Misallocation, productivity and firm dynamics in Vietnam

Long Thai

Thesis for the Degree of Doctor of Philosophy in Economics

49,470 words

School of Economics
University of Kent
Canterbury
CT2 7FS
United Kingdom

August 2020

In loving memory of my Dad...

Acknowledgement

Foremost, nothing could express my great gratitude dedicated to Professor Miguel León-Ledesma who offered me his mentorship and devoted care. His consistent patience, guidance and overall insights in this field have made this an inspiring and valuable experience for me. He constantly encourages me and was always willing to assist in any way he could throughout the project. I wholeheartedly thank him and will be truly indebted to him throughout my life.

I gratefully acknowledge the School of Economics, University of Kent, United Kingdom, the 911 Project of Vietnam's Ministry of Education and Training, the Foreign Trade University (FTU, Ha Noi) and the Faculty of International Economics (FTU) for having provided me with the best working environments, for their financial and administrative support.

I am extremely grateful to my mother, who means everything to me, for her selfless love, patient support emotionally and financially, for always being behind me when I have trouble. I sincerely thank my mother-in-law for her unreserved trust, encouragement and endless patience. I wish to thank my brothers, sisters-in-law for their silent support when I am away from home.

Last but not least, I would like to express my deepest gratitude to my dearest wife, Minh-Hong Phi, my great companion who always encourages me to move forward, for her unfailing love and understanding, her constant support and help throughout our challenging period.

Abstract

This PhD thesis is composed of three chapters on macroeconomics to investigate the aggregate productivity dynamics in Vietnam using plant-level data and focusing on the role of inefficient allocation of resources in an emerging economy.

In the first chapter, entitled *"Misallocation in the Vietnamese manufacturing sector"*, we focus on the static misallocation and distortions with an attempt to study the question of resource misallocation in Vietnam during the period 2000-2013 by adopting the framework of Hsieh and Klenow (2009) and the decomposition method of Chen and Irarrazabal (2015) with minor modification. The study is based on data from the annual Enterprises Survey collected by the General Statistics Office of Vietnam. We find significant misallocation of resources in Vietnamese manufacturing. Our results show that the indicator of allocative efficiency, the variation of revenue productivity, obviously surpasses that of the US economy. In the case that all output and capital distortions are eliminated, potential TFP gains could be equal to 89.6–106.7%. In addition, if the Vietnamese manufacturing sector could move to the same level of misallocation as the US, the aggregate TFP could increase by 32.68–44.65%. Besides, about 88% of the TFP growth can be attributed to the changes in efficient TFP and allocational efficiency accounts for 12%.Further, there is an improvement in allocational efficiency, judged by the fact that TFPR distributions became less dispersed. Also, state-owned and collective plants have lower TFPR than private and foreign ones, bigger plants have higher TFPR and older plants have lower TFPR; collective and state-owned plants tend to face output distortions, private plants tend to face more capital distortions; plants of all sizes seem to face the prevailing output distortions and plants of all ages face capital distortions. When studying the plant size distribution, we find that the actual plant size distribution is less dispersed than efficiently sized ones and the efficient plant size distribution becomes less dispersed over time. Compared to efficient size distribution, many plants overproduced in this period and should be reduced to reach their optimum size. Looking into the key forces of misallocation, the decomposition similar to Chen and Irarrazabal (2015) shows that the total allocation gain is primarily led by TFPR variation. In a modified decomposition and an investigation of TFP gain with the absence of each type of distortion, we find that the TFPR variation is driven mostly by output distortions, with capital distortions playing a smaller role. On the other hand, the TFPR variation is determined mainly by the between-group component of different groups of TFPQ. That between-group TPFR variation happens mostly in the lowest and the highest productivity plants.

