RETURN TO SCALE, THE MISSING MIDDLE AND POLICY IMPLICATIONS FOR SUPPORTING FIRMS' DEVELOPMENT AND TECHNOLOGY UPGRADING

Vu Hoang Đat[1]
The Centre for Analysis and Forecasting, Viet Nam Academy of Social Sciences

Pham Thu Hien
CSIRO, QLD, Australia 4006

Nguyen Nam Hai
Department of Planning- Finance, Ministry of Science and Technology

Abstract:

In this study, we investigate the "missing middle" phenomenon of firm size distribution in Viet Nam. Results imply existence of both the "missing middle" as well as the increasing return to scale in aggregate production function of most industries in Vietnam. Such co-existence suggests that there are forces other than those traditionally mentioned in economic literature (Tybout, 2000) affecting the firm size. However, there are heterogeneities in the return to scale within industries that the middle-size firms have the lowest return to scale, compared with those of their small or large size counterparts. This result implies when Viet Nam's firms develop into middle-sizeones they face significant challenges, not only in term of further developing into bigger sizes but also in term of remaining size efficiency comparing with small ones.

Keywords: *Economy; Technology development; Technology upgrading; Policy; Viet Nam.*

Code: *20122301*

1. Introduction

Firm size distribution has been a particular concern of economists for nearly a century. On entering markets, firms face a selection process that will cause less efficient firms to decrease their size and eventually exit while more efficient ones start to grow. Survivors will then choose to allocate resources and maximize their profit in the given macro environment and within a certain size. In other words, the accumulation of the underlying dynamics that results from a firm's decision on entry, development or exit evolves a certain firm size distribution (hereafter FSD). FSD is, therefore, an endogenous choice of firms and its analysis can provide a possible interpretation of the behavior of firms in the market, given their opportunities, constraints, and levels of efficiencies.

[1] Author's contact email address: dat.hoangvu@gmail.com

Returns to scale (RTS) have long been known to have significant impacts on the structure of FSD of the economy. Ijiri et al. (1977) and Lucas (1978) developed models where constant returns to scale (CRS) imply Gibrat's law (i.e. the probability of a given proportional change in size during a specified period is the same for all firms in a given industry, regardless of their initial sizes *(Mansfield, 1962)*). More recent studies also assume CRS when deriving a linear Zipf plot of FSD *(Luttmer, 2007)*. The relaxation of the CRS assumption can have significant effects on the shape of FSD. However, in the literature, there is no clear conclusion on the impact of scale elasticity onto FSD. Among the few related studies, Vinning (1976) finds that such a Gibrat-style process generates a convex rather than concave Zipf line.

Hence, departure from Zipf's law or deviation from a Pareto firm size distribution may imply "something inherent in the nature of size that causes a progressive decline in the growth rate of a firm as it expands its activities" *(Vinning, 1976)*. This linkage may provide a hint to explain the "missing middle" phenomenon in developing countries. In particular, in such economies nonconstant returns to scale may contribute to a deviation from the Pareto firm size distribution (FSD).

This study is dedicated to the examination of the RTS technology structure in Vietnam and its possible linkage to the characteristics of FSD. Vietnam is a transitional developing economy that is undergoing fundamental changes from a central planned to market oriented mechanism. After 20 years of reform, the market is gradually shaping. However, the economy is still characterized by numerous imperfections. Moreover, the deviation of FSD from Zipf's law and the log-normal distribution as well as the existence of a "missing middle" has been more and more characterized.

Apart from external reasons such as financial constraint or corruption, such a "missing middle" phenomenon in FSD may arise as the reflection of the industrial composition of the economy as a whole. In particular, if an industry is characterized by increasing returns to scale (IRTS), one may expect a larger proportion of large-sized firms, as firms with more capital investment will reap more from economies of scale in that industry. One example is the automobile industry. This industry requires large-scale assembly lines, thus, scale economies are far more important and the average size of automobile firms is much larger compared to those of labor-intensive industries such as the apparel sector. Such heterogeneity in industries certainly has significant influence on the distribution of firms across size groups.

We employ numbers of labourers as the measure of the firm size given its popularity in the literature. We aslo define middle-sized firms based on relative sizes across all firms rather than definitions for regulations such as that in Decree No. 56/2009ND-CP dated 30/06/2009 of the Government of Viet Nam for supporting small and medium firms. According to the Decree, eligiblely

supported firms are classified into micro, small and medium ones. Medium firms in the agricultural and manufacturing industries have a size of 200 to 300 labourers. Meanwhile, the corresponding firms in the service sector have a size of 50 to 100 labourers.

The middle size is understood as those in the middle of the firm size distribution. Based on the real distribution of Viet Nam's firms in the study period, we define the middle-sized firms as those having 100 to 200 laborers regardless of studied industries. Meanwhile, firms having less than 100 or more than 200 are regarded as small or large size ones respectively. The definitions are applied for the whole study period to assure comparability and consistency.

This study focuses on investigating the nature of RTS at the sectoral level. As far as we know, there have not been any published works which investigate the "missing middle" phenomenon of developing countries at the sectoral level. In addition, there have been no attempts in comprehensively examining the nature of RTS in Vietnamese industries. Therefore, it is of great interest to quantify the RTS of different sectors in Vietnam as well as any correlation that may exist between the FSD and RTS in Vietnamese industries.

The contribution of this study to the literature is twofold. The first contribution lies in its empirical findings. Since it is the first time that the "missing middle" is analyzed at the sectoral level, the study provides an opportunity to deepen the understanding on the nature and origin of the phenomenon. It offers a comprehensive picture of the nature of returns to scale in different industries of Vietnam. The study, thus, captures the heterogeneity of RTS among industries and shed light on the relationship between RTS and the emergence of FSD as well as the "missing middle". Although the time-span of the dataset is relatively short, the analyzed period has been a process of intense structural changes and our regression results can be compared with the existing estimates.

