Productivity, Efficiency and Firm Size Distribution: Evidence from Vietnam*

Hien Thu Pham † Nhan Buu Phan‡ Shino Takayama§

March 26, 2020

Abstract

By applying recently developed methodologies to Vietnamese data from 2000 to 2016, this paper studies production efficiencies and total factor productivity in manufacturing industries, using two separate empirical methodologies. The paper shows that middle-sized firms' production efficiencies tend to be lower than those of small-sized or large-sized firms in most of the manufacturing industries, and that the middle-sized firms are quite diverse in terms of efficiencies. Further, we show that the level of productivity is also quite diverse across different firm sizes in most industries. Given these observations, we present a simple theoretical analysis to indicate how our empirical findings could affect a firm's decision in expanding business size. Over the past few decades, the Vietnamese government has conducted policies to promote small-sized and middle-sized firms. Our findings indicate the importance of adopting policies that reduce the uncertainty that those firms may possibly face, such as financial support.

*We would like to thank Andrew McLennan, Prasada Rao, Haishan Yuan, Valentin Zelenyuk and participants at the Productivity Analysis Workshop. We are grateful to the General Statistical Office of Vietnam for providing permission to use the microdata of the Enterprise surveys. Any remaining errors are our own.

†Pham; CSIRO, QLD, Australia 4006; e-mail: `Hien.Pham@data61.csiro.au`.

‡Phan; School of Economics, University of Queensland; email: `b.phan@uq.edu.au`

§Takayama; School of Economics, University of Queensland; email: `s.takayama1@uq.edu.au`; tel: +61-7-3346-7379; fax: +61- 7-3365-7299.

1 Introduction

Business environments vary across nations. Although it is difficult to compare how well a firm is managed across different environments, the level of production efficiency can reflect the management of an organization. In this paper, we study how heterogeneous firms' efficiency and total factor productivity (hereafter TFP) are at the firm level using Vietnamese manufacturing data from 2000 to 2016. We apply different recently developed methodologies, which allow us to analyze the relationships between production efficiency, productivity and firm sizes. We further provide a theoretical analysis of possible consequences from this heterogeneity we find by the empirical analysis on decision by a potential investor. This theoretical analysis brings potentially important policy implications on promoting small and middle sized firms.

Vietnam is an interesting case to study. Since the emergence of the "Doi-Moi" (Renovation) process in 1986, the economic system has shifted toward a market-oriented economy and the economy has grown rapidly. Decades of rapid growth and structural change have transformed Vietnam into a low-middle-income country. As a part of the Doi-Moi process, the Vietnamese government implemented policies to promote competition and to enhance economic growth, including large firms' privatization, and tried to develop small-sized and middle-sized firms (hereafter SMEs), which include adopting policies to promote resource accumulation and technological diffusion (i.e., pro-investment and innovation policies) as well as financial support. For instance, Decree 90/2001/ND-CP in 2001 offered financial support for traditional industries, or for developing employees' skills. In addition, SMEs are encouraged to export by receiving funds to connect to foreign partners, conduct foreign market research or to participate in international exhibitions.

However, our preliminary data analysis shows that these policies may not be that effective in promoting middle-sized firms. Examining continuing firms in 15 distinct manufacturing industries, our main empirical investigation shows that both the efficiency and productivity of middle-sized firms is quite diverse, which can be thought of as the uncertainty when SMEs try to grow as our theoretical analysis shows.

After the recent development of empirical methodologies, there has been an increasing interest in the relationship between firm size and characteristics such as productivity, innovation and market structure (see Acs and Audretsch, 1987), growth and productivity (see Bentzen et al., 2011), and job creation (see Dalton et al., 2011). However, there are few empirical studies on the relationship between firm size and efficiency level, those examining developing countries are particularly rare with some exceptions (see Little, 1987; Van Biesebroeck, 2005), perhaps due to data availability at a firm level, and most of them are about advanced countries (see Liu and Yang, 2000; Angelini and Generate, 2008; Lawrance and Marks, 2008; Yuan et al., 2016). Further, there are not many analyses on the efficiency of middle-sized firms even using advanced countries' data, although there are some works that compare the efficiencies of different size groups (see Diaz and Sanchez, 2008; Leung et al., 2008; Kato and Kodama, 2014).

This paper is one of the few that applies recent methodologies to the analysis of firms in a developing transitional economy. To estimate both the firm-specific TFP and efficiency terms, we conduct two separate empirical analyses, the procedures of which will be explained in details in the Appendix. The first part adopts the stochastic frontier (henceforth SF) approach from Aigner et al. (1977) and Meeusen and Van den Broeck (1977). We consider both fixed and random-effect analyses using the Gibbs sampling algorithm, which are discussed at length in Tsionas (2002), O'Donnell and Coelli (2005) and Feng and Zhang (2014).

In the second empirical analysis, we employ the two-stage GMM model from Ackerberg et al. (2015) to eliminate the simultaneity bias. In the literature of productivity analysis, the problem of simultaneity bias, which was first examined in Marschak and Andrews (1944), has been a persistent issue. The problem comes from the assumption that firms can observe their productivity value before making decisions on inputs. However, the productivity term is unobserved by the econometrician, thereby creating multicollinearity between the inputs and the TFP value, which is in essence a part of the residual term of the production function. Ackerberg et al. (2015) contribute to a subset of literature which uses a proxy variable to eliminate the bias. In particular, they use the intermediate input as a proxy for the productivity term. Then, assuming the monotonicity of the relationship between the two variables, the productivity term can be "inverted out" of the production function and estimated in the second stage of the procedure.

Overall, our analysis shows that in most of the manufacturing industries that we study, middle-sized firms' production efficiencies tend to be lower than those of small-sized or large-sized firms. Interestingly, our estimation results also show that middle-sized firms are quite diverse in terms of efficiencies. In addition, throughout all sectors, there is also a high variation in the distribution of the TFP term. There is a clear monotone trend for the average TFP value, which is constant for some sectors and decreasing for others. Dhawan (2001) separates the sample firms into two groups, namely small-sized and large-sized firms, and shows, both theoretically and empirically with U.S. data, that small firms tend to be more productive than large ones; however, they also face significantly higher risks. Our results, in the context of a developing country, agree with those of Dhawan (2001).

These two methodologies follow two different strands of literature on productivity analysis. The two-stage GMM approach from Ackerberg et al. (2015) intends to measure productivity for each firm (when $F(K, L) = a_i K^{\beta} L^{1-\beta}$ is a production function, then it measures a_i for firm i) while the SF approach measures how far away a firm is operating from its possible best production practice (distance to $F(K, L)$), and this distance is called a firm's *technical efficiency score*. In this paper we employ both approaches to test the robustness of our results and find that both productivity and efficiency show a high variation across firm sizes in Vietnam.

Finally, we present a theoretical analysis, which provides an insight into a firm's behavior under heterogeneous distributions of production efficiency. In particular, our model shows that as the variation of production efficiency increases, it becomes more unlikely that a firm expands in size.

The remainder of the paper is organized as follows. Section 2 explains the industrial policies in Vietnam. Section 3 describes our data set. Section 4 explains the two empirical methodologies used to find efficiency and productivity. Section 5 provides the results and lists the robustness checks used in the paper. Section 6 discusses the result and provides further theoretical analysis. Section 7 provides final remarks.

2 Industrial Policy and Firm Size in Vietnam

After the implementation of the policies to promote competitiveness and economic growth, a large number of small-sized firms have emerged, as we see in Tables 1 and A-1. The former details the change over the period, while the latter shows the year-by-year change in size groups. On the other hand, the number of middle-sized firms is relatively low, compared with the numbers in the other size categories.[1] In addition, the number of firms who grow in size to become middle-sized throughout the period in the sample is very small. Table A-1 details the number of transitional firms that changed their size category from one year to the next between 2000 and 2016 in our sample. Compared with the

[1] The Vietnamese government classified a manufacturing firm by size as micro if it no more than 10 employees, small if the number of employees is between 11 and 200, middle if it has between 201 and 300 employees, and large if there are more than 301 employees. In 2018, it revised the thresholds; however, as our data set falls between 2000 and 2016, we use the previous thresholds.

number of firms in each size category, the number of firms that change their size category, irrespective of its direction of upward or downward, is remarkably small.

In addition, Figure 1 plots the firm size distribution in three years, 2000, 2008 and 2016. From the figure, the firm size distribution seems to be relatively stable over time. Thus, one can come to the conclusion from here that, despite the consistent increase in the number of firms in every size category, the aforementioned policies may not be that effective in promoting middle-sized firms.

Table 1: Total and Transitional Numbers of Firms, by Periods

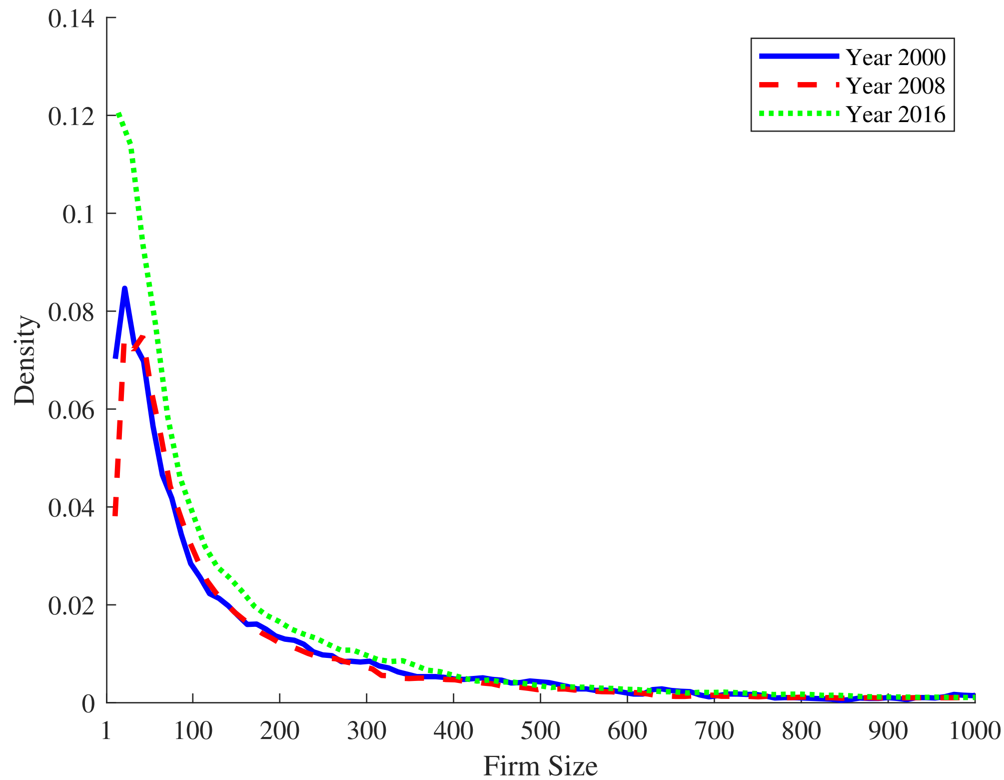
Period	2000 – 2016	2000 – 2008	2009 – 2016
# Continuing Firms	1541	2095	5282
Increase Size Group	257	349	466
Increase from Micro	40	81	40
Increase from Small	170	201	281
Increase from Middle	47	67	145
Decrease Size Group	295	205	713
Decrease from Small	79	44	242
Decrease from Middle	74	64	184
Decrease from Large	142	97	287
Unchanged Size Group	989	1541	4103

3 Data Description

We use the Enterprise Census data from the General Statistics Office of Vietnam (see Table A-2 for the details of our data), which defines an enterprise as "an economic unit that independently keeps a business account and acquires its own legal status." We focus on the period from 2000 to 2016. Table 2 lists the 15 manufacturing sectors we studied in this paper.[2] As we are focusing on changes in firm size, we only keep a firm if it has reported for at least 10 years of the sample period. In other words, we are examining the performance of continuing firms in terms of firm size, productivity

[2]These sector codes are based on the Vietnam's Standard Industrial Classification (VSIC) 1993. We also apply our empirical analyses on four other sectors, namely Agriculture (1), Forestry (2), Fishing (5) and Other mining and quarrying (14). However, as we are focusing on Vietnam manufacturing industries, these results are omitted.

Figure 1: Firm Size Distribution in Selected Years



and efficiency. In addition, we also remove all firms who reported negative or zero values for their revenues and spending.

The 15 sectors studied in this paper are the highest contributors to Vietnam's total manufacturing output. For instance, the total revenues from these sectors account for approximately 83% and 75% of Vietnam's total manufacturing output in 2010 and 2016, respectively. In addition, by nature these sectors are also quite diverse in terms of input intensities, market structures and technologies. Therefore, we can argue that our results provide a good representation of the Vietnamese manufacturing industries as a whole, especially in terms of continuing firms' size distributions, efficiencies and productivity.

In our analyses, the size of a firm is the total number of workers at the end of the fiscal year. Labor is the total income of employees in a firm including total wages and other employee-related costs such as social security and insurance. Intermediate materials include raw materials and other costs such as fuel. Capital is the total assets used in production. All variable values are adjusted for inflation. The annual inflation data are from the World Bank. Table 3 provides the summary statistics of the firm size for each sector.

Table 2: ISIC Description

VSIC	Description	VSIC	Description
15	Food products and beverages	17	Textiles
18	Wearing apparel, etc.	20	Wood and wood products
21	Paper and paper products	22	Publishing, printing, etc.
24	Chemical products, etc.	25	Rubber and plastic products
26	Cements and nonmetallic mineral products	27	Iron and steel products
28	Fabricated metal products	29	Machinery and equipment, etc.
31	Batteries, electrical cables and motors	35	Motorbikes, cars and ships
36	Furniture		

Table 3: Firm Size Summary Statistics

Sector	# Observations	Mean	Median	SD
15	17747	255.31	77	510.58
17	5521	326.24	118	655.11
18	7781	833.66	415	1243.47
20	5858	123.97	55	221.41
21	6357	108.90	54	157.24
22	4469	88.16	42	127.13
24	6783	173.09	69	318.87
25	8585	179.83	76	302.90
26	10353	218.92	105	347.07
27	2428	176.10	51	615.70
28	9746	137.62	50	299.26
29	3126	153.42	65	259.54
31	2258	477.02	114	1140.70
35	2097	434.42	157	874.95
36	6909	410.44	175	697.86

4 Empirical Methodologies

4.1 Stochastic Frontier Approach

In this section, we adopt the Bayesian procedure for estimating the random and fixed coefficient stochastic frontier model from O'Donnell and Coelli (2005) and Feng and Zhang (2014). The procedure is explained in details in Appendix A. The Bayesian procedure can utilize prior information to update the posterior estimation, while the SF method allows for statistical noise resulting from events outside the firm's control as well as other types of specification error and the omitted variable problem.

We estimate the following production function:

$$\ln y \quad = \quad a_0 + \textstyle\sum_n^3 b_n \ln x_n + \sigma_t t + u + v, \tag{1}$$

where u measures *production efficiency*, y denotes quantity produced, x_n denotes labor, capital and intermediate inputs for $n = 1, 2, 3$, respectively.

We use both the fixed and random coefficient models. In particular, we allow the coefficients of the standard fixed coefficient SF model to vary by decomposing the firm-specific coefficients into two parts: the mean vector and the random vector. The random vector is drawn from a multivariate normal distribution of mean zero. The fixed coefficient model can be obtained easily from the random model by excluding the random components.

In general, the advantage of using the random coefficient model over the fixed version is that it provides a better approximation of the underlying heterogeneity of technologies across different firms and thus allows us to estimate the returns to scale of each firm more accurately. With this method, we can obtain, for each firm, a separate production frontier and thus estimate the returns to scale of each firm according to its own frontier. This method is especially useful when there exists a high degree of heterogeneity across observations. To determine the relevance of both models, we compute both the deviance information criterion (DIC) and the Spearman rank correlation, which is used to test the association between two ranked variables without requiring knowledge of the joint probability distribution of the variables. We test the rank correlation of the firm-specific efficiency scores between the fixed and random coefficient models.

To estimate the model using the Markov Chain Monte Carlo (MCMC) method, we use the Gibbs sampling algorithm, which draws from the joint posterior density by sampling from a series of conditional posteriors. Once draws from the joint distribution have been obtained, any posterior feature of interest can be calculated. When estimating the model for each sector, we discard the first 10,000 as a *burn-in* and then repeat the procedure 400,000 times with checking convergence using a scale reduction factor (SRF). All the results are based on the last 200,000 times after ensuring convergence.

Regarding the value of the prior, following Van den Broeck et al. (1994), Kleit and Terrell (2001), Bezemer et al. (2005) and Balcombe et al. (2006), we use 0.875 as an informative prior for all sectors. As shown from our results from the SRF and simulation inefficiency factor (SIF), our MCMC converges well to a stable distribution and the sample size in each sector is large.

Finally, we compute RTS using (17) and categorize firms into three groups, namely, Increasing Returns to Scale (IRS), Decreasing Returns to Scale (DRS) and Constant Returns to Scale (CRS) by examining whether the 95% credible intervals are strictly greater than 1.0 (IRS), strictly less than 1.0 (DRS) or include 1.0 (CRS). These definitions are also adopted in Feng and Zhang (2014).[3]

4.2 Two-Stage GMM

In this section, we first explain the motivation for the two-stage GMM model from Ackerberg et al. (2015), which eliminates the simultaneity bias. Consider the logarithm form of a Cobb-Douglas production function for a firm i at time t given by

$$y_{it} = \beta_{it} + \beta_k k_{it} + \beta_l l_{it} + \epsilon_{it}, \tag{2}$$

where y_{it}, k_{it} and l_{it} are the logs of output, capital input and labor input, respectively and $\beta_{it} = \beta_0 + \omega_{it}$. In the regression (2), as the residual term, we obtain ω_{it} and ϵ_{it}, which represent observed and unobserved productivity shocks to the firm before it makes a decision on inputs. As described in Ackerberg et al. (2015), the productivity shock term ω_{it} captures expected changes to productivity such as "the managerial ability of a firm, expected down-time due to machine breakdown, expected defect rates in a manufacturing process, soil quality, or the expected rainfall at a particular farm's location." The exponential of β_{it} is the firm-specific TFP value examined in the literature.

However, there exists an endogeneity issue between production inputs k_{it}, l_{it} and the observed productivity shock ω_{it}. As a profit maximizer, a firm's optimal choice of inputs depends on its productivity shock ω_{it}, which is assumed to be observable to the firm but not to us. For instance, if a steel manufacturing company knows that its production is to be halted for a few days for maintenance

[3] For a more detailed discussion, see Caves et al. (1982) and Färe and Grosskopf (1994).

on its machinery, it is expected to reduce the labor input l_{it} accordingly, resulting in a reduction of revenue y_{it}. However, we could not see the correlation between ω_{it} and l_{it}, and thus may attribute the reduction of y_{it} to a lower value of β_l. Thus the input coefficients are biased and the productivity value is incorrectly estimated.

In general there are two approaches to solve the simultaneity bias, which are using dynamic panel data such as Blundell and Bond (2000) and using proxy variables. The second approach is more commonly employed in the literature including Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg et al. (2015), Wooldridge (2009) and Kasahara et al. (2015). We focus on the latter approach.

To eliminate the bias, Olley and Pakes (1996) assume the labor input to be non-dynamic, i.e. the current labor input does not determine future profits. On the other hand, capital is dynamic through investment and determined by the firm; thus capital is determined from the previous period and as such uncorrelated with the productivity term ω_{it}. To avoid collinearity between l_{it} and ω_{it}, the authors assume investment to be strictly increasing in productivity. Another approach by Levinsohn and Petrin (2003) uses the intermediate input demand function instead of investment as a proxy for capital input. The idea behind this is that it is more plausible for intermediate inputs to be strictly increasing in productivity.

Ackerberg et al. (2015) build upon Levinsohn and Petrin (2003) and retain the specifications where the proxy variable is the intermediate materials and the capital stock is dynamic and decided in the previous period. On the other hand, they assume that labor is also a determinant of intermediate input. Thus the coefficient for labor is estimated in the second stage of the GMM method, in contrast to the first stage in Levinsohn and Petrin (2003). We can find the coefficients for labor and capital inputs, as well as the firm-specific productivity term from this model. The assumptions for the model together with the procedure are described in Appendix B.

In addition, it should be noted here that the TFP value from this approach is very different from those of the SF methodology. The main reason is that this proxy variable strand of productivity literature assumes no inefficiencies in the production process. In other words, all firms are operating on their respective frontiers. Any changes in ω_{it} result in the movement of the frontier itself, rather than the movement of a firm to below the frontier.

Finally, note that as examined in Foster et al. (2008) and Katayama et al. (2009), there exist two different measures of productivity, which are physical and revenue productivity values. Consider the typical situation where a firm uses capital, labor and intermediate materials to produce goods, all of which are measured in monetary terms. We use real revenue as a proxy for output, which can result in a different production efficiency measure than our desired physical production efficiency value. To overcome this problem, first, we check the calculated Herfindahl–Hirschman Index (HHI) for the sectors we study[4]. The resulting HHIs are well below 1,500, which implies low market concentrations. We further conduct a robustness check using various markups. Second, to make sure that our revenue productivity results are consistent with the physical ones, in our robustness check, we also allow for various degrees of markup whereby the price increase is proportional to a firm's revenue. We find no significant change in our conclusion.

5 The Empirical Results

5.1 Summary

In this section, we describe the findings, whose key points are summarized below:

- In all sectors, there are many firms whose efficiency scores are low across different firm sizes.
- In all sectors, production efficiencies are most variable in the middle-sized firm group, and the firms with the lowest production efficiency are middle-sized.
- Most size groups in all sectors have constant returns-to-scale technologies.
- There is significant heterogeneity in firms' technologies across different size groups.
- Firms' total factor productivity is also quite diverse, and the trend of the average TFP value is either constant or monotone decreasing as the firm size increases.

To keep the presentation of figures as clear as possible, the results for six sectors (sectors 15, 18, 25, 26, 28 and 36) are shown in this paper. These sectors are chosen as they have the highest number of observations in the sample. In addition, they are also within the highest contributors to the total

[4]Assume a market with N firms, where s_i is the market share in percentage of firm i. Then, the HHI is:

$$HHI = \sum_{i=1}^{N} s_i^2.$$

The index can range from close to zero to 10,000.

output of all Vietnamese manufacturing industries. For instance, between 2010 and 2016, these six sectors contribute from 44% to 48% of Vietnam's total output from manufacturing industries.

The results and graphs for the other sectors are available upon request. The estimated coefficients from the SF model are listed in Table A-3 (see the Appendix) together with the SRF and SIF for each coefficient, and DIC statistics. The first two factors help us check the convergence of MCMC, while the last factor helps us choose which model (random or fixed effect) is more appropriate for our analysis.

Throughout all sectors, the DIC scores are consistently and significantly negative, confirming the superiority of the random effect model over the fixed effect one. Furthermore, we study the penalty of using the fixed effect model by calculating the Spearman coefficient. Table 4 shows how different each group's Spearman coefficient is, compared with the one for the whole sector in the rankings of efficiency scores. In this table, to group firms, we use the same categorization by the Vietnamese government (see Footnote 1). We find that there are some positive correlations, while the numbers vary depending on the group and sector. In addition, across the size group and sector, in terms of efficiency scores, the results of Spearman coefficients for each group are consistent with that of the total sector. Overall, there is very small correlation of the rankings between the two model's efficiency scores. Given that in the fixed-effect model, the error term includes both efficiencies and actual errors in the random-effect model, this indicates not considering the firm-specific heterogeneity can lead to mis-specification of inefficiencies.

Figure A-2 presents a scatter plot of the efficiency scores computed using the SF method for some of our sample industries for the entire sample period, and Figures A-3 and A-4, respectively, show a scatter plot of average efficiency scores computed using the SF method for each size under the fixed-effect and random-effect models. We further study what portion of the lower efficiency group of each sector belongs to the middle-sized group. In most sectors, the proportion of middle-sized firms in the low-efficiency group is higher than the ratio of the middle-sized firms relative to all observations. In addition, Figure A-3 shows the larger-sized firms are estimated to be significantly more efficient than their peers, which we do not observe in Figure A-4. Finally, Figure A-5 presents a scatterplot of RTS computed using the SF method for each size. Our analysis shows that the majority of most size groups in all sectors have constant returns-to-scale technologies.

Table A-4 details the coefficients for labor and capital inputs using the two-stage GMM from Ackerberg et al. (2015). It should be noted here that due to the requirement of a balanced panel for

Table 4: Comparison of Efficiency by Spearman Coefficient

Sector	Group 1	Group 2	Group 3	Group 4	Total
Sector 15	0.3231	0.2610	0.3541	0.2139	0.3063
Sector 18	0.4286	0.3328	0.3622	0.3119	0.3334
Sector 25	0.1203	0.2947	0.4086	0.3190	0.3388
Sector 26	0.8080	0.2847	0.1703	0.1160	0.2730
Sector 28	0.3565	0.3122	0.3566	0.2799	0.3517
Sector 36	0.5588	0.4106	0.3450	0.3826	0.3434

Note: Groups 1 to 4 consist of micro-sized (between 1 and 10 employees), small-sized (between 11 and 200 employees), middle-sized (between 201 and 300 employees), and large-sized (more than 301 employees) firms, respectively.

this methodology, we split the sample into two separate periods to achieve the optimal number of observations as well as to provide comparative analysis. The first period is between 2000 and 2008, and the second period is from 2009 to 2016. Another reason for this separation is to negate the effect of the Global Financial Crisis of 2008. Comparing between the two periods 2000 – 2008 and 2009 – 2016, with the exception of sector 18, all sectors experience a significant increase in capital intensity in the latter period. In addition, most industries exhibit constant returns to scale technology.

Finally, the distributions of the TFP across firm sizes are plotted in Figures A-6 and A-8 for two periods. As seen from the two figures, the productivity distribution greatly varies across sectors and firm sizes. The main thing to note is that the trend for the average TFP is either constant or decreasing as firm size increases. For instance, while the trend is mostly constant for sector 15 for both periods, we can see a very high variation of the TFP distribution across firm size, especially in the 2009 – 2016 period. In addition, we also see a very clear lower average TFP value for the middle-sized firms in sector 18. As a whole the period 2009 – 2016 has a higher variation of the TFP values than the 2000 – 2008 period.

The high variation and the decreasing trend of the TFP values become clearer for many sectors once we control for the time trend. Figures A-7 and A-9 plot the smoothed TFP values across firm size and time. First, for both periods, given the same year, there seem to be a decreasing tendency of productivity as firm size increases for most industries, with the exception of sector 25 between 2009 and 2016. In addition, the TFP growth values seem to vary sector by sector; for instance, the smoothed values for sectors 15 and 26 are increasing in years. The interesting case is sector 18: while there is an increasing trend of productivity between 2009 and 2016, it instead seems to stay constant or falling during the previous period. Recall that sector 18 is the apparel manufacturing industry, which is one of the main exports of Vietnam. As such, the effects of the GFC appear pronounced in this sector.

5.2 Robustness

To confirm the robustness of our results, we perform four different sets of exercises and explain each procedure. The first two exercises check the robustness of the SF approach, while the latter two test our two-stage GMM approach. We find that the results detailed above still hold in these analyses of different settings.

5.2.1 Trans-log functions

We redo the stochastic frontier approach using trans-log functions, instead of the Cobb-Douglas production functions in the main analysis. This allows interactions between inputs and is widely used in the literature. Instead of equation (1), the estimated trans-log production functions follows

$$\begin{aligned}\ln y =& a_0 + \sum_n^3 b_n \ln x_n + \frac{1}{2}\sum_n^3\sum_j^3 b_{nj}\ln x_n \times \ln x_j \\ &+ \sigma_t t + \frac{1}{2}\sigma_{tt}t^2 + \sum_n^3 \rho_n t \ln x_n + u + v.\end{aligned}$$

Throughout all sectors, we still see the high variation of the efficiency score in the middle-sized groups, and that the majority of firms display constant return-to-scale technology.

5.2.2 Firms' characteristics

We group firms based on their locations and ownership types and study whether our findings are still robust. The results suggest our findings are robust. The groups of locations consist of nine groups: northern midland and mountainous area, Red River Delta, north-central coast, south-central coast, central highlands, southeast area, Mekong River Delta, Hanoi and Ho Chi Minh City. The ownership types we consider are the following three types: state-owned firms, private firms and foreign firms. We still find that for the middle-sized groups, the efficiency scores are relatively diversified across firms and regions, and different ownership types still remain.

5.2.3 Different Periods

As a robustness check for the two-stage GMM model, we perform the analysis using different periods. Apart from the two intervals in the main analysis, we focus on two other periods with less length, which are 2004 – 2008 and 2009 – 2014. An advantage of reducing the period length is that we are able to capture more firms than the main analysis, either because a firm missed out a report in a year, or because it did not exist, or already closed down outside this smaller length. In addition, we also perform the analysis on the period 2005 – 2011, which also captures the effect of the Global Financial Crisis. We still see the high labor intensity relative to the capital intensity, and also the average TFP trend is monotone throughout the period.

5.2.4 Markups

We study whether the results are robust to various rates of markups in pricing and adopt the idea from Martins and Scarpetta (2002) and Bartelsman et al. (2013) who used a markup of 25 percent. We are interested in each firm n's physical productivity ν_n given their revenue productivity μ_n, which we obtain from the SF approach. We estimate μ_n using the relationship $R_n = \mu_n f_n(k_n, l_n, i_n)$ where R_n is firm n's revenue, f_n is its production function, while k_n, l_n and i_n denote firm n's capital, labor and intermediate inputs, respectively. Let q_n and p_n denote firm n's quantity produced and price, respectively. Then, $q_n = \frac{\mu_n}{p_n} f_n(k_n, l_n, i_n)$, and thus $\nu_n = \frac{\mu_n}{p_n}$.

Let $\bar{p}$ denote a base price. Consider two firms, firms 1 and n. When firm 1's price $p_1 = (1 + m \cdot R_1) \cdot \bar{p}$ where m is a markup, the physical productivity ν_1 can be expressed as $\frac{\mu_1}{(1+m \cdot R_1)\bar{p}}$ by the argument in the previous paragraph. Then, the physical productivity ratio between firm 1 and firm n can be expressed as

$$\frac{\nu_1}{\nu_n} = \frac{(1 + m \cdot R_n)}{(1 + m \cdot R_1)} \cdot \frac{\mu_1}{\mu_n}. \tag{3}$$

Although we cannot directly observe the base price $\bar{p}$, we can eliminate it by studying the ratio as in (3). We compared the plots of physical productivity ratios for $m = 0, 0.1, 0.15, 0.2$ and 0.25 by setting ν_n to be the physical productivity of firm n who has the highest revenue productivity. As a large firm's score goes down more as m increases, the monotone trend of the average TFP still persists.

6 A Further Investigation of Our Findings

6.1 Discussion

In this section, we further investigate our findings by developing a theoretical explanation to show how a firm's decision of whether to grow in size differs under different distributions of efficiency. The efficiency term from our SF approach measures how far away a firm is operating from its production frontier. In this section we propose a model to study how the difference in these production conditions could affect a firm's decision. In general, we find that as a firm's production efficiency becomes more varied, it is less likely that the firm expands in size.

Ever since the proposition of Gibrat's Law in Gibrat (1931), there has been a plethora of literature on firm size distribution and firm size growth. A particular strand of literature, including Cooley and Quadrini (2001), Cabral and Mata (2003), Albuquerque and Hopenhayn (2004), Clementi and Hopenhayn (2006) and Alfaro et al. (2019) focus on financing constraints as one of the reasons for small-sized firms being unable to expand. Given that financing constraints are more prevalent in developing countries and Vietnam in particular, we envisage that this issue might be related to our empirical result that the variation of the efficiency term is highest in the middle-sized group.

In what follows, we demonstrate how the variation of the efficiency term affects the decision of the investor within a theoretical framework. More specifically, between, say, a large firm and a small one, both of whom wish to borrow money from the investor for expansion, we show that the high variation in efficiency that firms face if they expand their business can act as a disincentive for investors to lend money. Going back to our example, between a large-sized firms whose efficiency score is less uncertain than a small-sized firm trying to expand and become middle-sized, it can easily be seen that the investor prefers to invest in the large firm. Our theoretical framework models the inefficiency as a firm's management cost.

We model the distance from the production frontier, which our empirical analysis from the SFA measures, as a management cost. Then, to model the variation of a firm's efficiency score, we use the concept of a *mean-preserving spread* from Rothschild and Stiglitz (1970). First, we define this term. Let x be a random variable with distribution X. Construct another random variable y such that

$$y = x + z,$$

where $\mathbb{E}[z|x] = 0$ for all x. Let Y be y's distribution. Then, Y is a *mean-preserving spread of* X.

In words, one distribution is a mean-preserving spread of another if they have the same mean and the difference is the random noise whose conditional mean is 0. Then, X is considered less risky and is always preferred to Y by risk-averse individuals (see Rothschild and Stiglitz, 1970).

6.2 Theoretical Analysis

We study the situation where one investor consider investing to two different types of firms where the distribution of one firm's (denoted by firm 1) management cost is more spread than the one of the other (denoted by firm 0). We assume that the investor is risk averse. The firm needs funds from the investor to expand their business and hire an extra worker. The firm is endowed with wealth ω, which can be used as the collateral for the loan from the investor. The original profit of the firm is given as

$$\bar{\pi} = aL^{\beta}K^{1-\beta} - wL - rK, \tag{4}$$

where L and K are its current labor and capital inputs, respectively; w is the wage and r the cost of capital. We assume all four variables to be exogenous, as the firm is currently in operation.

If a firm expands, it experiences a transaction cost of m. The cost m is stochastic and its mean is given by M. Suppose that M has the cumulative distribution function F. In addition, assume that both the firm and the investor does not know the value of m until the firm has hired the extra worker; they only know its distribution. With probability q, the project fails and does not produce enough profits. Then the firm fails. In this case, the firm goes bankrupt and the investor only gets ω as collateral. Assume q is exogenously given.

The investor's decision problem is to decide whether to lend the amount p to the firm as well as the associated interest rate i, using the firm's wealth ω as collateral to one of the two types of firms. Suppose p is exogenously given. The profit if a firm expands is

$$\pi_M = q\left(a(L+1)^{\beta}K^{1-\beta} - i \cdot p\right) - w(L+1) - rK - M - (1-q)\omega. \tag{5}$$

We assume the investor's utility u to be a function of his wealth, and is strictly concave. If the investment is successful and the firm continues to operate, he receives both the principal and interest of the investment amount. On the other hand, if the firm experiences a negative profit due to the expansion, it closes and the investor seizes ω. Then, if the investor lends the money to the firm, his expected utility is

$$q \cdot u(i \cdot p) + (1-q) \cdot u(\omega), \tag{6}$$

while if he does not lend the money, his utility is $u(p)$.

Suppose that the market is perfectly competitive and then the investor offers an interest rate such that $\pi_M = \bar{\pi}$. Let F_j be a cumulative distributions of firm j's management cost M for each $j = 0, 1$. Let I_j be the cumulative distribution of i_M that directly results from the distributions of M in F_j for each $j = 0, 1$. The following lemma looks at the relationships between the distribution functions.

Lemma 1. *Suppose that F_1 is a mean-preserving spread of F_0. Then I_1 is also a mean-preserving spread of I_0.*

Proof. Using Rothschild and Stiglitz (1970), F_1 is a mean-preserving spread of F_0 if and only if its associated costs M_1 and M_0 follows, respectively

$$M_0 = M_1 + \epsilon,$$

where $\mathbb{E}[\epsilon|M_1] = 0$. We consider an interest rate such that π_M equals $\bar{\pi}$. Given the realized M, this is the maximum interest rate that makes the firm to be willing to borrow funds.

$$i_M = \frac{aK^{1-\beta}\left[q(L+1)^{\beta} - L^{\beta}\right] - w - (1-q)\omega - M}{q \cdot p}. \tag{7}$$

Using equation (7), the two interest rate variables i_1 and i_0 under the two cumulative distributions I_1 and I_0, respectively, also follow

$$i_0 = i_1 - \frac{\epsilon}{q \cdot p}. \tag{8}$$

Since $\mathbb{E}[\epsilon|M_1] = 0$, it follows that $\mathbb{E}[\frac{\epsilon}{q \cdot p}|i_1] = 0$. Thus I_1 is a mean-preserving spread of I_0. □

The following proposition extends the lemma to consider the investor's decision under two different distributions of the transaction cost.

Proposition 1. *Suppose that F_1 is a mean-preserving spread of F_0. Then,*

$$\mathbb{E}_{I_0}[u(p \cdot i_M)] \geq \mathbb{E}_{I_1}[u(p \cdot i_M)].$$

Proof. We have shown in the previous lemma that if F_1 is a mean-preserving spread of F_0, I_1 is also a mean-preserving spread of I_0. Then, for the risk averse investor, it follows that

$$\mathbb{E}_{I_0}[u(p \cdot i_M)] \geq \mathbb{E}_{I_1}[u(p \cdot i_M)]. \tag{9}$$

□

For given p, we have shown that the risk-averse investor would prefer firm 0 over firm 1 due to less uncertainty in management costs. Then the probability that the investor invests to firm j for $j = 0, 1$ equals to the probability that the following holds:

$$q\mathbb{E}_{I_j}[u(p \cdot i_M)] + (1-q)u(\omega) \leq u(p). \tag{10}$$

From Proposition 1, for any given p, the probability that (10) holds for $j = 0$ is higher than the one for $j = 1$. Therefore, we conclude that the investor is more likely to invest in firm 0 than in firm 1 even though their expected cost is the same.

Finally, we note that in our empirical analysis, we have found that the productivity term a (in our theoretical analysis) is decreasing with size. Here even when we assume that it is constant, the investor is more likely to invest in smaller firms. It is easy to see that when a decreases with firm size, this tendency becomes stronger. In the empirical analysis, we have also observed a higher variation of TFP term in most sectors. We can show that a higher variation of a would also have a negative influence to a risk averse investor's decision to expand their firm's size. We provide another simple model with heterogenous productivity and analyze the decision making by a risk averse manager in Section C of the appendix. The key point is that under the empirical features we have found about the relationship between productivity or efficiency and firm sizes, our theoretical investigation even in a simple setting indicates the existence of unwillingness for small firms to expand their size.

7 Concluding Remarks

Recently, there is also an increasing interest in measuring productivity in developing countries (see Tybout, 2000; Hsieh and Klenow, 2009; Bloom et al., 2010). Our paper contributes to this line of literature that studies the relationship between production efficiency and firm size using the data from Vietnam, a transitional developing country. We showed how heterogeneous are production efficiency and technologies across sizes using a recently developed method.

Our empirical analysis shows that there is great variation in productivity across firm sizes in all sectors, while efficiency is more diverse mostly in the middle-sized firm group. Further, Proposition 1 in our theoretical analysis shows that a more diverse distribution of the efficiency could be a risk for an investor, who in turn is not willing to provide financial loans for firms to expand their business. By using the U.S. data, Dhawan (2001) categorized firms into two groups, either small or large, and showed that small-sized firms face more risks. In this sense, our results add to the findings of Dhawan

(2001) by showing that more diverse efficiency score could be another risk factor for the firm, although it appears that this is more relevant to the middle-sized firms rather than the small-sized firms in the case of Vietnam.

Our study in this paper poses future research questions. First, in the literature, some empirical studies show that resource misallocation can also be an important factor to influence economic growth and aggregate productivity such as Hsieh and Klenow (2009), Jones (2011), and Bento and Restuccia (2017). For instance, Hsieh and Klenow (2009) show that misallocation elimination can improve total factor productivity by 30–50% in China, 40–60% in India. Bento and Restuccia (2017) also estimate the loss in aggregate outputs and average firm size of 53% and 86%, respectively, once the productivity elasticity of distortions increases from 0.09 in the United States to 0.5 in India. As future research, it is interesting to study if Vietnamese firms have the problem of resource misallocation.

In the literature on firm-size distributions, one of the issues that has attracted attention recently is the "missing middle" phenomenon, which refers to the empirical fact that most employment in developing countries is located in either small-sized or large-sized firms. It was first documented by Liedholm and Mead (1987), and there is an ongoing debate on whether this phenomenon exists as well as what its underlying causes are (see, for example, Little, 1987; Tybout, 2000; Hsieh and Olken, 2014) [5]

If the firm-size distribution in developing countries shows a different feature from the one in advanced countries, we would envisage that there are some reasons for this that apply to developing countries in particular but less so to developed countries. One possible reason, as explained in Tybout (2000), is that strong business regulation could be related to the existence of too many small firms. Intuitively, small firms are reluctant to grow in size because they would be subject to more regulations or hassling from the officials. A direct consequence of that is the variation of the productivity, as explained in Saastamoinen and Kuosmanen (2014).

Second, because our analysis is a snapshot of the Vietnamese manufacturing industry, it would be worthwhile to see why the middle-sized groups' efficiencies vary more than the other groups. In addition, one may ask whether our findings from a transition economy, Vietnam, provide a reasonable foundation for generalization because the business environment and its features vary significantly across countries. Comparison of our findings with the same analyses by using other countries' data

[5] In relation to this literature, we have also conducted bimodality coefficients tests and Dip tests (the details are available upon request). Our sample rejects the null hypothesis of unimodality.

would be interesting.

Finally, a natural question to ask is: why do we observe such variations in productivity and efficiency in the middle-sized firms? Empirically investigating causes and theoretically developing a model that induces this phenomena in equilibrium are interesting venues for future research. Such analyses would provide a deeper understanding on what brings out the different firm size distributions in different countries and environments.

Appendices

A Stochastic Frontier Methodology

Let $\ln y = q_{it}$ for some i, t, and we rewrite (1) as:

$$q_{it} = \mathbf{x}_{it}\boldsymbol{\beta}_i + u_{it} + v_{it},$$

where $\boldsymbol{\beta}_i$ follows a multivariate normal distribution: $\boldsymbol{\beta}_i \sim i.i.d.N(\bar{\boldsymbol{\beta}}, \boldsymbol{\Omega})$. In addition, an error term $v_{it} \sim i.i.d.N(0, \sigma_v^2)$ and a firm-specific production efficiency $u_{it} \sim i.i.d.\exp(\lambda^{-1})$.

Decomposing $\boldsymbol{\beta}_i$ into its mean vector of $\bar{\boldsymbol{\beta}}$ and a random vector $\boldsymbol{\delta}_i$ yields

$$\boldsymbol{\beta}_i = \bar{\boldsymbol{\beta}} + \boldsymbol{\delta}_i,$$

where $\boldsymbol{\delta}_i \sim i.i.d.N(\mathbf{0}, \boldsymbol{\Omega})$.

Given these specifications, we can rewrite the SF model as

$$q_{it} = \mathbf{z}_{it}'\bar{\boldsymbol{\beta}} + \mathbf{z}_{it}'\boldsymbol{\delta}_i + u_{it} + v_{it},$$

where u_{it} and v_{it} are defined as above. When $\boldsymbol{\delta}_i = \mathbf{0}_{s\times 1}$, the random coefficient SF model reduces to the fixed coefficient SF model.

We now describe the Bayesian procedure for estimating the random coefficient SF model. We use a flat prior for $\bar{\boldsymbol{\beta}}$, as

$$p(\bar{\boldsymbol{\beta}}) \propto 1, \tag{11}$$

and the following distribution for $h_v \equiv \frac{1}{\sigma_v^2}$,

$$p(h_v) \propto \frac{1}{h_v}. \tag{12}$$

The prior for u_{it} can be written as

$$p(u_{it}|\lambda^{-1}) = f_{Gamma}(u_{it}|1, \lambda^{-1}).$$

To obtain the prior, we use:

$$p(\lambda^{-1}) = f_{Gamma}(\lambda^{-1}|1, -\ln \tau^*).$$

For priors of $\boldsymbol{\delta}_i$ and $\boldsymbol{\Omega}$, we use the following:

$$\boldsymbol{\delta}_i \sim i.i.d.N(\mathbf{0}, \boldsymbol{\Omega}) \text{ and } \boldsymbol{\Omega} \propto IW(\mathbf{1}, \boldsymbol{\Omega}), \tag{13}$$

where $IW(1, \boldsymbol{\Omega})$ is an inverse Wishart density with 1 degree of freedom and scale matrix $\boldsymbol{\Omega} = 10^{-6} I_s$.

Let K be the total number of firms in one sector of the data set and T_i be the set of time periods for which firm i's sample is taken in one sector of the data set. Let N be the total number of observations in the data set. The likelihood function is

$$\begin{aligned} \mathcal{L}(q|\bar{\boldsymbol{\beta}}, h_v, u, \lambda^{-1}, \boldsymbol{\delta}, \boldsymbol{\Omega}) &= \prod_{i=1}^{K} \prod_{t \in T_i} \left[\sqrt{\frac{h_v}{2\pi}} \exp \left(q_{it} - \mathbf{z}'_{it}\bar{\boldsymbol{\beta}} - \mathbf{z}_{it}\boldsymbol{\delta}_i - u_{it} \right)^2 \right] \\ &\propto (h_v)^{\frac{N}{2}} \exp \left[-\frac{h_v}{2} \mathbf{v}'\mathbf{v} \right], \end{aligned} \tag{14}$$

where $\mathbf{v}$ is an $N \times 1$ vector such that $\mathbf{v} = q_{it} - \mathbf{z}'_{it}\bar{\boldsymbol{\beta}} - \mathbf{z}_{it}\boldsymbol{\delta}_i - u_{it}$.

By Bayes' Theorem, using the likelihood function (14) and the prior distributions (11) to (13), we obtain the posterior joint density function

$$\begin{aligned} f(\bar{\boldsymbol{\beta}}, h_v, \mathbf{u}, \lambda^{-1}, \boldsymbol{\delta}, \boldsymbol{\Omega}|\mathbf{q}) \propto h_v^{N/2-1} \exp \left[-\frac{h_v}{2} \mathbf{v}'\mathbf{v} \right] \prod_{i=1}^{K} \left[|\boldsymbol{\Omega}|^{-1/2} \exp \left(-\frac{1}{2} \boldsymbol{\delta}'_i \boldsymbol{\Omega}^{-1} \boldsymbol{\delta}_i \right) \right] \\ \times \prod_{i=1}^{K} \prod_{t \in T_i} \left[\lambda^{-1} \exp(-\lambda^{-1} u_{it}) \exp(\lambda^{-1} \ln \tau^*) \times |\boldsymbol{\Omega}|^{\frac{t+s+1}{2}} \exp \left[-\frac{1}{2} \mathrm{tr}(\boldsymbol{\Phi}\boldsymbol{\Omega}^{-1}) \right] \right]. \end{aligned} \tag{15}$$

Then, the system of the full conditional posterior distribution in the random-effect analysis is:

$$\begin{aligned} p(\bar{\boldsymbol{\beta}}|\mathbf{q}, h_v, \mathbf{u}, \lambda^{-1}, \boldsymbol{\Delta}, \boldsymbol{\Omega}) &\propto f_{Normal}(\bar{\boldsymbol{\beta}}|\bar{\boldsymbol{b}}, h_v^{-1}, (\mathbf{z}\mathbf{z}')^{-1}) \\ p(h_v|\mathbf{q}, \bar{\boldsymbol{\beta}}, \mathbf{u}, \lambda^{-1}, \boldsymbol{\Delta}, \boldsymbol{\Omega}) &\propto f_{Gamma}(h_v|\tfrac{N}{2}, \tfrac{1}{2}\mathbf{v}'\mathbf{v}) \\ p(y_{it}|\mathbf{q}, \bar{\boldsymbol{\beta}}, h_v, \lambda^{-1}, \boldsymbol{\Delta}, \boldsymbol{\Omega}) &\propto f_{Norma}(u_{it}|q_{it} - \mathbf{z}'_{it}\bar{\boldsymbol{\beta}} - \mathbf{z}'_{it}\boldsymbol{\delta}_i - (h_v\lambda)^{-1}, h_v^{-1}) \cdot I(u_{it} \geq 0) \\ p(\lambda^{-1}|\mathbf{q}, \bar{\boldsymbol{\beta}}, h_v, \mathbf{u}, \boldsymbol{\Delta}, \boldsymbol{\Omega}) &\propto f_{Gamma}(\lambda^{-1}|N+1, \mathbf{u}'\boldsymbol{\iota}_N - \ln \tau^*) \\ p(\delta_i|\mathbf{q}, \bar{\boldsymbol{\beta}}, h_v, \mathbf{u}, \lambda^{-1}, \boldsymbol{\Omega}) &\propto f_{Normal}(\boldsymbol{\delta}_i|\bar{\boldsymbol{\delta}}, (\boldsymbol{\Omega}^{-1} + h_v\mathbf{z}\mathbf{z}'_i)^{-1}) \\ p(\boldsymbol{\Omega}|\mathbf{q}, \bar{\boldsymbol{\beta}}, h_v, \mathbf{u}, \lambda^{-1}, \boldsymbol{\Delta}) &\propto IW(r+K, \boldsymbol{\Phi} + \boldsymbol{\Delta}\boldsymbol{\Delta}'). \end{aligned}$$

In the fixed-effect analysis, the above system excluding the last two conditions is applicable.

The *efficiency score* is given by $\exp(-u_{it})$ for each i, t.

The Spearman rank correlation is calculated as follows. For a sample of size n, let $\mathrm{rg}(X_i)$ denote the ranking of firm i's efficiency score using the fixed-effect estimation and $\mathrm{rg}(Y_i)$ denote the ranking of firm i's efficiency score using the random-effect estimation. Then, the Spearman correlation coefficient ρ_s is given by

$$\rho_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)},$$

where $d_i = \mathrm{rg}(X_i) - \mathrm{rg}(Y_i)$, is the difference between the two ranks of each observation and n is the number of observations. A Spearman correlation of zero indicates that there is no tendency for a variable to either increase or decrease when the other one increases. The Spearman correlation increases in magnitude as the two variables become closer to being perfect monotonic functions of each other.

Finally, we compute the RTS. To do so, we take the first derivative of $\ln y$ with respect to each $\ln x$. This gives us the measure of RTS. Now,

$$\frac{\mathrm{d}\ln y}{\mathrm{d}\ln x_i} = b_i + \rho_i t, \tag{16}$$

and RTS is given by

$$RTS_i = \sum_i^3 \frac{\mathrm{d}\ln y}{\mathrm{d}\ln x_i}. \tag{17}$$

B Two-Stage GMM Methodology

In this section, we explain the two-stage GMM procedure from Ackerberg et al. (2015), which is used to eliminate the simultaneity bias between productivity and the production inputs. Essentially, the model uses the intermediate materials as a proxy for the productivity term and the capital stock is dynamic and decided in the previous period. We can find the coefficients for labor and capital, as well as the firm-specific TFP term from this model. We begin with listing the assumptions for the model, after which the procedure will be explained in detail.

B.1 Assumptions

There are five assumptions for the model, which outline the nature of the dynamically optimizing firms.

Assumption 1 (Information Set). *A firm i's information set I_{it} follows $I_{it} = \{\omega_{it}, \omega_{it-1}, ..., \omega_{i0}\}$. In addition, for any firm i, the unobserved shocks are i.i.d., and, $E[\varepsilon_{it}|I_{it}] = 0$.*

This assumption implies that the firm i's information set consists of all the current and past observable productivity shocks related to the firm, and so all firms know some part of the productivity shocks that happen in every period up to the current period.

Assumption 2 (First-Order Markov)**.** *The productivity shock ω_{it} follows a first-order Markov process, that is,*

$$p(\omega_{it}|I_{it-1}) = p(\omega_{it}|\omega_{it-1}).$$

This distribution is known to firms and stochastically increasing in ω_{it-1}.

As a result of the previous two assumptions, a firm does not know its future productivity shocks; it only knows its distribution, which depends solely on the previous period's value.

Assumption 3 (Timing of Capital and Labor Choices)**.** *Firm i's labor input l_{it} is chosen at period t while its capital accumulation follows*

$$k_{it} = \kappa(k_{it-1}, i_{it-1}), \tag{18}$$

where investment i_{it-1} is chosen in period $t-1$.

Assumption 3 implies that a firm's capital is dynamic and its stock in period t is a result of the firm's previous capital stock k_{it-1} and investment decision from the previous period i_{it-1}. This is in direct contrast to the firm's labor decision, which is stationary. It should be noted here that Ackerberg et al. (2015) also allows for the case of dynamic labor decision; that is, l_{it} can also be chosen at period $t-b$, where $b \in [0, 1]$. However, because we are analyzing manufacturing industries, which arguably do not require extensive training or marketing for the jobs, we can relax the assumption and allow labor to be chosen during the same period.

Assumption 4 (Scalar Unobservable)**.** *Firm i's intermediate input demand at time t follows*

$$m_{it} = \widetilde{f}_t(k_{it}, l_{it}, \omega_{it}). \tag{19}$$

Thus, a firm's intermediate input is nondynamic and chosen at the same time period as production and depends on a firm's capital, labor stocks as well as its productivity level.

Assumption 5 (Strict Monotonicity)**.** *The demand function for intermediate input $\widetilde{f}_t(k_{it}, l_{it}, \omega_{it})$ is strictly increasing in ω_{it}.*

This assumption implies that the more productive a firm is, the more it wants to accumulate intermediate inputs and, through these, to increase production. It is required to invert the intermediate input demand function.

B.2 The procedure

Using Assumption 5, we can invert the input demand function to

$$\omega_{it} = \widetilde{f}_t^{-1}(k_{it}, l_{it}, m_{it}), \tag{20}$$

which gives productivity as a function of the capital, labor and intermediate inputs. We substitute this into the production function:

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \widetilde{f}_t^{-1}(k_{it}, l_{it}, m_{it}) + \varepsilon_{it} = \widetilde{\Phi}_t(k_{it}, l_{it}, m_{it}) + \varepsilon_{it}. \tag{21}$$

This gives the first stage moment condition

$$\mathbb{E}[\varepsilon_{it}|I_{it}] = E[y_{it} - \widetilde{\Phi}_t(k_{it}, l_{it}, m_{it})|I_{it}] = 0. \tag{22}$$

This stage can be simply represented as an OLS regression of the equation

$$y_{it} = \gamma_0 + \gamma_l l_{it} + \gamma_k k_{it} + \gamma_m m_{it} + \varepsilon_{it} = X_{it}\gamma + \varepsilon_{it}, \tag{23}$$

which we can estimate to find $\widehat{\widetilde{\Phi}}_t(k_{it}, l_{it}, m_{it})$.

In the second stage, we extend Assumption 2 and allow the productivity term to evolve in an AR(1) process according to the following equation:

$$\omega_{it} = \mathbb{E}[\omega_{it}|I_{it-1}] + \xi_{it} = \mathbb{E}[\omega_{it}|\omega_{it-1}] + \xi_{it} = g(\omega_{it-1}) + \xi_{it}, \tag{24}$$

where $\mathbb{E}[\xi_{it}|I_{it-1}] = 0$. This gives us the second stage moment condition:

$$\begin{aligned}\mathbb{E}[\xi_{it} + \varepsilon_{it}|I_{it-1}] =& \mathbb{E}[y_{it} - \beta_0 - \beta_k k_{it} - \beta_l l_{it} \\ & - g(\widetilde{\Phi}_{t-1}(k_{it-1}, l_{it-1}, m_{it-1}) - \beta_0 - \beta_k k_{it-1} - \beta_l l_{it-1})|I_{it-1}] = 0.\end{aligned}$$

In particular, we look at the evolution equation of the form

$$\omega_{it} = \rho\omega_{it-1} + \xi_{it}. \tag{25}$$

As a result, we have four moment conditions in the second stage that we can use to estimate the coefficients β_0, β_l and β_k in the original production function, as well as the evolution coefficient ρ of equation (25):

$$\mathbb{E}\left[\left(y_{it}-\beta_0-\beta_k k_{it}-\beta_l l_{it}-\rho\left(\widetilde{\Phi}_{t-1}(k_{it-1},l_{it-1},m_{it-1})-\beta_0-\beta_k k_{it-1}-\beta_l l_{it-1}\right)\right)\right.$$

$$\left.\otimes\begin{pmatrix}1\\k_{it}\\l_{it-1}\\\widetilde{\Phi}_{t-1}(k_{it-1},l_{it-1},m_{it-1})\end{pmatrix}\right]=0. \quad (26)$$

As we are looking to find the total factor productivity value, which is the sum of β_0 and ω_{it} in this case, we focus on the middle two moment conditions. For given values of β_l and β_k, we form a new variable

$$\beta_0+\widehat{\omega_{it}(\beta_l},\beta_k)=\widehat{\widetilde{\Phi}}_t(k_{it},l_{it},m_{it})-\beta_k k_{it}-\beta_l l_{it}. \quad (27)$$

We follow the intuition of the innovation equation (25); however, instead of using ω_{it} as the dependent variable, we regress the equation

$$\beta_0+\widehat{\omega_{it}(\beta_l},\beta_k)=\rho_0+\rho\left(\beta_0+\widehat{\omega_{it-1}(\beta_l},\beta_k)\right)+\xi_{it}. \quad (28)$$

Note that the error term $\widehat{\xi_{it}}(\beta_k,\beta_l)$ from this regression is identical to the one in (25). In addition, it can be implied from the regression that $\widehat{\xi_{it}}(\beta_k,\beta_l)$ is uncorrelated with $\widehat{\omega_{it-1}}(\beta_l,\beta_k)$ and has a mean of zero, meaning that the first and fourth moment conditions in (26) are satisfied. Thus, we search for the values of $\widehat{\beta_k}$ and $\widehat{\beta_l}$ that satisfy the following two moment conditions, which are the remaining two conditions from equation (26):

$$E\left[\widehat{\xi_{it}}(\beta_k,\beta_l)\otimes\begin{pmatrix}k_{it}\\l_{it-1}\end{pmatrix}\right]=0. \quad (29)$$

After finding the value of $\widehat{\beta_k}$ and $\widehat{\beta_l}$, we use their values in equation (28) to find ρ as well as the sum of productivity and technology terms from equation (27). Then, the exponential of that sum is the firm-specific total factor productivity $TFP_{it}=\exp(\beta_0+\omega_{it})$.

C Theoretical Model with Heterogeneous Productivity Distributions

In contrast with the theoretical model presented in Section 6.2, here we assume that there is no management cost. The model here is more associated with the empirical analysis in Section 4.2. Suppose that there is a firm whose production function is given by $A(l) \cdot F(k, l)$ where k and l are capital and labor inputs, respectively, and $A(l)$ represents a productivity term, which depends on l. Suppose that A is continuously differentiable. The manager of the firm is risk-averse, and his utility denoted by u is a strictly concave function of the firm's profit level. The profit of this firm is given by

$$\Pi(k, l) = pA(l)F(k, l) - rk - wl, \tag{30}$$

where p denotes the price of the product, and r and w represent the costs of capital and labor, respectively. We measure the disincentive for expanding by

$$D(l) = u(\Pi(k, l)) - \mathbb{E}[u(\Pi(k, l) + \Delta\Pi(K, L))], \tag{31}$$

and call $D(l)$ a *disincentive measure*. In words, a disincentive measure is the difference between the manager's current utility and the expected utility hypothetically assuming that the firm employs one extra unit of labor. If $D(L) \geq 0$, the manager decides to expand the firm; otherwise, it becomes unlikely for the firm to grow in size because the manager would experience a loss in his expected utility if he chooses to expand the business.

Now, suppose that the current production size is $(k, l) = (K, L)$. Let $A(L) = a$, of which the manager knows the value. However, he does not exactly know the value of the productivity term after marginally increasing the production scale. Consider the situation where the manager obtains more resources and considers expanding the business that he manages. We show that when the productivity term's distribution is more spread with the same mean, even if the expected profit does not change, the manager has a disincentive to expand.

Lemma 2. *Suppose that the manager knows that the productivity term after marginally expanding the business follows $\tilde{a}_1 = a + \tilde{\epsilon}_1$, where $\tilde{\epsilon}_1$ is a random noise and its distribution function is H_1 with mean 0. Then, the associated disincentive measure $D_1(L) > 0$.*

Proof of Lemma 2. If the manager marginally expands the business, the change in profit is

$$\Delta\Pi(K,L) = \Pi_K(K,L) + \Pi_L(K,L), \tag{32}$$

where the first derivatives of (30) with respect to k and l are respectively given by

$$\Pi_K(K,L) = paF_K(K,L) - r; \tag{33}$$

$$,\Pi_L(K,L) = \mathbb{E}[p\tilde{\epsilon}_1 F(K,L)] + paF_L(K,L) - w. \tag{34}$$

Because $\mathbb{E}[p\tilde{\epsilon}_1 F(K,L)] = 0$, the first-order condition gives us

$$paF_K(K,L) = r \text{ and } \mathbb{E}[p\tilde{\epsilon}_1 F(K,L)] + paF_L(K,L) = paF_L(K,L) = w. \tag{35}$$

Substituting (35) into (33) and (34) of (32), we obtain

$$\mathbb{E}[\Pi(K,L) + \Delta\Pi(K,L)] = \Pi(K,L) + pF_L(K,L)\int_{H_1} \tilde{\epsilon}_1 \mathrm{d}\tilde{\epsilon}_1. \tag{36}$$

Because the mean of H_1 is 0, $\mathbb{E}[\Pi(K,L) + \Delta\Pi(K,L)] = \Pi(K,L)$, and u is strictly concave, we obtain

$$u(\Pi(K,L)) > \mathbb{E}[u(\Pi(K,L) + \Delta\Pi(K,L))].$$

In other words, $D_1(L) > 0$. □

Now, we extend Lemma 2 to a more general case where we can study the manager's decision under different levels of uncertainties in the productivity term. We model the level of uncertainties by using the sum of sequenced random noises (in Lemma 2, it is denoted by $\tilde{\epsilon}_1$). Our next proposition shows that $D(L)$ increases as the prospect after the business expansion becomes riskier, which in our model is presented by the difference in the sum of sequenced random noises.

Proposition 2. *For a given L, take a sequence of random noises $\{\tilde{\epsilon}_n\}_n$ such that each ϵ_n follows a distribution function H_n with mean 0, and a sequence of productivity terms $\{\tilde{a}_n\}$ such that $\tilde{a}_n = a + \sum_{j=1}^{n} \tilde{\epsilon}_j$. For every n, the associated disincentive measures satisfy $D_{n+1}(L) > D_n(L) > 0$.*

Proof of Proposition 2. The simple intuition here is that a distribution with $\tilde{a}_n$ is itself a mean-preserving spread of a distribution with $\tilde{a}_{n-1}$. To show the result recursively, we will prove that when $\tilde{a}_2 = a + \tilde{\epsilon}_1 + \tilde{\epsilon}_2$, where $\tilde{\epsilon}_2$ is a random noise and its distribution function is H_2 with mean 0, the associated disincentive measures $D_1(L)$ and $D_2(L)$ for $\tilde{a}_1$ and $\tilde{a}_2$ satisfy $D_2(L) > D_1(L)$. Then recursively, we obtain the desired result.

Let $\Delta\Pi_n(K, L)$ denote the change in profit under the changes in productivity terms up to ϵ_n (namely, $\Delta\Pi_1(K, L) = \Delta\Pi(K, L)$ studied in Lemma 2). Note $\tilde{a}_2 = a + \tilde{\epsilon}_1 + \tilde{\epsilon}_2$ where $\tilde{\epsilon}_1$ and $\tilde{\epsilon}_2$ are distributed according to H_1 and H_2, respectively, both of which have a mean of 0. By (36),

$$\begin{aligned} \mathbb{E}[\Delta\Pi_2(K, L)] &= pF_L(K, L) \left[\int_{H_1} \tilde{\epsilon}_1 \mathrm{d}\tilde{\epsilon}_1 + \int_{H_2} \tilde{\epsilon}_2 \mathrm{d}\tilde{\epsilon}_2\right] \\ &= pF_L(K, L) \int_{H_2} \tilde{\epsilon}_2 \mathrm{d}\tilde{\epsilon}_2 + \mathbb{E}[\Delta\Pi_1(K, L)]. \end{aligned} \tag{37}$$

Then, by (37), the distribution of $\Delta\Pi_2(K, L)$ is a mean-preserving spread of the distribution of $\Delta\Pi_1(K, L)$. Thus, by Theorem 2 from Rothschild and Stiglitz (1970), the following holds:

$$\mathbb{E}[u\left(\Pi(K, L) + \Delta\Pi_1(K, L)\right)] > \mathbb{E}[u(\Pi(K, L) + \Delta\Pi_2(K, L))]. \tag{38}$$

By (38) together with Lemma 2, we obtain the required result. $\square$

D Tables and Figures

Table A-1: Total and Transitional Numbers of Firms, by Year

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
# Firms	3405	4172	4963	5820	6377	6759	7277	7005	6974	7194	7023	6840	6637	6455	4797	5920
# Micro-sized Firms	287	333	377	394	388	313	314	264	237	391	436	467	481	472	186	480
# Small-sized Firms	2099	2621	3138	3750	4217	4552	4865	4677	4700	4742	4605	4413	4247	4066	2773	3653
# Middle-sized Firms	294	327	377	456	472	527	598	628	585	584	605	606	592	552	531	494
# Large-sized Firms	725	891	1071	1220	1300	1367	1500	1436	1452	1477	1377	1354	1317	1365	1307	1293
# Continuing Firms	2254	3191	3935	4737	5501	5875	6375	6650	6490	6541	6565	6450	6252	6058	4569	4376
Increase Size Group	184	304	323	323	352	333	370	301	283	299	307	273	235	265	157	179
Increase from Micro	45	98	106	109	124	90	95	82	72	50	96	80	72	72	22	30
Increase from Small	88	118	128	138	146	151	171	156	124	147	142	123	98	106	85	80
Increase from Middle	51	88	89	76	82	92	104	63	87	102	69	70	65	87	50	69
Decrease Size Group	104	103	137	210	265	230	283	347	337	316	419	342	302	285	251	228
Decrease from Small	39	47	56	58	80	62	89	80	81	107	140	119	121	107	54	56
Decrease from Middle	36	31	34	70	100	76	88	135	148	103	134	108	98	103	95	86
Decrease from Large	31	27	51	85	90	97	107	133	110	110	148	118	85	76	103	87
Unchanged Size Group	29	25	47	82	85	92	106	132	108	106	145	115	83	75	102	86

Table A-2: Summary of Enterprise Census Data Statistics (Number of Observations)

Sector	# Observations	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
15	17747	466	671	815	938	1065	1148	1176	1250	1222	1230	1253	1220	1195	1144	1109	828	1017
17	5521	135	185	231	260	282	327	385	398	380	368	386	404	377	378	378	292	355
18	7781	186	249	301	366	438	476	470	546	531	542	556	548	558	527	531	462	494
20	5858	156	222	276	319	395	403	424	435	416	405	413	406	380	357	345	222	284
21	6357	144	213	259	319	369	409	438	455	433	425	455	461	451	440	429	278	379
22	4469	109	153	196	247	289	314	319	327	287	287	336	304	314	300	286	167	234
24	6783	174	234	270	322	370	393	457	481	466	462	473	472	472	459	456	373	449
25	8585	193	265	333	404	472	538	563	631	578	584	643	625	624	604	589	408	531
26	10353	281	404	471	528	630	673	708	756	745	745	743	712	660	648	619	454	576
27	2428	33	63	85	113	132	150	169	192	187	175	182	183	174	181	158	107	144
28	9746	184	272	363	440	542	599	656	736	715	721	727	710	689	679	658	444	611
29	3126	67	105	128	159	185	207	200	221	216	206	206	234	219	212	207	161	193
31	2258	70	87	105	117	132	142	161	166	158	158	163	148	134	140	131	115	131
35	2097	67	88	108	126	141	153	151	160	149	150	140	122	115	113	107	102	105
36	6909	135	194	231	305	378	445	482	523	522	516	518	474	478	455	452	384	417
# Firms		2400	3405	4172	4963	5820	6377	6759	7277	7005	6974	7194	7023	6840	6637	6455	4797	5920

Table A-3: The Results from the Stochastic Frontier Approach

	a_0	b_1	b_2	b_3	σ_t
	-0.210	0.453	0.144	0.184	0.008
95 % CI	-0.595	0.155	-0.092	-0.008	-0.040
	0.196	0.743	0.378	0.376	0.056
SRF*	1.005	1.012	1.012	1.148	1.001
SIF†	49.696	47.543	38.460	48.679	44.719
Difference of DICs‡: -633770.57					

Sector 15

	a_0	b_1	b_2	b_3	σ_t
	-0.259	0.534	0.090	0.121	0.010
95 % CI	-0.932	0.279	-0.013	-0.035	-0.035
	0.396	0.773	0.192	0.276	0.057
SRF*	1.485	1.029	1.322	1.191	1.056
SIF†	34.405	42.121	39.846	42.768	28.169
Difference of DICs‡: -108398.51					

Sector 18

	a_0	b_1	b_2	b_3	σ_t
	-0.099	0.416	0.117	0.222	0.006
95 % CI	-0.510	0.254	0.069	0.011	-0.034
	0.307	0.570	0.164	0.430	0.046
SRF*	1.011	1.734	1.554	2.878	1.304
SIF†	45.013	35.802	42.984	43.082	37.201
Difference of DICs‡: -195878.34					

Sector 25

	a_0	b_1	b_2	b_3	σ_t
	-0.447	0.548	0.131	0.129	0.008
95 % CI	-1.106	0.281	0.071	-0.052	-0.036
	0.228	0.817	0.191	0.314	0.052
SRF*	1.188	1.057	1.021	1.413	1.244
SIF†	49.118	33.773	46.564	47.551	34.324
Difference of DICs‡: -313324.41					

Sector 26

	a_0	b_1	b_2	b_3	σ_t
	-0.229	0.481	0.145	0.193	0.010
95 % CI	-0.440	0.212	-0.036	-0.008	-0.029
	-0.001	0.740	0.327	0.394	0.050
SRF*	1.841	1.238	1.124	1.001	1.095
SIF†	23.815	32.839	31.593	42.464	22.303
Difference of DICs‡: -146114.61					

Sector 28

	a_0	b_1	b_2	b_3	σ_t
	0.040	0.527	0.142	0.179	0.006
95 % CI	-0.519	0.263	-0.061	-0.060	-0.032
	0.595	0.788	0.348	0.415	0.044
SRF*	1.577	1.126	1.721	1.160	1.064
SIF†	29.268	33.230	34.135	32.104	23.500
Difference of DICs‡: -109719.21					

Sector 36

Note: b_1, b_2 and b_3 are the coefficients corresponding to the labor, capital and intermediate inputs, respectively. * SRF is *scale reduction factor*; † SIF is *simulation inefficiency factor*; ‡ Difference of DICs is the difference of *deviance information criterion* for the random-effect model and the fixed-effect model (DIC for the random-effect model – DIC for the fixed-effect model).

Table A-4: Coefficient Estimates from the Two-stage GMM Model for Two Periods

Sector	2000 – 2008			2009 – 2016		
	# Firms	β_l	β_k	# Firms	β_l	β_k
15	283	0.8247	0.0941	575	0.7643	0.2599
		(0.0521)	(0.0641)		(0.0528)	(0.0599)
18	102	0.8744	0.1791	314	0.8969	0.0958
		(0.0675)	(0.0694)		(0.0482)	(0.0451)
25	97	0.8805	0.1604	274	0.6986	0.2578
		(0.0731)	(0.0891)		(0.0418)	(0.0470)
26	194	0.9800	0.1746	319	0.8575	0.2291
		(0.0370)	(0.0430)		(0.0393)	(0.0428)
28	80	1.0519	0.1177	269	0.5971	0.2913
		(0.1123)	(0.1277)		(0.0481)	(0.0530)
36	56	0.9818	0.0692	245	0.8154	0.1489
		(0.0889)	(0.0959)		(0.0379)	(0.0384)

Note: β_l and β_k are the coefficients for labor and capital inputs, respectively. Standard errors are given in parentheses.

Figure A-2: Distribution of SF Efficiency Scores across Different Firm Sizes

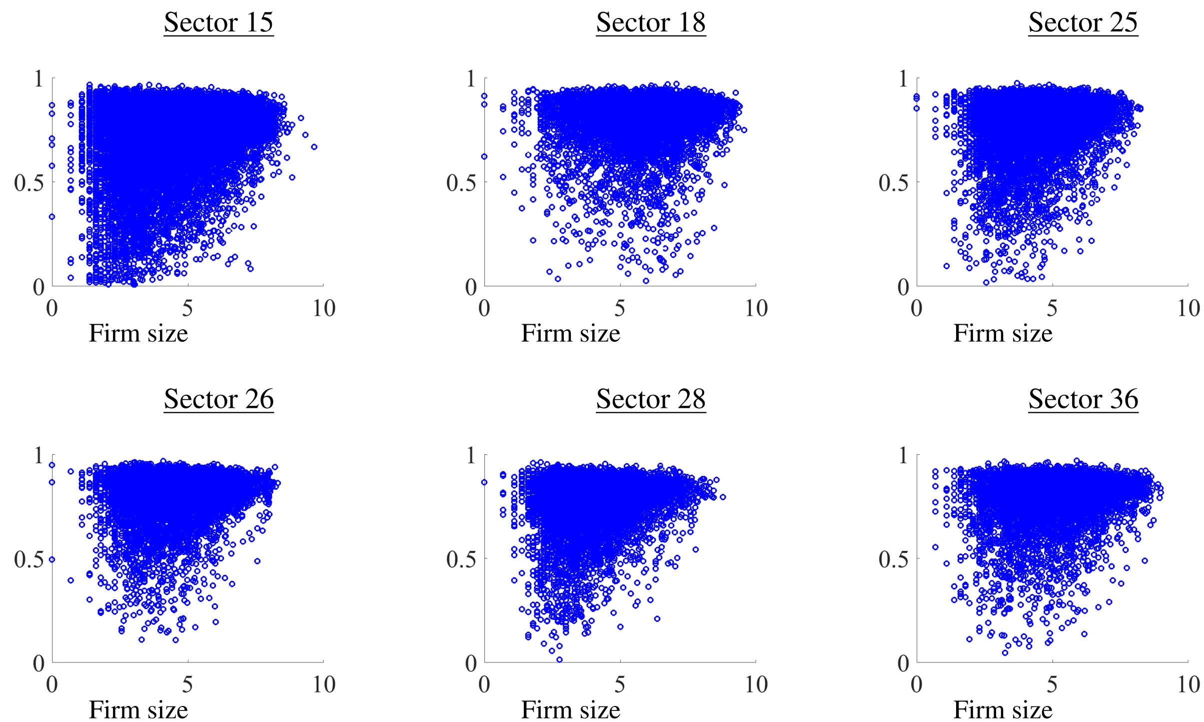


Figure A-3: Distribution of Average Efficiency Scores across Different Firm Sizes under the Fixed-Effect Model

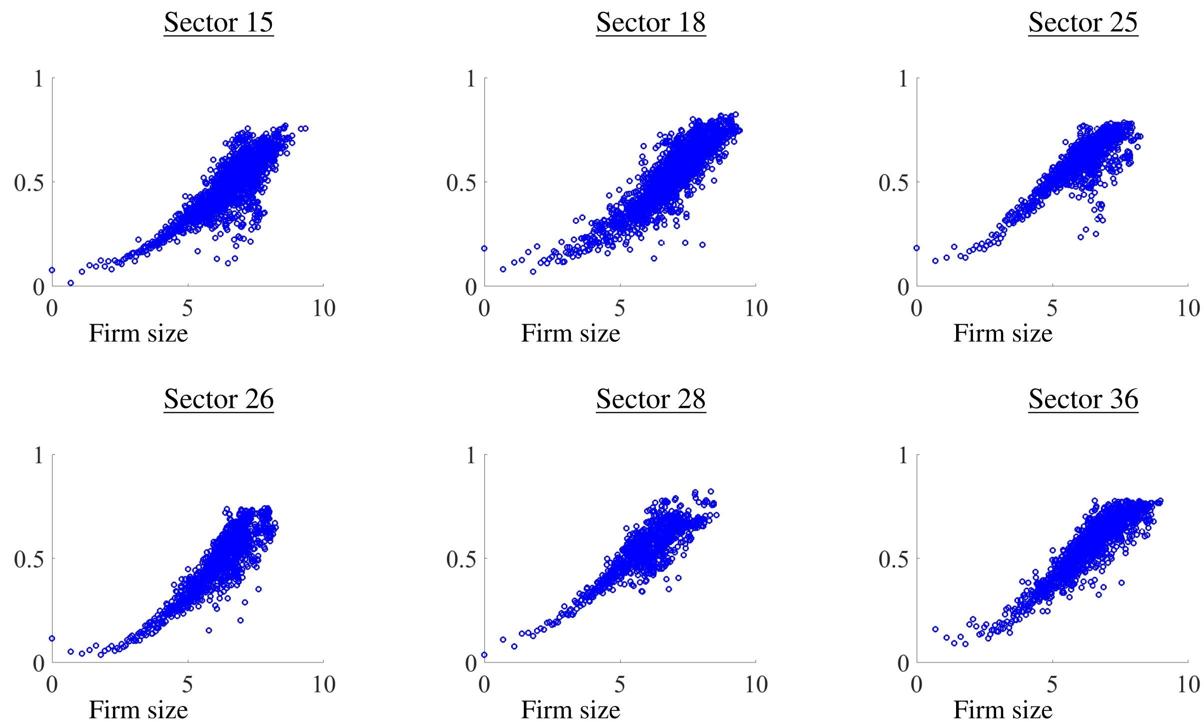


Figure A-4: Distribution of Average Efficiency Scores across Different Firm Sizes under the Random-Effect Model

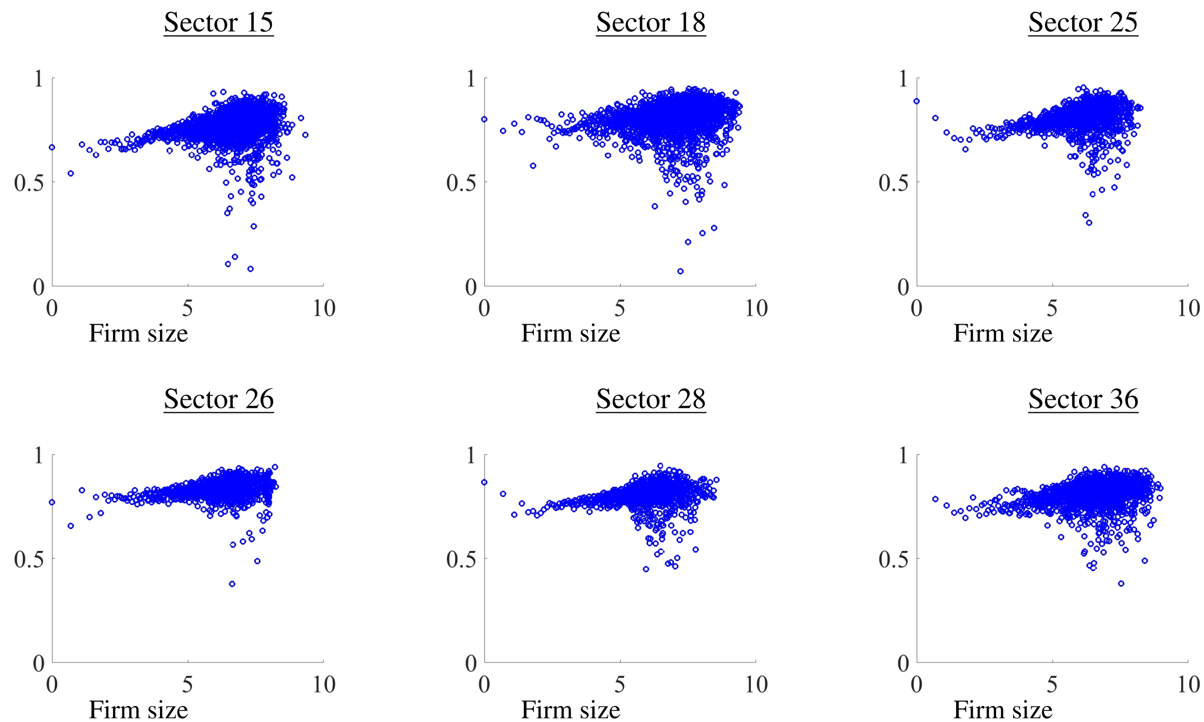


Figure A-5: Distribution of RTS across Different Sizes

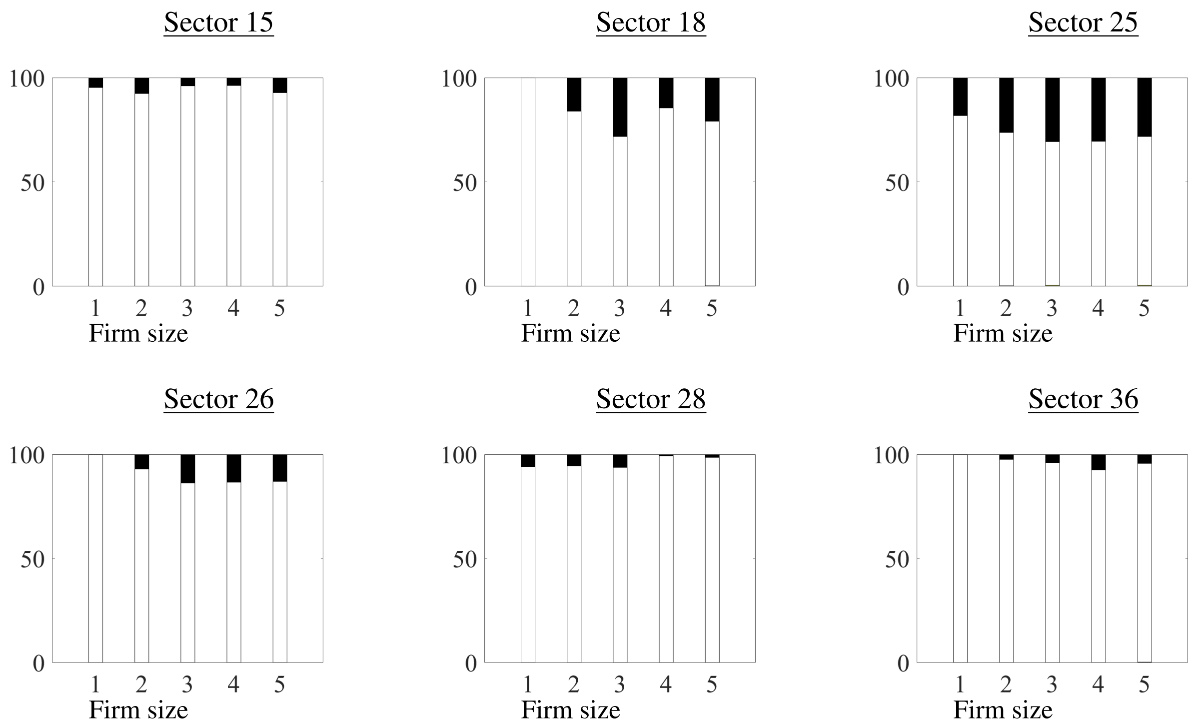
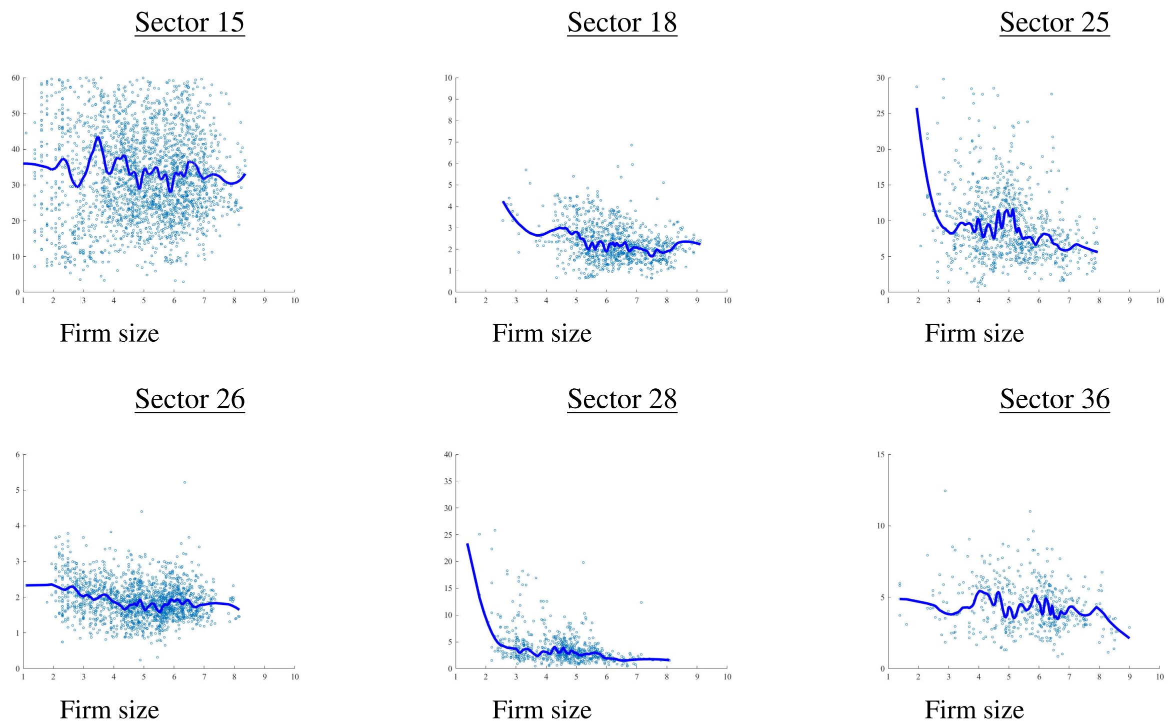


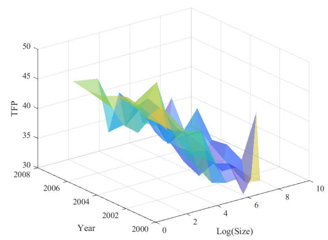
Figure A-6: Distribution of Total Factor Productivity across Different Firm Sizes – From 2000 to 2008



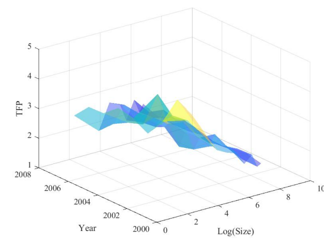
Note: The curve represents the smooth trend line of the efficiency scores for every 0.1 interval of $log(size)$.

Figure A-7: Distribution of Total Factor Productivity across Firm Sizes and Years – From 2000 to 2008