The second chapter titled *"Productivity dynamics in Vietnam: an application of dynamic Olley-Pakes productivity decomposition"* analyses the sources of aggregate productivity growth and firm dynamics. Using a similar dataset as in the previous chapter, we divide the whole sample into three groups of plants, that of survivous, entrants and exiters. Applying the Melitz and Polanec (2015)'s decomposition, we find that the increase in aggregate productivity relative to that in 2000

is mostly due to the growing contribution of surviving plants in which the improvement in productivity within plants plays a large part. Comparing the productivity of exiters and entrants to that of survivors in the corresponding period, we find a positive contribution of exiters and a negative contribution of entering plants. The positive contribution of surviving plants and the positive contribution of exiting plants are consistent when we use two alternative decomposition methods, such as those of Foster, Haltiwanger, and Krizan (2001) and Griliches and Regev (1995). Looking deeply into the plant dynamics, we explore the plant's productivity life cycle and find that new plants can learn from market functioning. Precisely, we observe that the group of plants surviving throughout the period of 2000-2013 has the highest productivity and new entrants could not catch up these plants. However, when we relax the definition of incumbents at time t as plants being at both time $t-1$ and t, we find that new entrants can catch up the incumbents 3 to 5 years after entering the market. Exiting plants are less productive than the incumbents and the gap in productivity is more pronounced when they are closer to their exit year. Further, we apply the method of Hashiguchi et al. (2015) to analyse the contribution of surviving plants and plant dynamics in some specific groups of plants. We divide our sample according to three alternative criteria: their two-digit industry, their ownership status and their export status. We find that the between-plant component could improve when the industries increase their foreign owners' value-added share. We also record a better market share reallocation for exporting plants during 2010-2013 and for private ownership plants from 2006.

The third chapter titled *"Markups and Trade liberalisation: the case of Vietnamese manufacturing sector"* focuses on the role of markups variation as a source of misallocation. Following the methodology of De Loecker and Warzynski (2012) in estimating plant markup, we firstly investigate the markups characteristics and find that the aggregate markup in Vietnamese manufacturing decreases gradually overtime. That suggests a higher competition between plants and makes plants change their prices to better reflect their costs, therefore benefiting the consumers. Moreover, we observe a lower markup dispersion in 2013 in comparison to 2000, implying a better market share reallocation between manufacturing plants. To better understand the source of the reducing aggregate markup, we decompose it at industry-level into the change of markup within industry, the change in the composition between industries and the joint change in markup and the industry composition. We find that the markup change within the industry contributes mostly to the fall in the aggregate markup. We then decompose the aggregate markup at plant-level using the insight of Melitz and Polanec (2015). Dividing plants into three groups (survivors, entrants and exiters), we find that the decline in markups within plants drives the aggregate markup change. Looking at the relationship between markups and export status, we find that the higher markup for exporting firms is the results of not only the selection process but also the learning-by-doing process. Indeed, plants having prior success (larger size, higher productivity, wagebill and markups) are likely to export. After entering export markets, they continue to charge higher markups. To analyse the reason behind the decrease in markup dispersion, we look at the role of trade reform: the accession into the WTO in early 2007. Following Lu and Yu (2015) in using the difference-in-differences methodology, we find evidence that trade liberalisation reduces markups dispersion measured as the Thiel-index. This effect is more pronounced in the two-digit industries with relatively more state-owned plants.

Contents

List of Tables

List of Figures

Chapter 1

Misallocation in the Vietnamese manufacturing sector

1.1 Introduction

One of the most attractive current macroeconomic issues is explaining the disparity of level and growth of output per capita for different countries. There is support for asserting that the productivity gap plays the key role in explaining the difference in standard of life among different countries (Klenow and Rodriguez-Clare, 1997). For a deeper understanding, many contributions have focused on finding the reasons why poorer countries face lower Total Factor Productivity (TFP) than richer ones. The TFP gap across countries may be explained by how these countries manage to allocate various resources across heterogeneous production establishments (Hsieh and Klenow, 2009; Restuccia and Rogerson, 2013).

Restuccia and Rogerson (2013) divide the study of misallocation problems into two main approaches. Studies following the direct approach measure the extent to which some factors distort the allocation and affect aggregate TFP. These factors could be policies (Hopenhayn and Rogerson, 1993; Lagos, 2006) or credit market imperfections (Banerjee and Duflo, 2005; Midrigan and Xu, n.d.). The indirect approach instead isolates the underlying sources of the wedges, focusing on the conditions under which misallocation affects TFP (Restuccia and Rogerson, 2008) and measuring the variation of wedges (Bartelsman, Haltiwanger, and Scarpetta, 2013; Hsieh and Klenow, 2009).

Among the papers using the indirect approach, Hsieh and Klenow (2009),using production unit-level data, provide a comprehensive method for measuring the possible effect of resource misallocation on aggregate TFP. Differentiating between physical and revenue productivity, their important finding is that revenue productivity should be equal across heterogenous firms if there is no distortion. Therefore, the dispersion in revenue-based TFP (TFPR) implies evidence of distortion, which can account for the TFPR pattern of countries. Moreover, the Hsieh–Klenow (HK) method allows the measurement of how much aggregate productivity could improve if capital and labour marginal products are equal across firms within an industry. Therefore, one advantage of this method is that the contribution of resource misallocation to aggregate productivity can be compared across countries. In light of the advantages of this methodology, we aim to use it to quantify the misallocation of resources in Vietnam.

Our contribution is twofold. First, we attempt to identify the key sources of misallocation and TFP gains by using decomposition methods which are still limited in application to less developed countries. Second, we compute misallocation not only in the case of no distortion but also when there is only one type of distortion (capital distortion or output distortion) in order to compare the impact of both types.

Our research has some key findings. First, we find considerable resource misallocation in Vietnamese manufacturing. In the case of no distortion, potential TFP gains could be equal to 89.6-106.7% or around 97.2% for the period. There is also an allocational efficiency improvement by 6% for the whole period or 0.43% on average per year. Besides, about 88% of the TFP growth can be attributed to the changes in efficient TFP, and allocational efficiency accounts for 12%. Second, there is an improvement in allocational efficiency, judged by the fact that TFPR distribution becomes less dispersed. Third, the actual plant size distribution is less dispersed than efficiently sized ones, and the efficient plant size distribution becomes less dispersed over time. Compared to the efficient size distribution, many plants overproduced in this period and should be reduced to reach their best size. Fourth, the total allocation gain is primarily led by TFPR variation, which is driven mostly by output distortions, with capital distortions playing a smaller role.

Our research is related to several empirical studies that are based on the HK methodology, especially for developing countries. This strand of literature confirms that for developing countries, the TFP would have increased significantly if resources had been reallocated to either the case of no distortions or "US efficiency". Machicado and Birbuet (2012) use this methodology to analyse the case of Bolivia and find that if their resources are hypothetically reallocated to US efficiency, the gains of manufacturing TFP range from -6% to 38%. Camacho and Conover (2010) explore that reallocation of resources by using the US as a benchmark boost manufacturing TFP by about 3%-8%. Oberfield (2013) investigate resource misallocation across plants in Chile over the period 1979–1996, including the crisis in 1982. To determine the role of allocational efficiency, Oberfield develops a measure of allocational efficiency derived from the framework of Hsieh and Klenow (2009) and finds that roughly one-third of the loss in TFP can be accounted by the allocational efficiency changes in between-industry. Dheera-Aumpon (2014) finds that the misallocation across firms in Thailand is higher than in China, India and the US. If resources are reallocated as in the US, manufacturing TFP increases by 74%.

Based on the HK model, some studies develop the decomposition method to clarify the sources of TFP gains and misallocation. Chen and Irarrazabal (2015) decompose the TFP gains in different components and find that, for Chile, the dispersion in TFPR within a sector contributes mostly to its TFP gains. In the case of Ukraine, Ryzhenkov (2016) decomposes the TFP gains into the variance of capital and output wedges. In the Ukrainian economy, the impact on the overall misallocation from both output and capital distortions are roughly equivalent in terms of scale. Some studies focus on the decomposition of the misallocation, measured as the dispersion of TFPR. Decomposing this variable into that within quintile and that between quintiles of firms according to physical productivity, Chen and Irarrazabal (2015) point out the major contribution to the variance of log TFPR between quintiles. In recent work, Haltiwanger, Kulick, and Syverson (2018) show that, under the specific assumptions of HK model (isoelastic demand, constant marginal costs), the variance of TFPR comes from distortions. When these assumptions are relaxed, the variance of TFPR could be decomposed into the contribution of fundamentals, the contribution of distortions and the contribution of terms involving both fundamentals and distortions.

Our research is related to some papers analysing the misallocation of resources in the case of Vietnam. Bach (2019) uses micro-level data of Vietnam during the period 2000-2008 to quantify the gains of TFP in the case of absence of distortions in the manufacturing sector, finding the sector's growth could increase by 226% in 2000, 181% in 2007, and 142% in 2008. Ha, Kiyota, and Yamanouchi (2016) extend the research of Bach (2019) and present some interesting results: (i) TFP of Vietnam would be boosted by 30.7% if Vietnam is hypothetically as efficient as the US in 1997, (ii) the distortion is related to firm size: distortions are more advantageous in small firms than large ones, (iii) "the efficient distribution of firm size is more dispersed than the actual size distribution". Our differences from the previous studies focusing on Vietnam are threefold. First, we compute the misallocation not only for the case of no distortion but also when there is only one type of distortion (capital distortion or output distortion). We also try to differentiate the impact of physical productivity and allocational efficiency on aggregate TFP. Second, we attempt to identify the sources of misallocation and TFP gain by using decomposition methods. Third, we investigate some characteristics related to the distribution of two measures of productivity, namely physical and revenue productivity, that have been only limitedly investigated in previous studies on Vietnam.

The rest of this paper is organised as follows. Section 1.2 states the theoretical framework for the misallocation measurement and the decomposition method. Section 1.3 describes the data and several parameter choices. Section 1.4 presents the main findings. Finally, Section 1.5 provides the conclusion.

1.2 Methodology

1.2.1 Measurement of misallocation

To evaluate misallocation, we use the theoretical model of Hsieh and Klenow (2009) (HK) which is mainly a monopolistic competition framework.

At the aggregate level, in a perfectly competitive market, a representative firm produces a final good using Cobb–Douglas production technology:

$$Y = \prod_{s=1}^{S} Y_s^{\theta_s} \tag{1.1}$$

where $\sum_{s=1}^{S} \theta_s = 1$, Y_s is the output of industry s, $s = 1, \ldots, S$ and θ_s is the output share of industry s.

M_s differentiated products are used to produce Y_s with CES technology:

$$Y_s = \left(\sum_{i=1}^{M_s} Y_{si}^{1-\frac{1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tag{1.2}$$

where firm i, $i = 1, \ldots, M_{si}$ produces Y_{si}, and the elasticity of substitution between firm output is σ.

Firm i in industry s produces Y_{si} using a Cobb-Douglas technology, which consists of total factor productivity A, labour L and capital K:

$$Y_{si} = A_{si} K_{si}^{\alpha_s} L_{si}^{1-\alpha_s} \tag{1.3}$$

where α_s are capital income shares which are equal for all firms in an individual industry but different across industries.

Firm's maximisation problem

Hsieh and Klenow (2009) assume that each firm has an output idiosyncratic tax rate, $\tau_{Y_{si}}$, and a capital idiosyncratic tax rate,$\tau_{K_{si}}$, which are called the "wedges". Hence, each firm maximises its profit by:

$$\begin{aligned} &\max_{P_{si},K_{si},L_{si}} (1-\tau_{Ysi}) P_{si} \underbrace{A_{si} K_{si}^{\alpha_s} L_{si}^{1-\alpha_s}}_{Y_{si}} - \omega L_{si} - (1+\tau_{Ksi}) R K_{si} \\ &st: Y_{si} = Y_s \left[\frac{P_s}{P_{si}}\right]^{\sigma} \end{aligned} \tag{1.4}$$

where P_{si}, ω and R denote firm's price, the common wage and rental costs, respectively. The parameters $\tau_{K_{si}}$ create distortions which increase the capital marginal product relative to labour and $\tau_{Y_{si}}$ create distortions which raise the capital and labour marginal products by the same proportion.

For the first-order condition, the firm's price and output are given by:[1]

$$P_{si} = \frac{\sigma}{\sigma-1} \left(\frac{R}{\alpha_s}\right)^{\alpha} \left(\frac{\omega}{1-\alpha_s}\right)^{(1-\alpha_s)} \frac{(1+\tau_{Ksi})^{\alpha_s}}{(1-\tau_{Ysi}) A_{si}} \tag{1.5}$$

$$Y_{si} = \frac{A_{si}^{\sigma} (1-\tau_{ysi})^{\sigma}}{(1+\tau_{ksi})^{\alpha_s \sigma}} \left(\frac{\sigma-1}{\sigma}\right)^{\sigma} \left(\frac{\alpha_s}{R}\right)^{\alpha_s \sigma} \left(\frac{1-\alpha_s}{\omega}\right)^{\sigma(1-\alpha_s)} P_s^{\sigma} Y_s \tag{1.6}$$

The marginal revenue product of capital and labour can be represented by:

$$MRPK_{si} = \alpha_s \frac{\sigma-1}{\sigma} \frac{P_{si} Y_{si}}{K_{si}} = \frac{(1+\tau_{Ksi}) R}{(1-\tau_{Ysi})} \tag{1.7}$$

$$MRPL_{si} = (1-\alpha_s) \frac{\sigma-1}{\sigma} \frac{P_{si} Y_{si}}{L_{si}} = \frac{\omega}{(1-\tau_{Ysi})} \tag{1.8}$$

These marginal values are different from the first best choice of firms in the presence of distortions. For example, if a firm receives a subsidy from the government, the $\tau_{Y_{si}}$ is negative, then the $MRPL$ is lower than firms with no subsidy. This incident creates a suboptimal allocation of resources.

Physical and revenue productivity

The HK model uses two measures of productivity: "revenue productivity" and "physical productivity". A firm's revenue productivity can be used to measure firm-specific distortions.

[1] All firms are assumed to have the same wage.

$$TFPQ_{si} \triangleq A_{si} = \frac{Y_{si}}{K_{si}^{\alpha} L_{si}^{1-\alpha}} \tag{1.9}$$

$$TFPR_{si} \triangleq P_{si} A_{si} = \frac{P_{si} Y_{si}}{K_{si}^{\alpha_s} L_{si}^{1-\alpha_s}} \tag{1.10}$$

In the case of distortions, those distortions translate into new marginal productivities, which again modify the TFPR, which is now proportionate to a geometric mean of the firm's marginal revenue products of capital and labour:

$$TFPR_{si} = \frac{\sigma}{\sigma - 1} \left(\frac{R}{\alpha_s}\right)^{\alpha_s} \left(\frac{\omega}{1 - \alpha_s}\right)^{(1-\alpha_s)} \frac{(1 + \tau_{ksi})^{\alpha_s}}{(1 - \tau_{ysi})} = \frac{\sigma}{\sigma - 1} \left(\frac{MRPK_{si}}{\alpha_s}\right)^{\alpha_s} \left(\frac{MRPL_{si}}{1 - \alpha_s}\right)^{1-\alpha_s} \tag{1.11}$$

Two points should be noted from the above equation. First, a firm's revenue productivity can be used to measure firm-specific distortions since TFPR are different across firms within each industry if there are output and capital distortions. If distortions are eliminated, labour and capital are reallocated from low TFPQ firms to high TFPQ firms which results in equalising TFPR within industry and therefore create a potential TFP gain. Second, if the distortions are large then the dispersion of TFPR might be large.

Aggregate TFPs and misallocation measurement

We need two aggregate TFP: industry and whole sector TFP. Using the aggregate labour and capital for industry $L_s = \sum_{i=1}^{M_s} L_{si}$ and $K_s = \sum_{i=1}^{M_s} K_{si}$, we have the TFP at industry level:

$$TFP_s \triangleq \frac{Y_s}{K_s^{\alpha_s} L_s^{1-\alpha_s}} = \left[\sum_{i=1}^{M_s} \left(A_{si} \frac{\overline{TFPR_s}}{TFPR_{is}}\right)^{\sigma-1}\right]^{\frac{1}{\sigma-1}} \tag{1.12}$$

where

$$\begin{aligned}\overline{TFPR_s} &\triangleq \frac{\sigma - 1}{\sigma} \left(\frac{R}{\alpha_s \sum_{i=1}^{M_s} \frac{1-\tau_{Ysi}}{1+\tau_{Ksi}} \frac{P_{si}Y_{si}}{P_s Y_s}}\right)^{\alpha_s} \left(\frac{\omega}{(1-\alpha_s) \sum_{i=1}^{M_s} (1-\tau_{Ysi}) \frac{P_{si}Y_{si}}{P_s Y_s}}\right)^{1-\alpha_s} \\ &= \frac{\sigma - 1}{\sigma} \left(\frac{\overline{MRPK_s}}{\alpha_s}\right)_s^{\alpha} \left(\frac{\overline{MRPL_s}}{1 - \alpha_s}\right)^{1-\alpha_s}\end{aligned} \tag{1.13}$$

Hsieh and Klenow (2009) prove that if the dispersion in TFPR is high, then industry TFP will be low. Besides, if there are no distortions, then $\tau_{Y_{si}} = \tau_{K_{si}} = 0$ and $TFPR_{si} = \overline{TFPR}_s$. The efficient productivity of the aggregate industry is:

$$TFP_s^E = \bar{A}_s = \left(\sum_{i=1}^{M_s} A_{si}^{\sigma-1}\right)^{\frac{1}{\sigma-1}} \tag{1.14}$$

At the whole sector level, the actual and efficient productivities are:

$$TFP = \prod_{s=1}^{S} TFP_s^{\theta_s} = \prod_{s=1}^{S} \left[\sum_{i=1}^{M_s} \left(A_{si} \frac{\overline{TFPR_s}}{TFPR_{si}} \right)^{\sigma-1} \right]^{\frac{\theta_s}{\sigma-1}} \tag{1.15}$$

and

$$TFP^E = \prod_{s=1}^{S} \left(TFP_s^E \right)^{\theta_s} = \prod_{s=1}^{S} \left(\bar{A}_s \right)^{\theta_s} \tag{1.16}$$

Then, the relationship between the actual and efficient productivities is:

$$\frac{TFP}{TFP^E} = \prod_{s=1}^{S} \left[\sum_{i=1}^{M_s} \left(\frac{A_{si}}{\bar{A}_s} \frac{\overline{TFPR_s}}{TFPR_{si}} \right)^{\sigma-1} \right]^{\frac{\theta_s}{\sigma-1}} \tag{1.17}$$

Therefore, the potential TFP gain from reallocation is:

$$Gain = \left(\frac{TFP^E}{TFP} - 1 \right) \times 100\% \tag{1.18}$$

1.2.2 Decomposition analysis

In this part, we provide the method used to identify the key forces that form the resource misallocation. In this analysis, we decide to use the technique suggested by Chen and Irarrazabal (2015) by decomposing the gain and, hence, single out the key parts in misallocation.

This decomposition is mainly based on the assumption by Chen and Irarrazabal (2015); Hsieh and Klenow (2009) of joint log-normal distribution of $1 - \tau_{Y_{si}}$, $1 + \tau_{K_{si}}$ and $TFPQ_{si}$. This assumption also entails a joint log-normal distribution of TFPR and TFPQ.

Applying the Central Limit Theorem, the aggregate TFP can be decomposed as:

$$logTFP = logTFP^E - \frac{\sigma}{2} var(logTFPR_{si}) - \frac{\alpha(1-\alpha)}{2} var\left[log\left(1 + \tau_{K_{si}}\right)\right] \tag{1.19}$$

From this equation, we can see that two factors result in changes in aggregate TFP: the first comes from the changes in the distribution of physical productivity (or the efficient TFP), and the second, the changes in the misallocation. This is also a difference between Chen and Irarrazabal's method and Hsieh and Klenow's model: the former tries to incorporate changes in both physical productivity and allocational efficiency while the latter focus only on allocational changes.

Therefore, the potential gains from the first-best allocation compared to the current allocation in the economy can be decomposed as follows:

$$logTFP^E - logTFP = \frac{\sigma}{2} var(logTFPR_{si}) + \frac{\alpha(1-\alpha)}{2} var\left[log\left(1 + \tau_{K_{si}}\right)\right] \tag{1.20}$$

The left-hand side represents the gains and the right-hand side (RHS) is a sum of two parts, or in our analysis of two forces: $var(logTFPR_{si})$ or the variation of TFPR, which represents distortions across firms in the industry; $var\left[log\left(1+\tau_{K_{si}}\right)\right]$ the variation of capital distortions, which represents the capital-labour ratio distortions.

We can further decompose the TFPR variation into three other components:

$$\begin{aligned} var(logTFPR_{si}) =& var\left[log\left(1-\tau_{Y_{si}}\right)\right] + \alpha_s^2 var\left[log\left(1+\tau_{K_{si}}\right)\right] \\ & - 2\alpha_s cov\left[log(1-\tau_{Y_{si}}), log(1+\tau_{K_{si}})\right] \end{aligned} \tag{1.21}$$

where the first term depicts the impact of output distortion on misallocation, the second term denotes the effect of capital distortion while the third captures the covariance between those two distortions.

Chen and Irarrazabal (2015) also take another direction to decompose TFPR variance into within-group and between-group components (with groups represented by quintiles of TFPQ distribution). This decomposition analyses the TFPR variation among groups of plants by five productivity levels. Specifically, the following formula provides this decomposition:

$$\begin{aligned} var\left(logTFPR_{si}\right) &= \frac{1}{M_s}\sum_q^Q \sum_i^{N_q}\left(logTFPR_{sqi} - \overline{TFPR}_s\right)^2 \\ &= \frac{1}{M_s}\sum_q^Q N_q var\left(logTFPR_{si}\right)_q + \frac{1}{M_s}\sum_q^Q N_q\left(\overline{TFPR}_{sq} - \overline{TFPR}_s\right)^2 \end{aligned} \tag{1.22}$$

where the within-group component is the first term in the second line, and between-group component is the second one, measured as the deviation of quintile q mean log TFPR from the mean log TFPR of industry s. In Equation (1.22), N_q is the number of plants in quintile q, $logTFPR_{sqi}$ is log TFPR of plant i, that lay in the $q-th$ quintile of industry s, $\overline{TFPR}_s$ is the industry s' average log TFPR, $\overline{TFPR}_{sq}$ is the average logTFPR of $q-th$ quintile of industry s,

1.3 Data and parameters

1.3.1 Data

We use plant-level data for the manufacturing sector from the Vietnamese Enterprise Survey collected annually for 2000-2013, provided by the General Statistics Office of Vietnam (VGSO). The dataset includes all registered enterprises, including all state-owned enterprises (SOEs) and foreign-owned firms, at the end of each year without any firm size threshold, plus a representative random sample of small registered firms with less than 10 employees. This is a comprehensive dataset for the period 2000–2013, which provides plant data such as the number of employees, capital, depreciation, pre-tax profit, wage bill, sales, etc. Since these variables are in nominal values, we deflate them using the Producer Price Index in manufacturing provided by the VGSO. In this study, we study the manufacturing industries defined under the International Standard Classification of Industry revision 3.1 (ISIC Rev.3.1) with a classification up to 2-digit industry. Because we want to use the data at plant-level, we combine firm identifiers provided with plant code and city to seperate plants.

However, as the dataset does not provide plant-level value added, we compute this variable as the sum of the wage bill, depreciation and pre-tax profit.[2] Since the HK method is based on the assumption that all plants have the same wage, we first consider in our main analysis a constant wage bill of unity. Then, we re-estimate the TFP gains with employment as the wage bill as in Ha and Kiyota (2014) for a robustness check.

1.3.2 Parameter calibration

In order to apply the methodology of Hsieh and Klenow (2009), we have to calibrate the model parameters. The parameters necessary for such a calculation consist of the rental price of capital, the elasticity of output with respect to capital and the elasticity of substitution between plants' value-added.