The second contribution lies in investigation of RTS of firms with different sizes within industries. The results exhibit that the middle-size firms have the lowest RTS, compared with their small or large counterparts. The lowest RTS of the middle-size firms imply that small sized firms in Vietnam face significant challenges to grow into middle-sized ones.

The study is divided into five sections. Apart from the introduction, the second section discusses the data source. The third section presents an industry level analysis of FSDs for Vietnam. The fourth section explores industry heterogeneity of RTS of industries and some possible correlations between IRTS and the "missing middle" as well as discussions of underling reasons for the phenomenon. The fifth section concludes the study.

2. Data description

In this study, we utilize the micro firm level dataset of the Enterprise Census conducted by the General Statistics Office of Vietnam (GSO). The census has been conducted annually since 2000 to provide information on firms and their activities including revenue, labor, capital, demographic data, firms' ownership and employment, among others. With such information, the study focuses on estimating RTS at both the sectoral level as well as across different size groups at the firm level within one industry. In general, the use of micro data not only provides more freedom to control for a larger number of variables, compared to traditional estimations but also enables me to control for specific sectoral or localization characteristics while avoiding aggregation related problems.

The dataset contains firm information for a period of 9 years - from 2000 to 2008. In this study, the exclusion of recent year data (from 2009 up to now) is due to the fact that the Vietnamese economy is suffering from the global financial crisis. We include all industries which have at least 200 observations for one given year. This means we focus our attention on 32 industries of the economy with the sample size ranging from approximately 1,500 to 160,000 observations per industry.

We exclude inconsistent data from our sample, i.e. observations which are recorded twice for the same firm in the same year, observations with negative or zero revenue values, negative or implausibly large number of employees, observations with missing information of output and inputs (labour, capital and intermediate material). We also identify and eliminate outliers if a rationale for exclusion is found. Outliers are identified using Cook's distance method (i.e., observations are identified as extremes according to their distance from the nearer quartile and the magnitudes of their influence are measured and considered).

In this study, an industry is defined as a group of firms producing a homogeneous product or a number of goods that are closely related. After disregarding industries with less than 200 observations in one year, the final dataset consists of 33 over 62 industries of the economy. Despite the fact that they cover only about 50% of total industries, the included industries represent more than 90% of the total number of firms in the Enterprise Census. The first 18 industries (ISIC 11-36) are manufacturing industries and the last 14 sectors are non-manufacturing ones (ISIC 45-93).

Real revenue is used as the proxy for output. There are the long-term debate of the choice between revenue or production function in RTS estimation. The major problem here is that at firm level, prices and quantities are not well measured and the revenue function, instead of gross output or cost function, is normally utilized to estimate RTS. In the literature, there are arguments

suggesting that elasticities of labor and capital, in a revenue estimate, may be downward biased and such the bias is inversely proportional to the mark-up.

One proposed solution is to deflate the nominal output of the firm by replacing the individual price with the common price index of the industry. This method, however, cannot solve the problem since changes in sector prices are substantially diversified and correlated with changes in labor and capital *(Klette and Griliches, 1996)*. There are, however, various reasons supporting the use of revenue function. According to Jacques and Jordi (2005), the introduction of individual output price into the production function does not markedly modify the estimate of RTS and therefore a revenue function does not cause major divergence in RTS 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 well measured and compared across firms. All considered, real revenue can be considered an appropriate proxy for output in this study.

Other input variable are defined in accordance with the literature. Labor input is measured by the total income of employees in a firm. This includes total wage and other employees' costs such as social security, insurances and other advantages. The value of intermediate material includes costs such as fuel and the value of other materials. Capital is measured as the total asset of the firm. All input values are, then, corrected with inflation to get the real value.

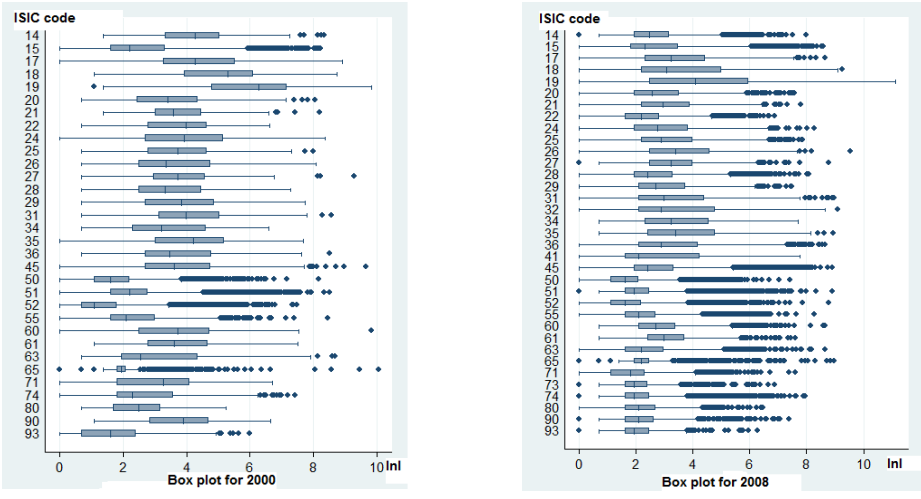
3. The firm size distribution at the sectoral level in Vietnam

This section is devoted to the FSD analysis at the sectoral level in Vietnam. We use the number of employees as the measure of firm size as discussed in the introduction.

