eta(P_a, Y_a) > \eta(P_f, Y_f)$, non-exporters also earn less profit in the open economy than in the closed economy. These two results together contradicts that the investor earns zero expected profit both in the closed economy and in the open economy. Therefore, managers in the smallest exporters exert more effort in the open economy than in the closed economy. QED.
References:


