

# Revisiting the “Missing Middle”: Productivity Analysis\*

Hien Thu Pham<sup>†</sup>

Shino Takayama<sup>‡</sup>

April 17, 2017

## Abstract

This paper investigates empirically the relationship between firm size and production efficiency and inefficiency associated with the production scale. We study the possible sources of the missing middle phenomenon, which refers to the fact that most employment in developing countries is located in either small-sized or large-sized firms. Using Vietnamese data, we show that middle-sized firms’ production efficiencies tend to be lower than small-sized or large-sized firms in most of the manufacturing industries, that the least efficient firm tends to be middle-sized, and that efficiency scores are more diverse for middle-sized firms, which is arguably associated with the uncertainty that a small firm faces when increasing its size. Our work also indicates that the large-sized firms may be unable to fully utilize their inputs.

**Key Words:** firm size distribution, missing middle, productivity, efficiency, data envelopment analysis, free disposal hull.

**JEL Classification Numbers:** D21, D22, L25.

---

\*A part of this paper was circulated as *Revisiting the “Missing Middle”: Production and Corruption*. We would like to thank Fabrizio Carmignani, Begona Dominguez, Jeff Kline, Do Won Kwak, Hideyuki Mizobuchi, Andrew McLennan, Flavio Menezes, Thanh Le, Antonio Peyrache, Rodney Strachan, Satoshi Tanaka, Yuichiro Waki, Haishan Yuan, Valentin Zelenyuk, and other participants at the UQ Macro Workshop. We are grateful to the General Statistical Office of Vietnam for providing permission to use the micro data of the Enterprise surveys. ST acknowledges funding from the Zengin Foundation for Studies on Economics and Finance. Any remaining errors are our own.

<sup>†</sup>Pham (corresponding author); School of Economics, University of Queensland, Level 6 - Colin Clark Building (39), QLD 4072, Australia; e-mail: [thu.pham3@uq.net.au](mailto:thu.pham3@uq.net.au)

<sup>‡</sup>Takayama; School of Economics, University of Queensland; email: [s.takayama1@uq.edu.au](mailto:s.takayama1@uq.edu.au); tel: +61-7-3346-7379; fax: +61- 7-3365-7299.

# 1 Introduction

The missing middle refers to the empirical fact that most employment in developing countries is located in either small-sized or large-sized firms. The missing middle was first documented in Liedholm and Mead (1987). Tybout (2000) also finds that there is a large spike in the size distribution for the small-sized category, and that it drops off quickly in the middle-sized category in the poorest countries, and then argues that strong business regulation could be a reason behind the existence of too many small firms<sup>1</sup>.

Our objective is to identify the reasons for the missing middle phenomenon, which is more evident in developing countries than in developed countries. To achieve this objective, we use firm-level data from Vietnam and examine whether there are any differences in production efficiencies across different sized firms. We estimate each firm's production efficiency and further study average product and production scale in each manufacturing sector.

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, and the scale at which minimum efficiency occurs and the level of this efficiency are different across industries. Furthermore, we show that the most inefficient firms tend to be middle-sized across different sectors. Interestingly, the middle-sized firms are quite diverse in terms of efficiency scores. Some middle-sized firms show high efficiency scores, while other firms show very low scores. This may be associated with the uncertainty about increasing firm size.

Business environments vary across nations. Although it is difficult to measure how good the business environment is, or how well each firm is managed, the level of production efficiency can reflect the management of an organization. We show that there is significant uncertainty for a small firm in increasing its size, as the middle-sized groups are a 'mixed bag'. We also provide empirical evidence that available resources might be less efficiently used in developing nations and that this could be an important factor for the missing middle phenomenon.

The data set that we use to measure productivity and efficiency is the Enterprise Census conducted by the General Statistics Office of Vietnam. The Enterprise Census contains information on all reg-

---

<sup>1</sup>See also Little et al. (1987) for South Korea and Steel and Webster (1992) for Ghana. The presence of the missing middle phenomenon and how it is defined have been extensively discussed in the literature (see Hsieh and Olken, 2014; Tybout, 2014a,b).

istered firms in Vietnam. Our data include capital, labor, and intermediate materials as inputs, and goods measured in monetary terms as output. Table 1 lists the sectors we study in this paper.<sup>2</sup>

To see if the missing middle phenomenon is observed in our data set and if there are any differences in the firm size distributions across sectors, we compute the bimodality coefficients for each sector.<sup>3</sup> We then consider the situation where each firm uses capital, labor, and intermediate materials to produce goods, and study production efficiency in various sectors of the manufacturing industry by using the nonparametric frontier approach.

Table 1: ISIC Codes and Industry Description

| ISIC | Industry Description          | ISIC | Industry Description        |
|------|-------------------------------|------|-----------------------------|
| 14   | Other mining and quarrying    | 15   | Food products and beverages |
| 17   | Textiles                      | 18   | Wearing apparel etc.        |
| 19   | Tanning and dressing leather  | 20   | Wood and wood products      |
| 21   | Paper and paper products      | 22   | Publishing, printing, etc.  |
| 24   | Chemical products, etc.       | 25   | Rubber and plastic products |
| 26   | Nonmetallic mineral products  | 28   | Fabricated metal products   |
| 29   | Machinery and equipment, etc. | 36   | Furniture                   |

In recent years, many methods have been developed to estimate production efficiency under the frontier approach.<sup>4</sup> Comparing average products across different inputs is a complicated task. As pointed out by Ray (2004), measures of a firm’s productivity relying on a single input disregarding other inputs may fail to reflect total factor productivity. Taking this into account, we use the frontier approach.

Our analysis relies on the following theoretical framework. Consider one sector in the manufacturing industry. Firm  $i$ ’s input and output combination is called a *productive unit*, denoted by  $(x_i, y_i)$  with  $x_i \in \mathbb{R}_+^3$  and  $y_i \in \mathbb{R}_+$ . Then, define firm  $i$ ’s production set by  $\Psi_i = \{(x, y) : x \geq x_i, y \leq y_i\}$ . The union of all the firms’ production sets in the sector,  $\bigcup_i \Psi_i^l$ , is the production set of this sector. The

<sup>2</sup>These sector codes are based on the International Standard Industrial Classification (ISIC) codes. We choose these sectors because each sector based on the two-digit-level ISIC code has more than 200 observations per year, so that we can ensure robustness of the nonparametric estimations.

<sup>3</sup>The logic behind the bimodality coefficient is that a bimodal distribution would have very low kurtosis, an asymmetric characteristic, or both, all of which increase this coefficient. The formula itself does not assume a particular distribution. However, the value for the uniform distribution or the exponential distribution is  $5/9$ . Values greater than  $5/9$  may indicate a bimodal or multimodal distribution in the data.

<sup>4</sup>The production frontier can be estimated using parametric estimates (deterministic or stochastic frontier analysis) or nonparametric ones (data envelopment analysis or free disposal hull). For more details, see Kalirajan and Shand (1994).

surface of this set is the production function. Using these data, we construct a production function for each sector.<sup>5</sup> Then, we compute each firm's efficiency score, which is the minimal proportional reduction of all inputs while maintaining the same output level within  $\bigcup_i \Psi_i^l$ . This is known as the input-oriented method of the free disposal hull (FDH hereafter) approach in the productivity analysis literature.

Second, by using the algorithm developed by Soleimani-Damaneh and Reshadi (2007), we study the relationship between production scale and average product at the firm level in each industry.<sup>6</sup> We find that the large-sized firms tend to operate at larger scales compared with the scale of the most efficient unit.

Our work indicates that the large-sized firms may be unable to fully utilize their inputs. Hsieh and Olken (2014) claim that “the problem of economic development in low-income and middle-income countries is how to relieve the differential constraints faced by large firms.” Our analysis of production scale finds that the large-sized category is quite mixed, including a few firms with high average product and many firms with low average product. Our analysis suggests that one of the problems, in Vietnam at least, is how to activate certain resources that these large-sized firms hold but tend not to fully utilize in production.<sup>7</sup>

Half a century ago, Leibenstein (1966) introduced the concept of *X-efficiency*, which is mainly caused by incentive misalignments in a workplace, unlike allocative efficiency. Furthermore, he provides a review of the empirical literature, showing that allocative inefficiency such as that caused by a tax could cause very small losses in total production, while improving management or motivational

---

<sup>5</sup>To clarify, we do not impose any functional form on the estimation, and a production function is estimated as a frontier of a production set.

<sup>6</sup>Diewert et al. (2011) use the following definition of *returns to scale*. If all inputs are multiplied by a positive scalar,  $t$ , and the consequent output can be represented as  $t^\gamma y$ , the value of  $\gamma$  may be said to indicate the magnitude of returns to scale. If  $\gamma = 1$ , there are constant returns to scale; if  $\gamma > 1$ , there are increasing returns to scale and if  $\gamma < 1$ , there are decreasing returns to scale. In the literature on productivity analysis, a production function is constructed from the data, and we apply this definition to each firm's production by comparing it to the production of the firm with the highest average product among the set of firms with similar production scales. Applying this definition to each firm's production might be confusing, because in the classical production theory of economics, returns to scale is a property of the production technology. Thus, we only use the term returns to scale when we talk about production technology itself.

<sup>7</sup>For example, in Vietnam, sector 17 (Textiles) and sector 21 (Paper and paper products) include many large-sized firms that were previously or are currently state-owned. As we will see later in Figure 4, these industries exhibit this feature more visibly.

efficiency in a workplace could increase production by approximately 50%. We also find that there are many firms whose efficiency is low—less than half of the most efficient firms—across different firm sizes in all sectors. To examine the source of this inefficiency, we study average product and production scale. As Feng and Zhang (2014) note, not considering the technological heterogeneity of an individual firm even within a single sector may cause bias in measuring returns to scale. Pham and Takayama (2017) measure returns to scale considering technological heterogeneity for each firm using a random coefficient stochastic frontier analysis and find that most firms show constant returns to scale technologies. Given their analyses, the finding of this paper that large-sized firms tend to operate at larger scales than the most efficient unit may be arguably caused by inefficiencies only associated with scale. Our findings in this paper complement the work of Pham and Takayama (2017). To our knowledge, our study is one of the few that examines production efficiencies and production scale across different firm sizes in various sectors of a developing country in the context of the missing middle phenomenon.

Finally, we also study the standard deviation of the efficiency scores for each size. Our result shows that the size increases, the standard deviation increases across sectors. This relationship is more eminent for middle-sized firms compared with small-sized firms. Our analysis so far indicates that the middle-sized group is very mixed in the sense that firms with very high and very low production efficiencies coexist. Although our analysis is a snapshot of the Vietnamese manufacturing industry, the diverse production efficiencies in the middle-sized groups can be thought of as a risk that small-sized firms would face in expanding their businesses.

As noted, we adopt the FDH approach to compute the efficiency scores. The FDH approach allows us to deal with a nonconvex production set unlike data envelopment analysis (DEA hereafter). These are well-established nonparametric methods for efficiency analysis and is widely used in the literature (for example Hanson, 2016). However, analyses under both FDH and DEA tend to be sensitive to outliers in the data set. Moreover, the number of observations required to obtain robust estimation results increases dramatically with the number of inputs and outputs under the FDH and DEA approaches. Cazals et al. (2001) develop an *order-m* method to overcome this difficulty (see also Simar, 2003). To check the robustness of our findings, we use the *order-m* approach. Under this approach, for each observation, we randomly draw  $m$  samples and compute efficiency scores for many different time periods. In this way, we see how extreme values of variables distort the results. Our results show that our findings are still robust under this approach.

The remainder of the paper is organized as follows. Section 2 describes our data set and provides a summary of Vietnam’s firm size distribution and average products. Section 3 describes our methodology for measuring productivity and efficiency, and then presents an analysis of production efficiency and inefficiency associated with scale. Section 4 concludes.

## **2 Firm Size Distribution in Vietnam**

### **2.1 Data Description**

In Vietnam, the Enterprise Census has been conducted annually since 2000. In the survey, an enterprise is defined as “an economic unit that independently keeps a business account and acquires its own legal status.” In this paper, we focus on the nine-year period from 2000 to 2008. We exclude recent years (from 2009 to the present) because the Vietnamese economy has suffered from the effects of the global financial crisis, and a 30% reduction in the corporate income tax rate for qualifying entities was implemented in the fourth quarter of 2008 and all of 2009<sup>8</sup>. In addition, small-sized and middle-sized firms involved in labor-intensive production and processing activities will also benefit from the subsequent tax reduction under Decree 60/2011. We exclude inconsistent data from our sample, such as observations that are recorded twice for the same firm in the same year, those with negative or zero revenue values. We include 14 sectors of the manufacturing industry that have at least 200 observations for each year, and the number of observations ranges from approximately 3,300 to 29,000 per industry. Tables 1 and 2 (in the appendix) provide summary statistics of our data. For the whole manufacturing sector, the number of observations ranges from 7,777 (year 2000) to 23,832 (year 2008).

In the data, the size of a firm is a head-count of the number of workers in the firm at the end of the fiscal year. Labor is measured by the total income of employees in a firm. This includes total wages and other employee-related costs such as social security, insurance, and other benefits. Intermediate materials include costs such as fuel and the value of other materials. Capital is measured as assets to be used in production (in terms of present value). We use the deflated revenue as a proxy for physical output.<sup>9</sup> All input and output values are adjusted to account for inflation to obtain real values. Inflation rates are obtained from the World Bank.

---

<sup>8</sup>See Circular No. 03/2009/TT-BTC published in January 2009 by Vietnam’s Ministry of Finance.

<sup>9</sup>At the firm level, prices and quantities may not be accurately measured, and therefore revenue, instead of gross output or cost, is normally used. It has been argued that the elasticities of labor and capital, in a revenue estimate,

## 2.2 Summary

To examine the firm size distribution of the manufacturing sector in Vietnam, we start with the figures for each sector. Here we present two sets of analyses for all 14 sectors. The detailed discussion for these two sets of analyses is also found in Pham and Takayama (2017) which includes the results for 5 out of the 14 sectors. In what follows, we summarize the methods.

First, we conduct the bimodality coefficient test, which examines the relationship between the bimodality, skewness, and kurtosis of the distribution. The bimodality coefficient is defined as  $\frac{m_0^2 + 1}{m_1 + \frac{3(n-1)^2}{(n-2)(n-3)}}$  where  $m_0$  is skewness and  $m_1$  is the excess kurtosis. The bimodality coefficient (BC) statistics range from 0 to 1. The maximum value of 1 is obtained if the population has only two distinct values. A BC that exceeds 0.55 (the value of a uniform population) suggests bimodality. As shown, all BC statistics are larger than 0.55, which indicates the existence of bimodality.<sup>10</sup> The BCs for each sector are presented in Table 3.

Second, we conduct the Dip test for each sector as well as for each year of the entire sample. Hartigan and Hartigan (1985) developed the dip test to distinguish between unimodality and bimodality. By using the Dip test, we can study the null hypothesis  $H_0$  that the sample distribution has a unimodal density, against the alternative hypothesis  $H_1$  that the sample distribution has more than one mode. Hartigan and Hartigan (1985) define the Dip statistics as follows. Let  $\rho(F, G) = \sup_x |F(x) - G(x)|$  for any bounded functions  $F$  and  $G$ . Define  $\rho(F, \mathcal{A}) = \inf_{G \in \mathcal{A}} \rho(F, G)$  for any class  $\mathcal{A}$  of bounded functions. Let  $\mathcal{U}$  be the class of unimodal distribution functions. The Dip statistic of a distribution is defined by  $D(F) = \rho(F, \mathcal{U})$ . According to the results of the Dip test, the null hypothesis is rejected at the 5% significance level for all years, which indicates the presence of bimodality or multimodality.<sup>11</sup> The Dip statistics we obtained are all smaller than 0.05, which indicates the presence of bimodality or multimodality. Unimodality is rejected for all sectors, according to the Dip test.

---

may be downward biased (Basu and Fernald, 1997). Klette and Griliches (1996) report that changes in sector prices are substantially diversified and correlated with changes in labor and capital. However, according to Mairesse and Jaumandreu (2005), the introduction of individual output prices into the production function does not markedly affect the estimate. In addition, the estimation of a production function in terms of “physical quantities” is, in fact, meaningless, unless we confine the analysis to a very precisely defined industry where goods are so homogeneous that firm outputs can be accurately measured and compared across firms. Accordingly, we use this measure in the analysis.

<sup>10</sup>Detailed discussions are also found in Pham and Takayama (2017), or on page 1258 in SAS Institute Inc. (2008).

<sup>11</sup>A more detailed interpretation of this statistic is found in Pham and Takayama (2017); Freeman and Dale (2013).

Table 3: The Bimodality Coefficients in Each Sector

| Sector | 14   | 15   | 17   | 18   | 19   | 20   | 21   | 22   | 24   | 25   | 26   | 28   | 29   | 36   |
|--------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| BC     | 0.57 | 0.68 | 0.74 | 0.64 | 0.68 | 0.68 | 0.62 | 0.61 | 0.72 | 0.71 | 0.36 | 0.66 | 0.58 | 0.69 |

The BCs in sectors 14, 26, and 29 indicate that these sectors' firm size distributions are not bimodal, unlike other sectors. Table 3 indicates that the features of the firm size distributions vary substantially across sectors.<sup>12</sup> In what follows, we study productivity and efficiency across different firm sizes in each sector. The results presented in the next section indicate that the sectors with lower BCs in Table 3, particularly sectors 14, 26, and 29, show different features in terms of productivity and efficiency from other sectors.

### 3 Efficiency Scores and Production Scale

This section consists of two parts. We first measure the efficiency of production, using the efficiency score;<sup>13</sup> in the second part, we study the inefficiency associated with scale at the firm level using nonparametric methods, which impose no or very limited assumptions on the data. The following summarizes our findings.

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

The next subsections detail our methodologies and the above-mentioned points. We consider the situation where a firm uses capital, labor, and intermediate materials to produce goods, which are

<sup>12</sup>We also checked the distributions against Zipf's law, and the hypothesis that they follow Zipf's law was rejected in our data set. The details of the analysis are available upon request.

<sup>13</sup>Various programs to compute the efficiency scores are available. Wilson (2008) provides *FEAR* and Simar (2003) provides Matlab programs.

measured in monetary terms. Real revenue is used as a proxy for output. Three inputs are included in our estimation: intermediate inputs, labor, and capital. Then, we construct a production function for each sector without imposing any restrictions on the parametric relationship between inputs and outputs.

### 3.1 Efficiency Score

Suppose that there are  $L$  industrial sectors in the economy. Take a sector  $l \in \{1, \dots, L\}$  (we repeat the same procedure for each sector) where inputs  $x \in \mathbb{R}_+^3$  are used to produce an output  $y \in \mathbb{R}_+$ . Let  $N^l$  denote the number of firms in sector  $l$  of the data set. Firm  $i$ 's input and output combination is called a *productive unit*, denoted by  $(x_i, y_i)$  with  $x_i \in \mathbb{R}_+^3$  and  $y_i \in \mathbb{R}_+$ .<sup>14</sup> In the same way as discussed in the Introduction, for firm  $i$ , we formally define

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

Furthermore, we define  $\Psi_{FDH}^l = \bigcup_i \Psi_i^l$ , and the convex hull of  $\Psi_{FDH}^l$  is defined as  $\Psi_{DEA}^l$ . The efficiency score for a productive unit  $(x_i, y_i)$  is defined by

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

for each  $\Psi^l = \Psi_{FDH}^l$  under *FDH* or  $\Psi_{DEA}^l$  under *DEA*, respectively. Furthermore, the PS  $\Psi^l$  is assumed to satisfy the regularity conditions; namely, boundedness, closedness, no free-lunch<sup>15</sup>, and free disposability<sup>16</sup>. We say that  $\Psi^l$  exhibits *convexity* if for every  $(x_1, y_1), (x_2, y_2) \in \Psi^l$ , and any  $\alpha \in [0, 1]$ ,  $\alpha(x_1, y_1) + (1 - \alpha)(x_2, y_2) \in \Psi^l$  holds.

The efficiency score<sup>17</sup> lies between 0 and 1, and represents the minimal proportional reduction of all inputs while maintaining the same output level within the production set (hereafter, PS). Solving Problem (2) requires constructing the PS using linear programming.

<sup>14</sup>To keep the notation simple, we do not explicitly denote a period and an industry for each productive unit.

<sup>15</sup>A positive amount of production cannot occur without a positive amount of inputs.

<sup>16</sup>The increase in inputs must lead to increased or constant outputs, and a smaller output vector than a feasible vector is also feasible.

<sup>17</sup>This is called the “input-oriented” efficiency score. There is another definition called the “output-oriented” efficiency score. In this analysis, we use the input-oriented efficiency score because we have computed output-oriented scores for some sectors, and these did not change substantially.

Figure 1 is a scatterplot of efficiency scores for the FDH method in our sample industries for the entire nine-year period, and the curve represents the smooth trend line of the efficiency scores for every 0.1 interval of  $\log(\text{size})$ .<sup>18</sup> The  $x$ -axis represents the logarithm of firm size.

According to Figure 1, except in sectors 14 (Other mining and quarrying), 26 (Nonmetallic mineral products), and 29 (Machinery and equipment, etc.), the relationship between firm size and efficiency score is clearly U-shaped, which indicates that smaller firms and larger firms produce efficiently compared with middle-sized firms. On the other hand, Figure 1 for sectors 14, 26, and 29, for which low BCs are indicated in Table 3, shows that the smooth trend lines of the efficiency scores are mostly decreasing as the size increases. In sector 29, there is a spike in the trend line for very large-sized firms, although this seems to be caused by a few efficient large-sized firms. Moreover, firm size at the bottom of the U-shape—that is, the lowest efficiency score—differs across sectors. The heterogeneity of firm size at the bottom of the U-shape across industries indicates the importance of sector-level analysis in investigating the missing middle.

Figure 2 shows the ratio between the DEA scores and the FDH scores and indicates the existence of some nonconvexities in production technologies.<sup>19</sup> The ratios between the DEA scores and the FDH scores are less than one for most firms, which indicates that there is nonconvexity in the production technologies, and the shapes of the trend lines are quite different across sectors. This indicates that the nonconvexity of the technologies is also significantly different across sectors. This result supports our choice of using FDH rather than DEA.

Figure 3 shows the standard deviation of the efficiency scores for each size. As the size increases, the standard deviation increases across sectors. This relationship is stronger for middle-sized firms compared with small-sized firms. We can interpret this result as the existence of uncertainty that a small firm faces about increasing its size.

## 3.2 Production Scale

In this section, we study the relationship between production scale and average product at the firm level in each industry. Similar to the previous section, we consider manufacturing sector  $l$ . Let  $D^l$

<sup>18</sup> We use Matlab's command `smooth`. This inserts a smooth trend line by using the local regression method.

<sup>19</sup> Farrell (1957) emphasizes indivisibilities and economies of scale as important sources of nonconvexity and emphasizes that convexity can only be justified in terms of time divisibility (ignoring not only any setup times but also indivisibilities, increasing returns to scale, positive or negative production externalities, etc., that can lead to nonconvexity). For some recent theoretical and empirical studies, see Brown (1991) or (Ramey, 1991).

denote a set of all the observations in sector  $l$ . Then, we consider the following problem for each productive unit  $(x_i, y_i) \in D^l$ :

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

We use the FDH algorithm developed by Soleimani-Damaneh and Reshadi (2007).<sup>20</sup> The algorithm is as follows:

**Step 1.** Compute  $\lambda^{ij}$  for all  $i, j \in N^l$  by  $\lambda^{ij} = \frac{y_i}{y_j}$  where  $y_i, y_j \in D^l$ .

**Step 2.** Compute  $\theta^{ji}$  for all  $i, j \in N^l$  by  $\theta^{ji} = \max_{s \in S} \left\{ \frac{x_{js} \cdot \lambda^{ij}}{x_{is}} : (x_i, y_i) \in D^l \right\}$ .

**Step 3.** Compute  $\theta_i$  and  $A_i$  for all  $i \in N^l$  by  $\theta_i = \min_{j \in N^l} \theta^{ji}$  and  $A_i = \{j \in N^l : \theta^{ji} = \theta_i\}$ .

**Step 4.** Compute  $\lambda_i^+$  and  $\lambda_i^-$  for all  $i \in N^l$  by  $\lambda_i^+ = \max_{j \in A_i} \lambda^{ij}$  and  $\lambda_i^- = \min_{j \in A_i} \lambda^{ij}$ .

Soleimani-Damaneh and Mostafaei (2009) prove that  $\lambda_i^+$  and  $\lambda_i^-$  are equivalent to the ones defined in Problem 3.

Now, let  $\theta_i$  denote the solution  $\theta$  associated with observation  $i$ , and  $\lambda_i^+$  and  $\lambda_i^-$  denote the maximal and the minimal of the solutions  $\sum_{j \in N^l} \lambda_j$  to Problem 3 for each  $i$ , respectively.<sup>21</sup> When  $\lambda_i^+$  is smaller than 1, it indicates that there is a firm with higher average products whose scale is larger than firm  $i$ .

Conceptually, this method studies whether there is some management inefficiency related to production scale or decreasing returns to scale technology (or both), although the method does not identify the percentage of such a difference between Firm 1's and Firm 2's production levels that is attributable to management inefficiency and returns to scale technology. For instance, suppose that there are two firms, where Firm 1 uses 10 units of labor to produce five units of goods and Firm 2 uses 15 units of labor to produce seven units of goods. If the technology is constant returns to scale, then by simply extrapolating Firm 1's production, Firm 1 should be able to produce 7.5 units by using 15 units of labor. Then, comparing Firm 2's production, the technology may be decreasing returns to scale, or there may be some inefficiency in managing the resources in Firm 2.

<sup>20</sup>See also Soleimani-Damaneh and Mostafaei (2009)

<sup>21</sup>A more detailed description of the numerical algorithm used to calculate  $\lambda^+$  and  $\lambda^-$  is found in the online Appendix. In the productivity analysis literature, a firm with  $\lambda_i^+ < 1$  is said to be *operating under increasing returns to scale* and a firm with  $\lambda_i^- > 1$  is said to be *operating under decreasing returns to scale*.

Figure 4 presents the bar graphs indicating the percentage of firms that have  $\lambda_i^+ < 1$  (dark color),  $\lambda_i^- > 1$  (light color), and other (white) in each group of firm sizes. To grasp this easily, we group the samples by number of employees into five groups, where groups 1 to 5 are firms with 1 – 10, 11 – 50, 51 – 100, 101 – 200, and 201 or more employees, respectively.<sup>22</sup> The white portions are the groups of firms with the highest average products. According to the categorization defined above, we calculate how many firms belong to each group and then divide these numbers by the total number of firms in each group, so that we can obtain the percentages of firms in each category of each group. We repeat this procedure for all sectors.

Figure 4 shows that in most of the sectors, the proportion of firms with the property  $\lambda_i^+ < 1$  is highest in Group 1. This is understandable, because their size is small and thus their production tends to be small compared with the optimal scale. Somewhat more strikingly perhaps, we observe that in many sectors, the large-sized firm category includes most firms with the property  $\lambda_i^- > 1$ , particularly in sectors 22 (Publishing, printing, etc.), 24 (Chemical products, etc.), and 25 (Rubber and plastic products), unlike in sectors 14 (Other mining and quarrying), 26 (Nonmetallic mineral products), and 29 (Machinery and equipment, etc.), which are the sectors with the low BCs in Table 3. The picture of the production functions that the analysis shows is typical of firms using large quantities of inputs than the one of the optimal level, and many large-sized firms are located in this region. From the analysis in Figure 1, we can see that many large-sized firms operate close to the frontier of the function. This suggests that even though the large-sized firms can produce in a relatively efficient manner, they may not be able to take full advantage of scale.

As pointed out by Feng and Zhang (2014), not considering the technological heterogeneity of an individual firm even within a single sector can cause bias in measuring returns to scale. Pham and Takayama (2017) measure returns to scale considering technological heterogeneity for each firm using a random coefficient stochastic frontier and find that most firms show constant returns to scale technologies. Given this finding, this disparity we find here could be due to inefficiencies only associated with scale. Our findings in this paper complement the work in Pham and Takayama (2017).

We should note that we have also computed the annual efficiency scores,  $\lambda_i^+$  and  $\lambda_i^-$ , for each sector. In this annual census, because the number of observations for each sector decreases dramatically, it becomes more difficult to obtain a general trend. As described in Chapter 3.3 of Daraio and

---

<sup>22</sup>This set of cutoff sizes ensures that we have enough firms in each group.

Simar (2007), under both the FDH and DEA methods, larger quantities of data provide more reliable estimates.<sup>23</sup> Nevertheless, similar to the analysis that includes all the years for each sector, the annual analysis also shows that the middle-sized firms tend to have low efficiency scores. On the other hand, the BCs of sectors 14, 26, and 29 tend to increase and the annual bar graphs of these three sectors are similar to those of other sectors. Finally, we also observe that there are still many low-efficiency firms in Figure 1.

### 3.3 Robustness Check

To check the robustness of our findings, we use a bootstrap method, called ‘order- $m$ ’.<sup>24</sup> In this exercise, for each firm, we randomly select  $m$  firms for  $m = 10, 25, 50, 75, 100, 150, 200, 1500, 3000, N^l$  for each sector  $l$ , and conduct the same analyses 500 times. In this way, we obtain 95% confidence intervals for each firm’s  $\lambda_+$  and  $\lambda_-$ . Note that even when  $m = N^l$ , we repeatedly draw the samples from the same set and thus we can obtain a confidence interval. Figure 5 shows the results from this exercise. We categorize firms into three groups such that the 97.5% quantile of  $\lambda_i^+$  is smaller than 1 (dark color), the 2.5% quantile of  $\lambda_i^-$  is greater than 1 (light color), and firms that do not belong to either group (white). It is reasonable that as  $m$  increases, the heterogeneity increases. Furthermore, the result that most firms fall into the category of the group with white color is consistent with Pham and Takayama (2017), although the methodology is different.

We also conduct the same procedure, order- $m$ , for computing efficiency scores. Our main findings still hold across different  $m$  exercises, namely that middle-sized firms tend to have lower efficiency scores and have more diverse scores compared with other groups. Finally, we also conduct the same analyses by using the output-oriented method, which show similar results.

## 4 Conclusion

In this paper, we have focused on the relationship between firm size and production efficiency using Vietnamese data. We have presented results indicating that middle-sized firms tend to produce less efficiently relative to small-sized and large-sized firms. Our study analyzes firms in a developing country in which small-sized firms dominate the economy and exhibit higher efficiency levels com-

---

<sup>23</sup>Detailed results are available from the authors.

<sup>24</sup>This is widely adopted method. For example, Hanson (2016) also uses this method to compute the efficiency score.

pared with middle-sized firms.<sup>25</sup> Moreover, the heterogeneity of firm size at the bottom of the U-shape across industries indicates the importance of sector-level analysis in investigating the missing middle. Our analysis also contributes to this point.

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

Furthermore, our findings suggest the necessity of sector-by-sector analysis when considering the underlying mechanism for the missing middle phenomenon. Merging data from multiple sectors might mix some features such as returns to scale and inefficiency associated with management for each firm in a different sector, and we may not be able to identify a driving force for the missing middle phenomenon. It is particularly important to study this measure at the sectoral level, as the underlying technology in each sector would not be that different in comparison with different sectors.<sup>26</sup>

Interestingly, our result shows that the size increases, the standard deviation increases across sectors and this feature is more eminent for middle-sized firms compared with small-sized firms. Although our analysis is a snapshot of the Vietnamese manufacturing industry, the diverse production efficiencies in the middle-sized groups can be thought of as a risk that small-sized firms would face in expanding their businesses. It would be interesting to see why the middle-sized groups' efficiencies vary more than the other groups.

---

<sup>25</sup>In the literature, there has been increasing interest in the relationship between firm size and characteristics such as 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, and those examining developing countries are rare. Furthermore, in previous studies, large-sized firms are found to be the most efficient (see Angelini and Generale, 2008; Leung et al., 2008).

<sup>26</sup>Another way to approach this problem is by assuming that the basic natures of the technologies are not so different between developed and developing nations; this measure could be studied at the sectoral level in developed nations and the outcomes compared with ours.

## List of Tables and Figures

Table 1: Summary of Enterprise Census Data Statistics (Firm Size)

| Sector    | # Observations | Size    |        |          | Capital  |         |           | Material |         |           |
|-----------|----------------|---------|--------|----------|----------|---------|-----------|----------|---------|-----------|
|           |                | Mean    | Median | St. Dev. | Mean     | Median  | St. Dev.  | Mean     | Median  | St. Dev.  |
| Sector 14 | 7799           | 83.78   | 32.00  | 222.45   | 743.10   | 205.91  | 50551.52  | 13143.63 | 1299.27 | 50551.52  |
| Sector 15 | 29245          | 108.83  | 14.00  | 315.35   | 1204.60  | 85.74   | 100334.98 | 17688.89 | 1045.80 | 100334.98 |
| Sector 17 | 6096           | 215.85  | 48.00  | 558.99   | 2134.34  | 316.06  | 189108.37 | 32257.84 | 2853.26 | 189108.37 |
| Sector 18 | 10086          | 361.67  | 86.00  | 738.70   | 3522.43  | 560.14  | 35198.79  | 11798.09 | 2071.88 | 35198.79  |
| Sector 19 | 3327           | 1200.54 | 248.00 | 3437.18  | 10714.11 | 1778.67 | 367323.46 | 47710.37 | 6023.83 | 367323.46 |
| Sector 20 | 10761          | 71.23   | 20.00  | 167.98   | 546.58   | 125.17  | 14636.06  | 4251.61  | 947.54  | 14636.06  |
| Sector 21 | 6905           | 75.54   | 30.00  | 185.63   | 845.78   | 219.67  | 60352.01  | 12597.05 | 2350.08 | 60352.01  |
| Sector 22 | 8378           | 37.98   | 10.00  | 82.01    | 623.51   | 102.64  | 30957.98  | 5725.01  | 711.91  | 30957.98  |
| Sector 24 | 6880           | 104.91  | 27.00  | 274.71   | 2077.98  | 270.06  | 121884.53 | 27718.49 | 3389.30 | 121884.53 |
| Sector 25 | 9414           | 90.76   | 27.00  | 212.75   | 1121.98  | 226.56  | 47436.30  | 14915.65 | 2705.25 | 47436.30  |
| Sector 26 | 11643          | 141.22  | 45.00  | 295.38   | 1704.91  | 298.53  | 183453.44 | 28664.16 | 1595.30 | 183453.44 |
| Sector 28 | 15662          | 55.59   | 16.00  | 147.56   | 706.88   | 143.32  | 35268.64  | 9276.54  | 1193.19 | 35268.64  |
| Sector 29 | 4170           | 94.52   | 26.00  | 212.15   | 1171.65  | 258.16  | 66127.21  | 15777.26 | 1821.86 | 66127.21  |
| Sector 36 | 10202          | 168.35  | 30.00  | 430.30   | 1827.53  | 245.96  | 56261.18  | 12378.17 | 1627.22 | 56261.18  |

<sup>1</sup> Size is measured by the number of employees.

<sup>2</sup> Capital and Material are measured in the home currency (not inflation adjusted).

<sup>3</sup> St. Dev., standard deviation.

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

| Year  | Total  | Sec 14 | Sec 15 | Sec 17 | Sec 18 | Sec 19 | Sec 20 | Sec 21 | Sec 22 | Sec 24 | Sec 25 | Sec 26 | Sec 28 | Sec 29 | Sec 36 |
|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 2000  | 7777   | 454    | 2197   | 324    | 468    | 217    | 705    | 350    | 238    | 363    | 417    | 950    | 505    | 196    | 393    |
| 2001  | 9147   | 549    | 2433   | 371    | 585    | 250    | 647    | 478    | 342    | 437    | 523    | 1034   | 683    | 266    | 549    |
| 2002  | 11155  | 720    | 2702   | 484    | 723    | 287    | 830    | 550    | 485    | 555    | 691    | 1095   | 1007   | 327    | 699    |
| 2003  | 12846  | 804    | 2819   | 529    | 867    | 318    | 1053   | 612    | 619    | 633    | 804    | 1174   | 1290   | 394    | 930    |
| 2004  | 15658  | 929    | 3298   | 640    | 1164   | 398    | 1142   | 771    | 892    | 754    | 1005   | 1354   | 1677   | 477    | 1157   |
| 2005  | 17316  | 916    | 3514   | 755    | 1229   | 436    | 1238   | 888    | 1027   | 846    | 1209   | 1429   | 1962   | 542    | 1325   |
| 2006  | 19132  | 863    | 3791   | 917    | 1341   | 390    | 1539   | 908    | 1471   | 980    | 1327   | 1402   | 2223   | 566    | 1414   |
| 2007  | 23832  | 1086   | 4495   | 1072   | 1842   | 490    | 1860   | 1152   | 1725   | 1130   | 1715   | 1689   | 3032   | 730    | 1814   |
| 2008  | 23705  | 1478   | 3996   | 1004   | 1867   | 541    | 1747   | 1196   | 1579   | 1182   | 1723   | 1516   | 3283   | 672    | 1921   |
| Total | 140568 | 7799   | 29245  | 6096   | 10086  | 3327   | 10761  | 6905   | 8378   | 6880   | 9414   | 11643  | 15662  | 4170   | 10202  |

*Note:* The number of observations is shown for each sector of each industry.

Figure 1: Efficiency Scores Averaged for Each Size in Each Industry

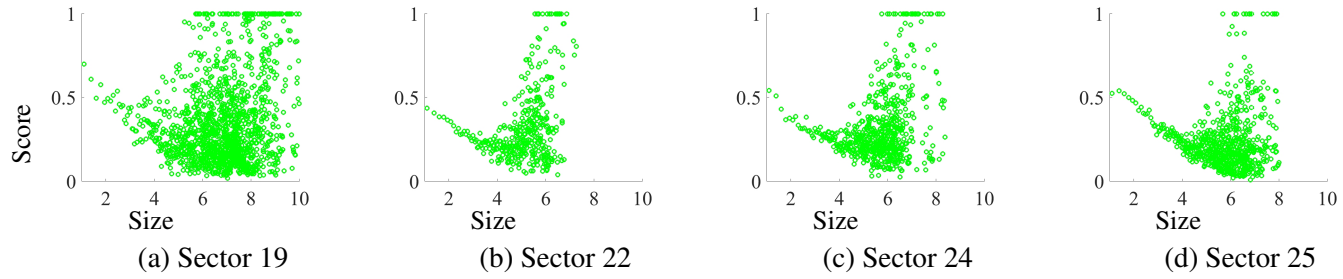


Figure 2: Ratio between FDH and DEA Scores in Each Industry

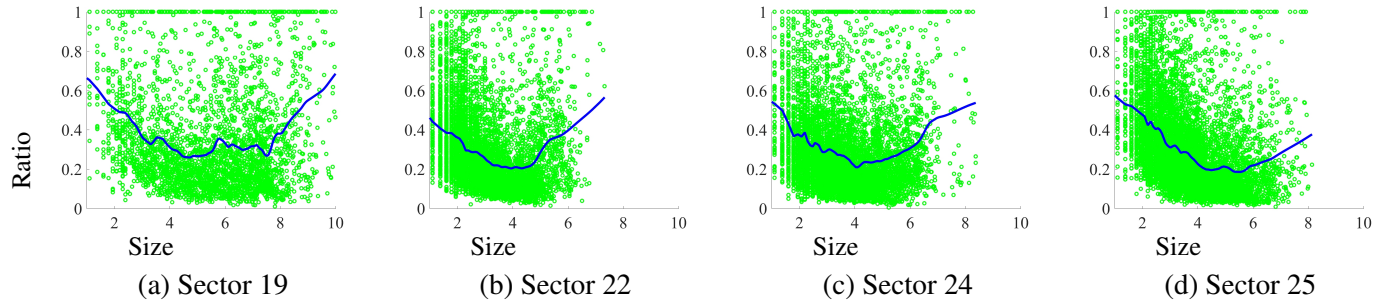


Figure 3: Standard Deviation and Efficiency Scores in Each Industry

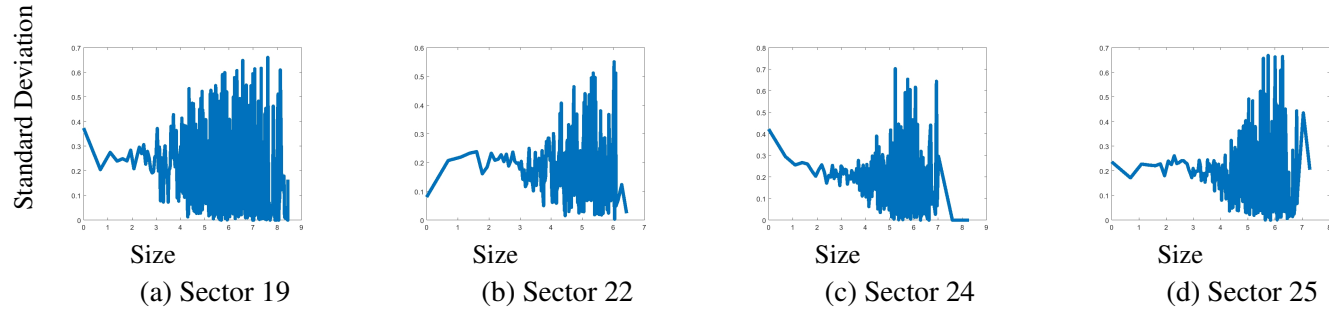
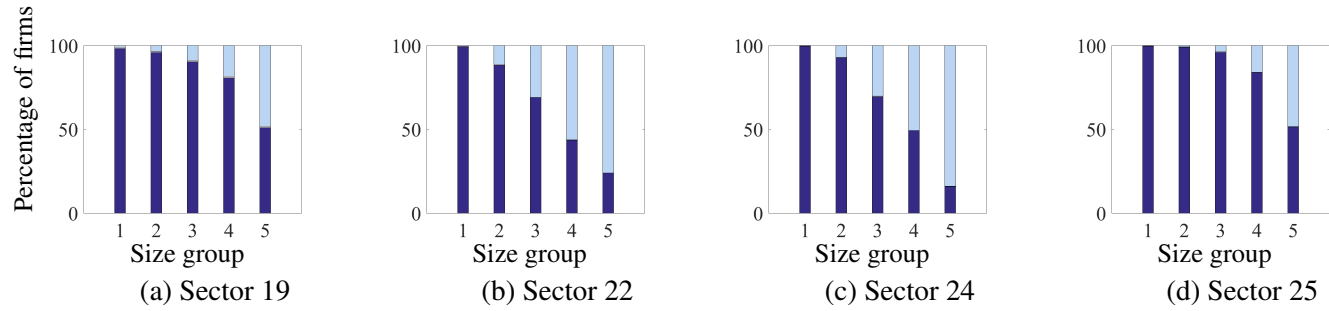
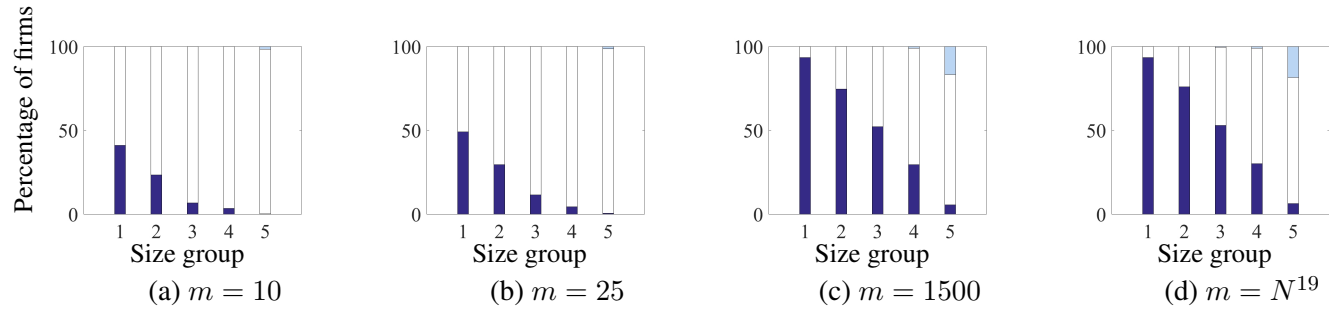


Figure 4: Production Scale in Each Industry



*Note:* The light color represents the portion of firms for which  $\lambda_i^- > 1$ ; the dark color represents the portion of firms for which  $\lambda_i^+ < 1$ ; the white color represents firms that do not belong to either group. The  $x$ -axis represents each group, namely  $n$  is Group  $n$  for each  $n = 1, \dots, 5$ .

Figure 5: Production Scale Using Order- $m$  in Sector 19



*Note:* The light color represents the portion of firms for which  $\lambda_i^- > 1$ ; the dark color represents the portion of firms for which  $\lambda_i^+ < 1$ ; and the white color represents firms that do not belong to either group. The  $x$ -axis represents each group, namely  $n$  is Group  $n$  for each  $n = 1, \dots, 5$ .

## References

- Acs, Zoltan J. and David B. Audretsch**, “Innovation, marketstructures, and firm size,” *Review of Economics and Statistics*, 1987, 69, 567–74.
- Angelini, P. and A. Generale**, “On the evolution of firm size distribution,” *American Economic Review*, 2008, 98, 426–438.
- Basu, S. and J. G. Fernald**, “Returns to Scale in U.S. Production: Estimates and Implications,” *Journal of Political Economy*, 1997, 105 (2), 249–283.
- Bentzen, J., E. S. Madsen, and V. Smith**, “Do firms’ growth rates depend on firm size?,” *Small Business Economics*, 2011, 39 (4), 937–47.
- Brown, D. J.**, “Equilibrium analysis with non-convex technologies,” *Handbook of Mathematical Economics*, 1991, 4.
- Cazals, Catherine, Jean-Pierre Florensa, and Léopold Simar**, “Nonparametric frontier estimation: a robust approach,” *Journal of Econometrics*, 2001, 106, 1–25.
- Dalton, Sherry, Erik Friesenhahn, James Spletzer, and David Talan**, “Employment Growth by Size Class: Firm and Establishment Data,” *Monthly Labor Review*, 2011, 134, 3–24.
- Daraio, Cinzia and Léopold Simar**, *Advanced Robust and Nonparametric Methods in Efficiency Analysis*, New York: Springer US, 2007.
- Diewert, W. Erwin, Takanobu Nakajima, Alice Nakamura, Emi Nakamura, and Masao Nakamura**, “Returns to scale: concept, estimation and analysis of Japan’s turbulent 1964-88 economy,” *Canadian Journal of Economics*, 2011, 44 (2), 451–485.
- Farrell, M. J.**, “The measurement of productive efficiency,” *Journal of the Royal Statistic Society*, 1957, 120, 253–282.
- Feng, Guohua and Xiaohui Zhang**, “Returns to scale at large banks in the US: A random coefficient stochastic frontier approach,” *Journal of Banking & Finance*, 2014, 39 (C), 135–145.
- Freeman, Jonathan B. and Rick Dale**, “Assessing bimodality to detect the presence of a dual cognitive process,” *Behavioral Research Methods*, 2013, 45 (1), 83–97.
- Hanson, Torbjørn**, “Efficiency and productivity in the operational units of the armed forces: A Norwegian example,” *International Journal of Production Economics*, 2016, 179, 12–23.
- Hartigan, J.A. and P. M. Hartigan**, “The Dip Test of Unimodality,” *The Annals of Statistics*, 1985, 13 (1), 70–84.
- Hsieh, Chang-Tai and Benjamin A. Olken**, “The Missing “Missing Middle”,” *Journal of Economic*

- Perspectives*, 2014, 28 (3), 89–108.
- Kalirajan, K.P. and R.T. Shand**, “Modelling and measuring economic efficiency under risk,” *Indian Journal of Agricultural Economics*, 1994, 49, 579–90.
- Klette, T. and Z. Griliches**, “The inconsistency of common scale estimators when output prices are unobserved and endogenous,” *Journal of Applied Econometric*, 1996, 11, 343–361.
- Leibenstein, Harvey**, “Allocative Efficiency vs. “X-efficiency”,” *American Economic Review*, 1966, 56 (3), 392–415.
- Leung, D., C. Meh, and Y. Terajima**, “Productivity in Canada: does firm size matter?,” *Bank of Canada Review*, 2008, 1.
- Liedholm, Carl and Donald C. Mead**, “Small Scale Industries in Developing Countries: Empirical Evidence and Policy Implications,” Technical Report, Michigan State University, Department of Agricultural, Food, and Resource Economics 1987.
- Little, I.M., D. Mazumdar, and J.M. Page Jr**, “Small manufacturing enterprises: a comparative analysis of India and other economies,” Technical Report 10118, A World Bank research publication 1987.
- Mairesse, Jacques and Jordi Jaumandreu**, “Panel-data estimates of the production function and the revenue function: What differences does it make?,” *Scandinavian Journal of Economics*, 2005, 107, 651–672.
- Pham, Hien Thu and Shino Takayama**, “Firm Size Distribution, Production Efficiency, and Returns to Scale: A Stochastic Frontier Approach,” *Working Paper*, 2017.
- Ramey, V. A.**, “Non-convex costs and the behavior of inventories,” *The Journal of Political Economy*, 1991, 99, 45–61.
- Ray, S. C.**, *Data Envelopment Analysis*, Cambridge, United Kingdom: Cambridge University Press, 2004.
- SAS Institute Inc.**, “SAS/STAT user’s guide 9.2,” 2008. Cary, NC: SAS Institute.
- Simar, L.**, “Detecting Outliers in Frontier Models: A Simple Approach,” *Journal of Productivity Analysis*, 2003, 20, 391–424.
- Soleimani-Damaneh, M. and A. Mostafaei**, “Stability of the classification of returns to scale in FDH models,” *European Journal of Operational Research*, 2009, 196, 1223–1228.
- **and M. Reshadi**, “A polynomial-time algorithm to estimate returns to scale in FDH models,” *Computer and Operations Research*, 2007, 34, 2168–2176.

- Steel, W. F. and L. M. Webster**, “How Small Enterprises in Ghana Have Responded to Adjustment,” *World Bank Economic Review*, 1992, 6, 423–438.
- Tybout, James R.**, “Manufacturing firms in Developing Countries: How Well Do They and Why?,” *Journal of Economic Literature*, 2000, 38, 11–44.
- , “Correspondence,” *Journal of Economic Perspectives*, 2014, 28 (4), 235–236.
- , “The Missing Middle, Revisited,” 2014. mimeo.
- Wilson, Paul W.**, “FEAR 1.0: A Software Package for Frontier Efficiency Analysis with R,” *Socio-Economic Planning Sciences*, 2008, 42, 247–254.