For an overview of FSD in Vietnam at a sectoral level, we first consider the box plot which reports the overall distribution of the number of employee variable in various sectors. A box plot is a very useful tool to demonstrate differences between the number of employees in different sectors without making any assumptions of the underlying statistical distribution. The bottom of each box indicates the25th percentile of the distribution, the top reports the 75^{th} percentile and the line in the middle is the 50^{th}percentile or the median. The spacings between the different parts of the box help indicate the degree of dispersion (spread) and skewness in the data, and identify outliers.

The overall impression is that all analyzed industries exhibit relatively similar distributions with a severely right skew trend in FSD. This right skew trend indicates the concentration of small firms in most sectors of Vietnam. Moreover, such a concentration tends to increase overtime in the development process.

The first overall impression of Vietnamese economy is that all analyzed industries exhibit relatively similar distributions with a severely right skew trend (i.e., most sectors are characterized with the concentration of small firms). In the majority of sectors, the boxes in the box plot are consistently located to the left so that the right tails are longer, and the medians are closer to the left line of the box. Moreover, the means are consistently to the right of their medians in all sectors. These indicate the obvious right skew. An important note here is that, the distribution of firm size seems to be more even in the early years of analysis (in 2000) and the skew become severe in recent years as the boxes continuously shift to the left of the distribution, strengthening the right skew of distributions. These right skew trends indicate the concentration of small firms in most sectors of Vietnam.



Source: Authors' calculation with data from Enterprise Censuses 2000 and 2008

Figure 1. The box plot of the employee number logarithm across industries in 2000 and 2008

There is, however, clear evidence of significant heterogeneity in firm size across industries. The median of firm size logarithm ranges from around 1.5 to 5 across sectors while the mean fluctuates in the interval 1.84 to 4.91. Also, most outliers are in the right end, indicating the existence of extremely large firms, compared to the average level of each sector. In general, service sectors tend to have much smaller average size and less fluctuation in size compared to manufacturing ones. Meanwhile, outliers tend to appear more often in service sectors. Again, the fluctuation in size tends to increase over time with the increase in the number of outliers in each sector throughout the analyzed period.

In Vietnam, the largest sector is the whole sale and commission trade service (ISIC 51), which contributes approximately 25% of the total firms. It is, then,

followed by two other major industries, construction and retail service (ISIC 45 and 52), which comprise 12.5% and 11% respectively of the firm population. Meanwhile, sectors like extraction of crude petroleum and natural gas or tobacco contribute just minor proportions of less than 0.3%. In other words, the majority contribution of employment comes from labor-intensive service industries, in particular from the wholesale and retail sector (ISIC 51, 52).

Since 1986, Vietnam has experienced a continuous shift of employment from agriculture toward manufacturing and service sectors. Differently from newly industrialized countries like Korea or Taiwan (China) during their initial stage of development, employment share of service sector increases much faster than those of manufacturing ones in Vietnam. In 2013, the proportion of service sectors is 32.0% while all manufacturing industries only employed 21.1% of the labor force. In other words, the tertiary sector has absorbed most of employment loss from the agriculture sector.

In the literature, the labor shares amongst sectors in the economy are expected to closely relate to income distribution and economic growth[2]. The rapid increase of service activities at the initial development stage may stagnate economic growth as the increase may concentrate mainly on labor intensive and low-income services like retail or labor-intensive transport. However, there is also a chance that such an increase leads to fast catch-up if such rapid growth comes from new business services such as finance or telecommunication.

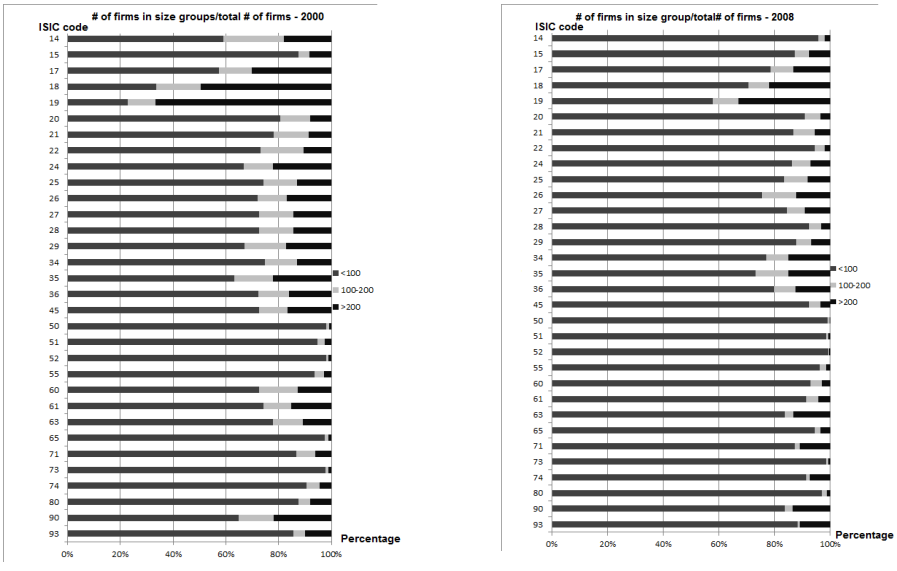
The data of Vietnam prevails evidence supporting the first hypothesis. Despite recent emergence of new services such as finance and IT, it appears that retail trade and small mechanic, the two biggest service sectors have achieved most gain of employment. In the 9 year period from 2000, more than 600,000 laborers find jobs in the small mechanic service sector (repair of motor vehicles, motorcycles and personal and household goods) while the retail sector creates nearly 300,000 job vacancies.

On the other hand, manufacturing industries, despite their minor contributions to the total firm population, take a significant part in the number of large firms. In 2008, approximately 20% of firms in industries such as extraction of crude petroleum and natural gas, mining, tobacco, garment, manufacture of furniture, garment, manufacture of other non-metallic mineral products and (ISIC 11, 14, 19, 18, 36, 26) are large-scaled. This trend is explainable since these industries require huge initial investment and are supposed to reap more from economies of scale than other sectors. In fact, one may find in the literature evidence of similar large scale bias in such sectors in both developed and developing countries. The distinct feature of Vietnam, however, is the pronounced skew

[2]Clark-Fischer-Kuznets, in their studies, suggest that at a low level of income, agriculture is dominant. As the economy grows, manufacturing industries capture more employment, followed by service sectors but at a slower rate. Service sectors only dominate in employment beyond a certain level of income.

toward very large firms (firms with more than 500 employees). This very large-sized firm group constitutes more than 50% of total employment of selected manufacturing industries in the period while levels of other neighboring countries like Malaysia or Thailand were only around 30%.

In general, the heterogeneity of FSD between service and manufacturing sectors supports the hypothesis that industrial decomposition contributes to the emergence of the "missing middle". In Vietnam, the majority of employment comes from labor-intensive service sectors which comprise of mainly small-scaled firms. Meanwhile, manufacturing sectors, despite their small contribution in the total number of firms, represent a concentration of large-scaled firms in the economy. As a result, the coexistence of industries with small-scaled firm dominance and sectors which favor large-scaled production provides a possible explanation to the existence of the "missing middle" in Vietnamese economy. The question to be raised here is whether the "missing middle" is just the result of sectoral decomposition? To answer this question, one may further look at the structure of FSD in each industry.



Source: Authors' calculation with data from Enterprise Censuses 2000 and 2008

Figure 2. The number of firms by size group - 2000 and 2008

Figure 2 presents the proportion of number of firms across size groups (the percentage of firms in each group over the total number of firms in each sector) in 2000 and 2008. Figure 1 provides a visual evidence of the existence of the dual distribution in various industries in Vietnam. Among different industries, manufacturing sectors (sectors with ISIC code from 11 to 33) show more obvious "missing middle". Industries that exhibit more pronounced "missing middle" include the extraction of crude petroleum and natural gas, garment and

tobacco. These are also industries which are supposed to reap the most from scale economies. In the literature, the dual structure has been mainly confirmed in manufacturing industries in many developing countries like India, Thailand or the Philippines. In fact, most evidence of the dualism of FSD in developing countries in the literature comes from manufacturing sectors, where scale effects are considered to be undoubtedly important. The "missing middle" observed in manufacturing industries in Vietnam, once again, confirms the importance of scale economies in FSD development of manufacturing sectors.

The "missing middle", however, can also be witnessed in service sectors in Vietnam. Different from previous research, this study contributes to the literature by analyzing the "missing middle" not only in manufacturing but also in non-manufacturing sectors. In Vietnam, although the trend is much less severe compared to manufacturing ones, the dualism structure exists in most service industries. Looking at Figure 1, the retail and whole sale service and hotel and restaurant (ISIC 51, 52, 55) are the only sectors where there is no clear evidence of the "missing middle", due to the overwhelming importance of small firms. Other service sectors such as health care, sewage and refuse disposal, recreational and cultural and transport activities exhibit much clearer sign of "missing middle". These are also sectors that are dominated by state-owned enterprises.

All considered, one may see that there is evidence of heterogeneity of firm distribution across sectors and among size groups within one sector in the economy. Such industrial decomposition can be a factor contributing to the emergence of the "missing middle". One possible explanation is the different nature of RTS of each industry or each size group. The next section, therefore, is devoted to examining the returns to scale nature of different industries in Vietnam as well as in each size group to capture any possible correlation between those factors and the "missing middle" in Vietnam.

4. Returns to scale and the "Missing middle"

Despite the simplicity of the RTS formulation, empirical estimations of RTS have been subjected to various issues over the years. An overview of the debate in the literature can be found in Karsten (2005).

Direct methods for measuring internal IRTS include survivor technique[3], engineering estimates and econometric estimates (both parametric and non-parametric). Econometric estimates offer another and, probably, the most popular way to estimate IRTS using accounting records. In such studies, actual costs, outputs and other characteristics of firms are gathered as cross-section, time series or panel data and are utilized to derive a relationship between costs

[3] Survivor technique is analysises of firm sizes with probabilities of development and decline.

and outputs. Being the most popular method, econometrics is widely applied, using both the parametric and non-parametric approach, at either the micro, regional or macro level.

Using parametric methods, different RTS structures have been found to prevail in different industries and different countries. For example, Christensen and Greene (1976) find significant IRTS in US electric power generation. Baldwin and Gorecki (1986), using ordinary least squares (OLS) estimate, also finds IRTS of about 10% in107 manufacturing industries in Canada. Hall (1990) reports considerably large RTS (using instrumental variables (IV) estimation). Empirical tests in Rumelt and Wensley (1981), however, suggest that "scale economies... are much less important than stochastic growth processes". Westbrook and Tybout (1992) also finds RTS values varying from 0.8 to 1.2 but the coefficients are not statistically different from unity and the null hypothesis of CRS cannot be rejected. The study compares various estimates namely OLS, between and within estimator, j^{th} difference and IV estimator and concludes that IV estimators are the most efficient estimators as they take into consideration sunk costs and are not biased by unobserved firm effects and measurement errors.

In this analysis, we also follow the similar approach to investigate RTS of industries in Vietnam. The first and foremost task in parametric RTS estimation is to choose an appropriate functional form for the production frontier. In the literature, firm technology can be represented by a number of production functions such as Cobb Douglas (CD), Constant Elasticity of Substitution (CES), generalized Leontief or transcendental logarithmic (trans-log) function. The translog function is attractively flexible since it contains both linear and quadratic terms with the ability of using more than two factor inputs. This allows a non-linear relationship between the output and the production factors. Moreover, the main characteristics of this type of production function is that the elasticity of substitution varies among factors, which allows us to access the substitutability and complementarity of different inputs in the production. Another advantage of this function, unlike Cobb- Douglas function, is that it does not require rigid assumptions namely: perfect substitution between production factors or perfect competition in the production market.

The translog function can be approximated by second order Taylor series *(Christensen et al., 1973)*

$$lnR = \beta_0 + \sum_{i=1}^{n} \beta_i lnX_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \beta_{ij} . lnX_i lnX_j \qquad (1)$$

where R is the total revenue of the firm; X_i is the inputs including real stock of capital, labor and material input respectively; β_0, β_1 are the constant term and the first derivatives, β_{ij} are second derivatives and cross second derivatives.

The output elasticity with respect to the i[th] input is given by:

$$\frac{\partial lnR}{\partial lnX_i} = \beta_i + \sum_{i=1}^{n} \beta_{ij} lnX_j \quad (2)$$

RTS are given by the scale elasticity at each specific point:

$$S = \sum_{i=1}^{n} \frac{\partial lnR}{\partial lnX_i} = \sum_{i=1}^{n} \beta_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ij} \, lnX_{ij} \quad (3)$$

Then, testing for constant returns (i.e., S = 1) is straightforward.

The first step of the analysis is to conduct the unit root test to check for the stationary status of variables[4]. The results of the standard Fisher-type test do not show any problem with unit root in the data.

Another issue is to check the presence of heteroskedasticity. Although OLS results are consistent with heteroskedasticity, its variance covariance matrix may change and any statistical test may be biased. We use Breusch-Pagan test to investigate the extent that the models suffer from heteroskedasticity. The results suggest the existence of heteroskedasticity in the dataset. To account for heteroskedasticity, we compute the heteroskedasticity consistent variance covariance matrix and use the result to test for CRTS.

The multi-collinearity problem[5] is also examined at this stage.4 The test results show that there is no evidence of severe multi-collinearity in the dataset.

The most obvious problem in OLS production function estimation, however, is the end ogeneity bias as OLS fails to be consistent when repressors and the error term are correlated. As pointed out by Marschak and Andrews (1944), production function estimation gives rise to an end ogeneity since inputs are assumed to be exogenous in the regression. The two standard and common methods to control for the end ogeneity problem are IV and FE/RE estimates.

The study begins with the traditional OLS estimator with robust standard error. The fixed or random effect is estimated next to control for individual heterogeneity by removing firm specific effects. The choice between fixed and random effect is decided based on Hausman test result. In general, most industries statistics are in favor of fixed effect estimation.

IV estimation is used to correct for potential end ogeneity problems for the capital input. In this study, we follow Levinsohn and Petrin (2003) and use intermediate input as the instrument for the endogenous variable - capital. Intermediate input seems to be appropriate for the dataset since it can provide

[4] Such the test is vital for co-integration and causality sequences and is carried out to avoid spurious regressions.

[5] Multi-collinearity may cause the model to be unidentified or it is harder to separate the impact of variables on output as it increases the standard error and widen the confidence interval. We check multi-collinearity by comparing the Variance Inflation Factor (VIF) score across variables. V IF > 10 indicate high collinearity.

robust standard errors and eliminates serial correlation problems to the estimation. The intermediate input, moreover, is less costly to adjust and may respond more fully to the entire productivity term.

In general, the chosen techniques are used to address three main possible causes of bias: (i) the presence of individual heterogeneity embodied in the error term, possibly correlated with all or parts of explanatory variables; (ii) the likely end ogeneity of some variables due to their simultaneous determination; and (iii) the presence of errors in variables such as measurement error.

Main empirical results

The main result is presented in Table 1. The columns of the table report different estimation results (OLS, IV, FE/RE) while the rows refer to different industries (denoted by the ISIC code). For each estimate, the table reports the RTS level as well as the result of the Wald test for CRTS. The "0" value indicates that it fails to reject the null hypothesis of CRS while the "+" and "-" sign imply that the CRS hypothesis is rejected in favor of IRTS/DRS respectively. For FE/RE model, the choice between random or fixed effect models is determined based on the result of Hausman test. In the table, "*" indicates the reported result comes from random effect estimation. The rest of the results in the column come from fixed effect estimation.

In general, the fit of all models is good with high level of R-squared (ranging from 0.5 to 0.9). Moreover, most coefficient estimates are economically meaningful and significant at 1% level. The RTS estimates are also plausible and lie well within the interval of (0.7, 1.7). As mentioned previously, the heteroskedasticity corrected variance covariance matrix is used in Wald test for CRTS[6].

Compared to the usual variance covariance matrix, the reported results tend to inflate the standard errors of the coefficient estimates and thus, CRS is thus rejected less often. It is also worth noting that heteroskedasticity correction does not change the results from constant to IRTS. From the Table, one may see that many industries operate under IRTS technology (23 over 32 sectors in OLS and FE/RE estimate and 12 over 32 in IV estimate).

Table 1. Parametric results of RTS in different industries in Vietnam

ISIC Code	OLS		IV		FE or RE		ISIC Code	OLS		IV		FE or RE	
14	0,97	0	0,97	0	1,00	0	36	1,12	+	1,03	0	1,04	+
15	1,13	+	1,16	+	1,10	+	45	1,01	0	0,84	-	0,85	-
17	1,02	+	0,95	0	1,07*	+	50	0,98	-	0,95	-	0,99	-

[6] The result of Wald test for CRS using the normal variance covariance matrix is available upon request.

ISIC Code	OLS		IV		FE or RE		ISIC Code	OLS		IV		FE or RE	
18	1,04	+	0,90	0	1,08	+	51	1,07	+	1,09	+	1,11	+
19	1,04	+	0,21	0	1,09	+	52	0,88	-	0,75	-	0,91	-
20	1,02	0	0,89	-	1,02	0	55	1,27	+	1,24	+	0,90	0
21	0,98	0	1,01	0	1,12	+	60	1,04	0	0,88	-	1,02	0
22	1,03	+	1,19	+	1,20	+	61	1,04	0	0,84	0	1,01	0
24	1,20	+	1,10	+	1,24	+	63	1,04	+	0,84	0	1,23	+
25	1,04	+	1,13	+	1,22*	+	65	1,70	+	1,70	+	1,09	+
26	1,16	+	1,06	+	1,06*	+	71	1,27	+	1,43	+	1,16*	+
28	1,07	+	1,08	+	1,13	+	73	1,03	+	0,98	0	1,15	+
29	1,07	+	0,96	0	1,05	+	74	1,23	+	1,03	0	1,22	+
31	1,11	+	0,83	0	1,17	+	80	1,22	+	1,24	+	1,21	+
34	1,23	+	1,52	+	1,17	+	90	1,00	0	0,94	0	1,00	0
35	1,08	+	0,97	0	1,16	+	93	1,13	+	1,06	0	1,11*	+

Source: Author's estimation with data from Enterprise Censuses 2000-2008

To facilitate the interpretation, we have summarized the incidence of RTS in different econometric specifications. Table 2 reports the percentage of sectors that exhibit IRTS, CRS or DRTS over the total number of sectors (32 sectors in this study). From the table, one may see that results of RTS are quite diversified across sectors. There are also differences in RTS of different econometric specifications. However, the switches are either between increasing and constant returns or between decreasing and constant returns and never between increasing and decreasing returns to scale. In general, OLS and FE/RE model report the highest number of IRTS industries (more than 70% of total sectors operates under IRTS when using these two estimates) while IV model shows the largest incidence of CRTS among the three. Among industries, manufacturing sectors have more uniformed bias toward IRTS. The incidence of scale economies can be witnessed in all models in many heavy industries (ISIC 15, 22, 24, 25, 26, and 28) and in OLS and FE model for most of manufacturing sectors. It is quite explainable since heavy industries tend to reap more from economies of scale. The results in service sectors are more diversified. While modern services like finance, machinery and equipment renting and education (ISIC 65, 71, 80) show IRTS, labor-intensive sectors tend to exhibit DRTS (ISIC 50, 52). Other services like water transport or sewage and disposal show CRTS technology in the production.

In addition, we also acknowledge the heterogeneity of input contribution across industries in the economy. In general, all inputs (labor, capital and intermediate inputs) make positive and highly significant contribution to revenue of a firm.

However, capital tends to have a much more important impact in the production process compared to labor and intermediate input in most manufacturing industries. Meanwhile, labor seems to be the major factor in service sectors. This finding goes well with theoretical predictions since capital tends to have more importance impacts on production in manufacturing industries while the role of labor tends to be dominant in service sectors.

Table 2. RTS Estimates Summary

Econometric specification	**OLS**	**IV**	**FE or RE**
IRTS (%)	71.8	37.5	71.8
CRS (%)	21.8	46.9	18.75
DRS (%)	6.4	15.6	9.4

Source: Author's estimation with data from Enterprise Censuses 2000-2008

At the first glance, the above results raise further questions rather than providing a consistent answer to the relationship between RTS and the "missing middle". Nevertheless, we are able to make some initial remarks while sorting out the econometric results.

First, estimated results suggest that IRTS do exist in Vietnam. More than two third of industries show evidence of IRTS in both OLS and FE model. However, the IRTS incidence is not similar across sectors. Evidence of IRTS can be seen more in manufacturing rather than service sectors in Vietnam.

Such a finding partly supports the hypothesis that the "missing middle" in the FSD of Vietnam is partly caused by industrial composition. In general, if one industry is characterized with IRTS – that is, the average cost of producing one unit of product decreases as total output increases - it is expected that large scale is preferable in such a sector. The more capital required in a production process, the greater the scope for reaping scale economies, and therefore the larger the optimum size of firms. Since most manufacturing industries require large scale investment, they tend to operate in large-scale firms while service firms tend to operate in small and medium size groups. In this way, the size distribution of firms in an economy depends, to a significant extent, on industrial composition of the economy. Since Vietnamese economy is dominated by labor intensive services such as whole sale or retail services, we can expect a high concentration of small firms in the economy while the existence of considerable very-large firm group in the economy may be explained by the bias toward large scale production of heavy industries.

Second, one may see the coexistence of the presence of IRTS and the "missing middle" at the sectoral level. In fact, all sectors that exhibit the "missing middle" (except for ISIC 90), also have evidence of IRTS in at least two out of three model specifications. Sectors which do not show the "missing middle"

like ISIC 50, 52 exhibit DRTS. In other words, there is also supporting evidence the correlation between the existence of the "missing middle" and scale economies of the industry.

However, we also note another fact that despite the existence of IRTS in almost all sectors in the economy, the level of IRTS is not high but just around 1.1% with many sectors' RTS being approximately unity. Does it mean that these sectors are actually operating near the optimal scale?

In other words, there is a possibility that IRTS are just results of the aggregation effect. The puzzle can probably be explained by the fact that the above parametric estimates only focus on the mean and leaving no opportunities to explore the difference in RTS within a particular sector. It is expected that firms with different scale may be subjected to different RTS which deviates from the mean.

To check the hypothesis, we use the quantile regression technique to assess the RTS of different size groups within one sector. The advantages of quantile regression include the robustness toward non-normal errors and outliers as well as the richer characterization of the data, which allows us to consider the impact of a covariate on the entire distribution of the dependent variable, not merely its conditional mean.

The regression results confirm the heterogeneity of RTS within one industry, especially in sectors which exhibit clear evidence of the "missing middle". In general, sectors with bi-modal FSD have the lowest RTS in the middle-sized firm group (firms with from 100 to 200 employees). The small sized firm group, on the other hand, yields the highest RTS level among the three size groups. Such U-shape RTS, however, cannot be found in sectors which do not exhibit "missing middle" like ISIC 73, 50, 51, 52. Such evidence again confirms the correlation between RTS nature and the "missing middle" phenomenon in Vietnam.

Challenges to middle-sized firms

Vietnamese firms face clear obstacles in growing as evidenced by the lowest RTS of the middle-size ones. When firms grow into the middle-sized ones, apart from popular challenges of changes in input and output markets, interactions with other players in the markets, we discuss two additional obstacles which potentially hinder performance of firms in the middle-sized group. They include insufficient investment and fails in technology upgrading. In the study, we only discuss mechanisms of adverse impacts of the two phenomenon. Empirical tests of these hypotheseses should be done in a separate study with data of investment and technology upgrading.

In the study period, Viet Nam was supposed to be featured with labour surplus. However, capital was limited. Growth in sizes often requires higher capita per

laborer as better technologies often need higher capita intensities. However, if firms cannot meet the requirement, efficiency of growth in sizes would be reduced. The challenge was figured out by Tybout (2000) for development of firms in developing countries.

The growth in sizes of firms is often accompanied by upgrades in technologies. Newly entered firms with given technologies would try to maximize their potentials. However, each technology has its own productivity frontier. Firms would not be able to further increase their productivities going beyond the frontiers of their technologies without replacements with better ones *Jovanovic and Nyarko, 1996)*. Given this phenomenon, when small firms reach potentially maximal productivities of their technologies, they would be faced two choices, keeping their sizes and productivities or upgrading their technologies.

However, technology upgrading is a risky process. Investment or technology upgrading does not secure improvements in productivity or returns to scale. When upgrading their technologies, firms face a cost of interruption *(Perez and Ponce, 2015)* with a temporary reduction in productivity *(Jovanovic and Nyarko, 1996)*. In technology upgrading period, competitors may increase their market shares. In addition, firms also face riskes of failing in mastering new technologies or gains from new technologies are lower than the costs. These risks all result in fails in technology upgrading.

It is a common expectation that RTS of the middle-sized firms should not be lower than that of small-sized ones if technology upgrading is in a consequential order and there are no risks of upgrading. This expectation is stemmed from two reasons: (1) better technologies would be more efficient; (2) it is a self-selection of firms to upgrade their technologies with an objective of a higher productivity. Meanwhile, the higher efficiency implies the higher RTS. However, the empirical results exhibit that Vietnamese middle-sized firms have the lowest RTS or their efficiencies are even lower than those of the small ones. Put differently, firms in Viet Nam fails to improve their efficiencies when grow into the middle-sized ones. The fails include technology replacement and upgrading process.

In fact, studies on technologies in general and technology upgrading in particular are scare as standard enterprise surveys do not often cover a technology section *(Verhoogen, 2020)*. Therefore, empirical evidence of risks of technology upgrading is also very scale. Results of Fernandes and Paunov (2015) for firms in Chile[7] in the period of 1996-2003 exhibited that a product diversification process would reduce probabilities of exiting. However, if firms only produce unique products, a product upgrading would result in higher probabilities of exiting. Biesebroeck (2005) indicated that technology

[7] Indeed, Chile is a high-income country and a member of OECD.

upgrading would not result in higher firms' growth in 9 Sahara economies in the period of 1992-1996.

In a different aspect, a study of World Bank (2017) exhibited that Vietnamese small and medium firms have often improved their management systems or optimized their given technologies rather than upgraded their technologies or products, compared with firms in other ASEAN's countries. This result implies that gains from technology upgrading are not enough to compensate risks or costs of the process in a majority of Vietnamese small and medium firms.

Therefore, the middle-sized firms in Viet Nam face a number of challenges, not only in term of growth into large ones but also in term of keeping efficiencies as small ones. This result implies that policies making or evaluation processes should take into account heterogeneities of firms with different sizes. Attentions should be paid to investment in general and replacement and upgrading of technologies in particular with the risks as discussed above. Taking into account heterogeneities of firms with different sizes would not only improve probabilities of success but also increase efficiencies of the policies. This feature is important in the policy making processes given limitation of resources in Viet Nam.

5. Conclusion

Firm size distribution has been of particular concern of economists for nearly a century. Given the macro environment, entrepreneurs choose to allocate the resources and emerge to the certain firm size to maximize their profit. FSD is indeed the endogenous choice of firms and specific FSD can give a possible interpretation on the differences underlying firm dynamics. FSD can also contribute to explain the industrial concentration of the economy, which in turn represents the level of competition in the market. In this study, we explore the existence of the "missing middle" phenomenon in Vietnam and provide initial studies on the correlation between this abnormal FSD and the existence of economies of scale in Vietnam's industries. We find evidence supports the existence of the "missing middle" in FSD of Vietnam as well as the IRTS in aggregate production function of most industries in Vietnam.

Such co-existence suggests that there exist forces other than those traditionally mentioned in economic literature at work. As mentioned by Tybout (2000), non-normal FSD can be resulted from in capital constraint, human resource limitation, the high corruption level, the different regulatory structure or the lack of transport infrastructure and the limited liberalization to international trade of the developing countries. In other words, micro firms are exposed to different sets of business environmental constraints compared to those of medium and large enterprises.

Results of RTS estimations for firms with different sizes within each industry exhibits that middle sized firms have the lowest RTS, compared with those of large or small ones. The lowest RTS implies of challenges when Viet Nam's firms develop into middle sized ones which include changes in operational environments as mentioned above.

Furthermore, we discuss two reasons of investment and risks of technology upgrading which can results in diminishing in RTS when firms grow into the middle size. Insufficient investment and fails in technology upgrading both cause decreasing in RTS. These are important reasons which should be investigated and directly tested in Viet Nam's context. In the meantime, the obstables implies that it should be taken into account the specific features of firm-sized groups. The attention should be paid to challenges in investment in general and investment in technology upgrading in particular with its specific risks of the interruption cost, temporary decreases in productivity, market shares and risks of fails in mastering new technologies.

On the other hand, increasing returns in the aggregate production function may be due to overhead (fixed) costs, diminishing marginal cost, positive spillovers from aggregate activity, the entry of new varieties of inputs or changes in the distribution of inputs across heterogeneous firms. Each channel has significant implications for models of growth, trade and business cycles.

Future studies can be focused on the causes of these phenomenons, including insufficient investment and fails in technology upgrading when growing into the middle size. Results will have important implications to policies toward creating a favorable macro environment for business development across different firm sizes./.

REFERENCES

1. Baldwin, J. and Gorecki, P. (1986). The role of scale in Canada-US productivity differences in the manufacturing sector: 1970-1979. University of Toronto Press.
2. Biesebroeck, J. V. (2005). Firm Size Matters: Growth and Productivity Growth in African Manufacturing. Economic Development and Cultural Change, 53(3): 545-583.
3. Brown, D. J. (1991). "Equilibrium analysis with non-convex technologies". *Handbook of Mathematical Economics*, Vol. 4.
4. Christensen, L. R. and Greene, W. H. (1976). "Economies of scale in U.S. electric power generation". *The Journal of Political Economy*, 84:655-676.
5. Christensen, L. R., Jorgenson, D., and Lau, L. (1973). "Transcendental logarithmic production frontier". *The Review of Economics and Statistics*, pages 28-45.
6. Farrell, M. J. (1957). "The measurement of productive efficiency". *Journal of the Royal Statistical Society*, 3:253-290.

7. Fernandes, A. M. and Paunov, C. (2015). "The risks of innovation: are innovating firms less likely to die?" *The Review of Economics and Statistics*, 97(3): 638-653.
8. Ijiri, Y., Simon, and Hebert (1977). Skew Distribution and the Size of Business Firms. North-Holland, Amsterdam, Minnesota.
9. Jacques, M. and Jordi, J. (2005). "Panel-data estimates of the production function and the revenue function: What differences does it make?" *Scandinavian Journal of Economics*, 107.
10. Jovanovic, B. and Yaw Nyarko, Y. (1996). "Learning by Doing and the Choice of Technology". *Econometrica*, Vol. 64 (6), pp. 1299-1310.
11. Karsten, J. (2005). Economies of scale: A survey of the empirical literature. Contemporary issues in urban and regional economics. New York: Nova Science Publ.
12. Klette, T. J. and Griliches, Z. (1996). "The inconsistency of common scale estimators when output prices are unobserved and endogenous". *Journal of Economic Behavior and Organization*, 11:343-346.
13. Levinsohn, J. and Petrin, A. (2003). "Estimating production functions using inputs to control for un observables". *Review of Economic Studies*, 70:317-342.
14. Lucas, R. E. (1978). "On the size and distribution of business firm". *Bell Journal of Economics*, 9:508-523.
15. Luttmer, E. G. J. (2007). "Selection, growth and the size distribution of firms". *Quarterly Journal of Economics*, 122:1103-1068.
16. Mansfield, E. (1962). "Entry, Gibrat's law, innovation, and the growth of firms". *American Economic Review*, 52:1023-1051.
17. Perez, C. and Ponce, C. J. (2015). "Disruption Costs, Learning by Doing and Technology Adoption". *International Journal of Industrial Organization*, Vol. 41, pp. 64-75
18. Ramey, V. A. (1991). "Non-convex costs and the behavior of inventories". *The Journal of Political Economy*, 99:45-61.
19. Rumelt, R. P. and Wensley, J. R. C. (1981). Stochastic and direct effect theories of the association between market share and profitability: an empirical discrimination. Working paper.
20. Tybout, J. R. (2000). "Manufacturing firms in developing countries: How well do they and why?". *Journal of Economic Literature*, 38:11-44.
21. Verhoogen, E. (2020). "Firm-Level Upgrading in Developing Countries". *CDEP-CGEG Working paper* No. 83. <https://cdep.sipa.columbia.edu/cdep-cgeg-working-paper-no-83>
22. Vinning, D. R. (1976). "Auto-correlated growth rates and the Pareto law: A further analysis". *Journal of Political Economy*, 84:369-380.
23. Westbrook, M. D. and Tybout, J. R. (1992). "Estimating returns to scale with large, imperfect panels: an application to Chilean manufacturing industries". *The World Bank Economic Review*, 7:85-112.
24. World Bank Group (2017). Vietnam: Enhancing Enterprise Competitiveness and SME Linkages. World Bank, Washington, DC. <https://openknowledge.worldbank.org/bitstream/handle/10986/28488/119861-WP-EnhancingEnterpriseCompetitivenessandSMELinkagesPPostD-PUBLIC.pdf?sequence=1&isAllowed=y >.