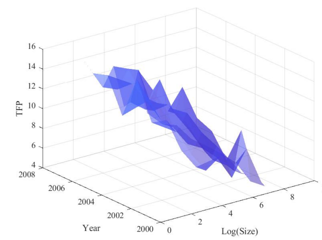
Sector 15



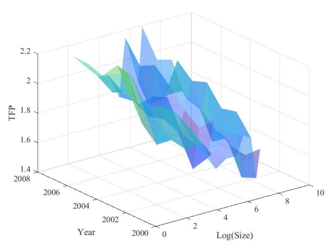
Sector 18



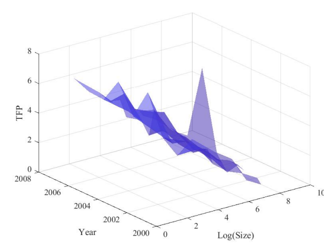
Sector 25



Sector 26



Sector 28



Sector 36

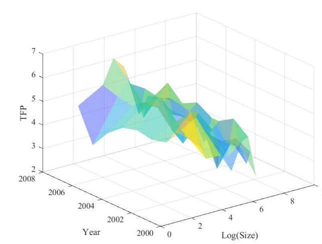
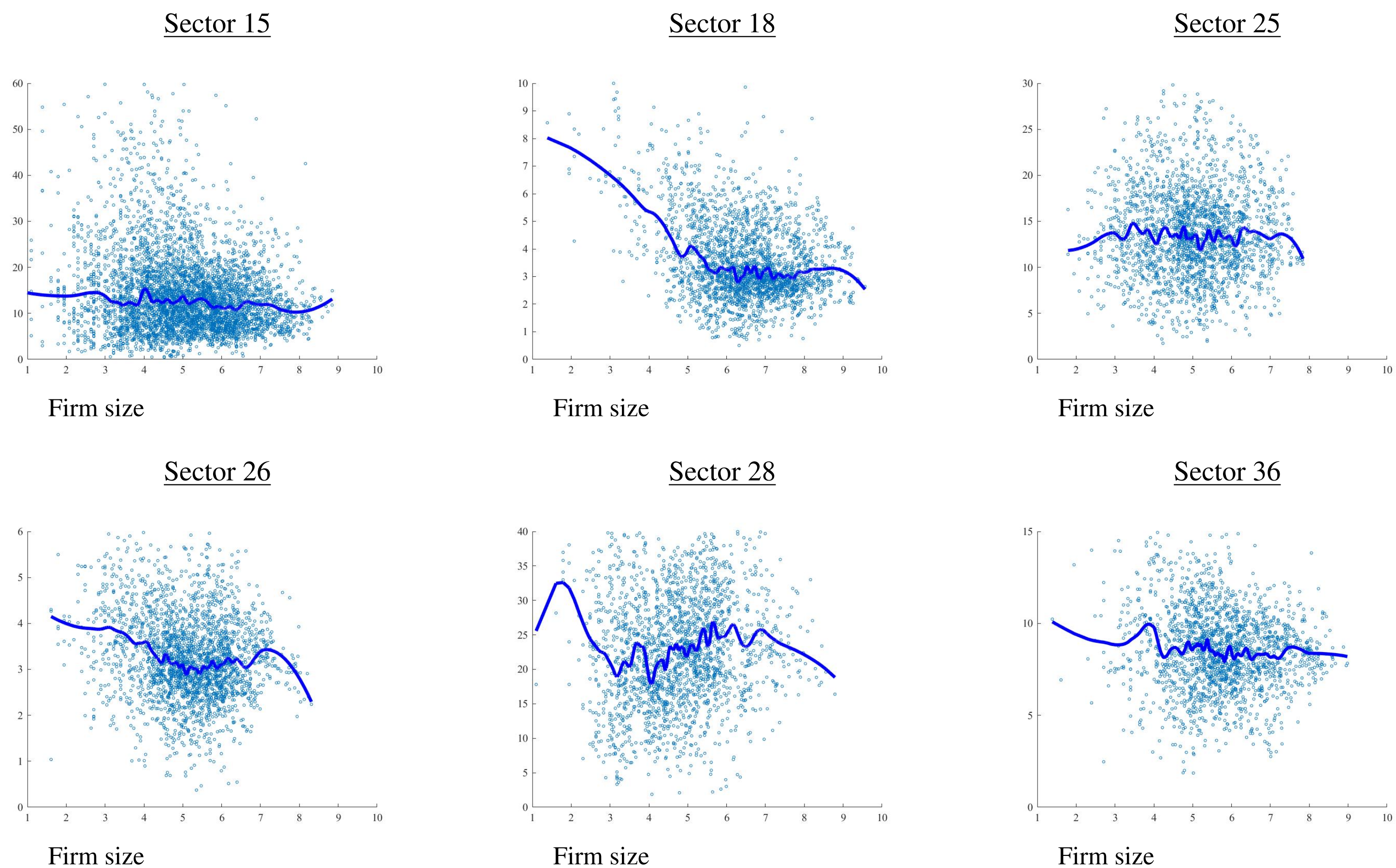
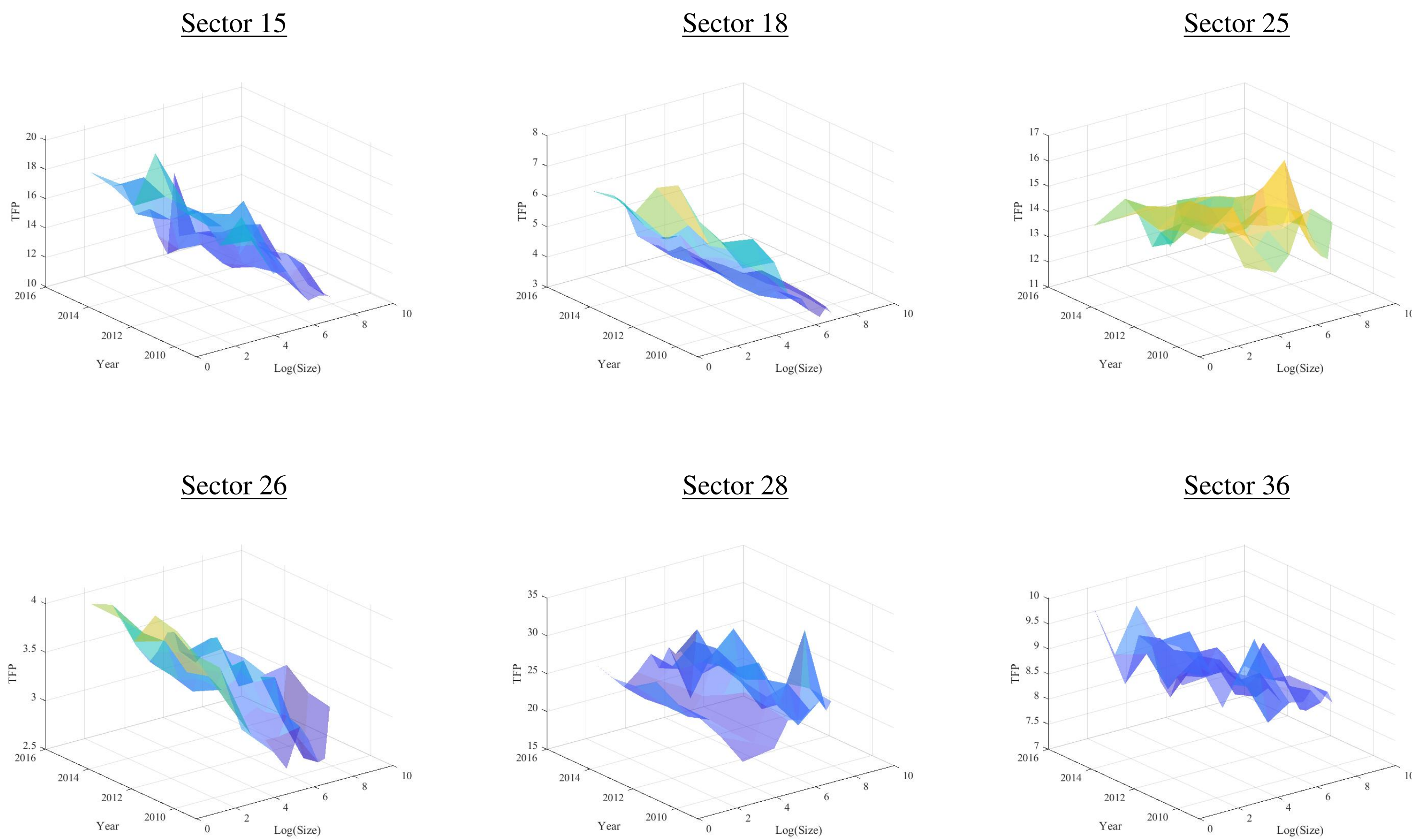


Figure A-8: Distribution of Total Factor Productivity across Different Firm Sizes – From 2009 to 2016



Note: The curve represents the smooth trend line of the efficiency scores for every 0.1 interval of $log(size)$.

Figure A-9: Distribution of Total Factor Productivity across Firm Sizes and Years – From 2009 to 2016



References

Ackerberg, D.A., K. Caves, and G. Frazer, “Identification properties of recent production function estimators,” *Econometrica*, 2015, *83* (6), 2411–2451.

Acs, Z.J. and D.B. Audretsch, “Innovation, marketstructures, and firm size,” *Review of Economics and Statistics*, 1987, *69*, 567–74.

Aigner, D.J., C.A.K. Lovell, and P. Schmidt, “Formulation and estimation of stochastic frontier production functions,” *Journal of Econometrics*, 1977, *6*, 21–37.

Albuquerque, Rui and Hugo A Hopenhayn, “Optimal lending contracts and firm dynamics,” *The Review of Economic Studies*, 2004, *71* (2), 285–315.

Alfaro, Laura, Gonzalo Asis, Anusha Chari, and Ugo Panizza, “Corporate debt, firm size and financial fragility in emerging markets,” *Journal of International Economics*, 2019, *118*, 1–19.

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

Balcombe, Kelvin, Iain Fraser, and Jae H Kim, “Estimating technical efficiency of Australian dairy farms using alternative frontier methodologies,” *Applied Economics*, 2006, *38* (19), 2221–2236.

Bartelsman, E., J. Haltiwanger, and S. Scarpetta, “Cross-country differences in productivity: The role of allocation and selection,” *American Economic Review*, 2013, *103* (1), 305–34.

Bento, Pedro and Diego Restuccia, “Misallocation, Establishment Size, and Productivity,” *American Economic Journal: Macroeconomics*, July 2017, *9* (3), 267–303.

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

Bezemer, Dirk, Kelvin Balcombe, Junior Davis, and Iain Fraser, “Livelihoods and farm efficiency in rural Georgia,” *Applied Economics*, 2005, *37* (15), 1737–1745.

Biesebroeck, Johannes Van, “Firm size matters: Growth and productivity growth in African manufacturing,” *Economic Development and Cultural Change*, 2005, *53* (3), 545–583.

Bloom, Nicholas, Aprajit Mahajan, David McKenzie, and John Roberts, “Why do firms in developing countries have low productivity?,” *American Economic Review*, 2010, *100* (2), 619–23.

Blundell, R. and S. Bond, “GMM estimation with persistent panel data: an application to production functions,” *Econometric Reviews*, 2000, *19* (3), 321–340.

Cabral, Luis and Jose Mata, “On the evolution of the firm size distribution: Facts and theory,” *American economic review*, 2003, *93* (4), 1075–1090.

Caves, D.W., L.R. Christensen, and W. Diewert, "The economic theory of index numbers and the measurement of input, output, and productivity," *Econometrica*, 1982, *50* (6), 1393–1414.

Clementi, Gian Luca and Hugo A Hopenhayn, "A theory of financing constraints and firm dynamics," *The Quarterly Journal of Economics*, 2006, *121* (1), 229–265.

Cooley, Thomas F and Vincenzo Quadrini, "Financial markets and firm dynamics," *American economic review*, 2001, *91* (5), 1286–1310.

Dalton, S., E. Friesenhahn, J. Spletzer, and D. Talan, "Employment growth by size class: firm and establishment data," *Monthly Labor Review*, 2011, *134*, 3–24.

Dhawan, R., "Firm size and productivity differential: theory and evidence from a panel of U.S. firms," *Journal of Economic Behavior & Organization*, 2001, *44* (3), 269–293.

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

Färe, R. and S. Grosskopf, *Cost and Revenue Constrained Production*, New York: Springer Verlag, 1994.

Feng, G. and X. Zhang, "Returns to scale at large banks in the U.S.: A random coefficient stochastic frontier approach," *Journal of Banking & Finance*, 2014, *39* (C), 135–145.

Foster, L., J. Haltiwanger, and C. Syverson, "Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?," *The American Economic Review*, 2008, *98* (1), 394–425.

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

Hsieh, Chang-Tai and Peter J Klenow, "Misallocation and manufacturing TFP in China and India," *The Quarterly journal of economics*, 2009, *124* (4), 1403–1448.

Hsieh, C.T. and B.A. Olken, "The missing "missing middle"," *Journal of Economic Perspectives*, 2014, *28* (3), 89–108.

Jones, Charles I, "Misallocation, Economic Growth, and Input-Output Economics," Working Paper 16742, National Bureau of Economic Research 2011.

Kasahara, H., P. Schrimpf, and M. Suzuki, "Identification and estimation of production function with unobserved heterogeneity," *University of British Columbia mimeo*, 2015.

Katayama, H., S. Lu, and J.R. Tybout, "Firm-level productivity studies: illusions and a solution," *International Journal of Industrial Organization*, 2009, *27* (3), 403–413.

Kato, Atsuyuki and Naomi Kodama, "Markups, productivity and external market development of the service SMEs," *Applied Economics*, 2014, *46* (29), 3601–3608.

Kleit, Andrew N and Dek Terrell, "Measuring potential efficiency gains from deregulation of electricity generation: a Bayesian approach," *Review of Economics and Statistics*, 2001, *83* (3), 523–530.

Lawrance, Anthony B and Robert E Marks, "Firm size distributions in an industry with constrained resources," *Applied Economics*, 2008, *40* (12), 1595–1607.

Leung, D., C. Meh, and Y. Terajima, "Productivity in Canada: does firm size matter?," *Bank of Canada Review*, 2008, *1*, 5–14.

Levinsohn, J. and A. Petrin, "Estimating production functions using inputs to control for unobservables," *The Review of Economic Studies*, 2003, *70* (2), 317–341.

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

Little, Ian MD, "Small manufacturing enterprises in developing countries," *The World Bank Economic Review*, 1987, *1* (2), 203–235.

Liu, P.W. and X. Yang, "The theory of irrelevance of the size of the firm," *Journal of Economic Behavior & Organization*, 2000, *42* (2), 145–165.

Marschak, J. and W.H. Andrews, "Random simultaneous equations and the theory of production," *Econometrica*, 1944, pp. 143–205.

Martins, J.O. and S. Scarpetta, "Estimation of the cyclical behaviour of mark-ups," *OECD Economic studies*, 2002, *2002* (1), 173–188.

Meeusen, W. and J. Van den Broeck, "Effciency estimation from Cobb-Douglas production functions with composed error," *International Economic Review*, 1977, *18*, 435–444.

O'Donnell, C.J. and T.J. Coelli, "A Bayesian approach to imposing curvature on distance functions," *Journal of Econometrics*, 2005, *126*, 493–523.

Olley, S.G. and A. Pakes, "The dynamics of productivity in the telecommunications equipment industry," *Econometrica*, 1996, *64* (6), 1263–1297.

Rothschild, M. and J.E. Stiglitz, "Increasing risk I: A definition," *Journal of Economic Theory*, 1970, *2* (3), 225–243.

Saastamoinen, Antti and Timo Kuosmanen, "Is corruption grease, grit or a gamble? Corruption increases variance of productivity across countries," *Applied Economics*, 2014, *46* (23), 2833–2849.

Tsionas, E.G., "Stochastic frontier models with random coefficients," *Journal of Applied Econometrics*, 2002, *17*, 127–147.

Tybout, J.R., "Manufacturing firms in developing countries: How well do they and why?," *Journal of Economic Literature*, 2000, *38*, 11–44.

Van den Broeck, Julien, Gary Koop, Jacek Osiewalski, and Mark FJ Steel, "Stochastic frontier models: A Bayesian perspective," *Journal of Econometrics*, 1994, *61* (2), 273–303.

Wooldridge, J.M., "On estimating firm-level production functions using proxy variables to control for unobservables," *Economics Letters*, 2009, *104* (3), 112–114.

Yuan, Qigang, Yanping Zhao, Hui Shang, Wei Zhang, and Zaghum Umar, "Financing constraints on the size distribution of industrial firms: the Chinese experience," *Applied Economics*, 2016, *48* (41), 3899–3911.