Hsieh and Klenow (2009) chose the value of 3 for the elasticity of substitution between plants value-added σ. The authors based this on the literature in this area, which gives an estimation of this parameter from 3 to 10, and argue that the gain from no distortions in their model is increasing with respect to the elasticity of substitution; therefore, they chose the conservative value of 3 as the baseline. This particular value is widely used in many studies derived from their framework. We follow this practice and for robustness checks, we set σ to be equal to 2 as well as 5.

In Hsieh and Klenow (2009) and several studies using this method, with a depreciation rate of 5% and real interest rate of 5% the rental price of capital R is determined at the value of 10%. We follow this "default" value of $R = 0.1$ which is used in many studies for other countries. However, for robustness and to make the data more relevant for Vietnam, we study Vietnamese legislation and related statistics to calibrate the rental price of capital. An analysis of Vietnamese Tax code gives us a depreciation rate at 10%. The real interest rate in Vietnam from 2000 to 2013 varied between -6.55% in 2005 to 6.91% in 2000,[3] with a median of 2.25% and mean of 1.53%. So we chose to use a real interest rate value of 2%. Therefore, the rental price of capital is set to be 12%, or $R = 0.12$.

The elasticity of output with respect to capital α_s is equal to 1 minus the share of labour, defined as labour compensation in the manufacturing sector's value added as in Hsieh and Klenow (2009). Since the US economy is widely considered to be less distorted than less developed countries such as Vietnam, we use the US labour shares as the benchmark case. US labour shares are determined using data from NBER Productivity Database, which specifies value added aggregated by industry as well as labour compensation following NAICS 1997 classification. Since our main dataset covers the period from 2000 to 2013, the US labour shares for each industry are calculated as the ratio of total labour compensation from 2000 to total value added from 2000. Following Hsieh and Klenow (2009), since the compensation can include non-wage forms, we multiply the calculated US labour shares by 3/2 to account for this observation. The Vietnamese data requires mapping the NAICS to VSIC classification, we then use the NAICS - ISIC Rev 3.1 - VSIC correspondence table.

[2]Unfortunately, we do not have data on interest paid on credit and loans, which is part of the payments to capital. Therefore, as in Ha and Kiyota (2014), we use this relaxed definition.

[3]Data collected from the World Bank's World Development Indicators database.

The plant's capital and output distortion can be determined from the elasticity of output to capital, the elasticity of substitution and the plant data as follows:

$$1 + \tau_{K_{si}} = \frac{\alpha_s}{1 - \alpha_s} \frac{\omega L_{si}}{R K_{si}} \tag{1.23}$$

$$1 - \tau_{Y_{si}} = \frac{\sigma}{1 - \sigma} \frac{\omega L_{si}}{(1 - \alpha_s) P_{si} Y_{si}} \tag{1.24}$$

Last, the physical productivity is computed as follows:

$$A_{si} = \frac{Y_{si}}{K_{si}^{\alpha_s} L_{si}^{1-\alpha_s}} = \lambda_s \frac{(P_{si} Y_{si})^{\frac{\sigma}{\sigma-1}}}{K_{si}^{\alpha_s} L_{si}^{1-\alpha_s}} \tag{1.25}$$

where $\lambda_s = (P_s Y_s)^{-\frac{1}{\sigma-1}} / P_s$. As in HK, we set $\lambda_s = 1$ as the reallocation gains do not depend on a concrete value of λ_s even we do not observe it.

1.4 Results

In this part of the chapter, we first report the potential gain of aggregate productivity growth. The next step involves the progression of different productivity distribution measures and the plant-size one and some further investigations related to different plant characteristics. We subsequently, perform the decompositions of the total gain (or total distortions), then within different levels of productivity, study the changes in allocational efficiency among plants. Finally, we perform several robustness checks of the results with different scenarios.

1.4.1 Potential gains

We consider potential gains for Vietnamese manufacturing in the case where resource misallocation is eliminated. The objective of this investigation is to estimate how much the economy would gain if there is no misallocation, meaning that there is no distortion among plants. The potential gains are calculated according to Equation (1.18). Table 1.1 shows that potential gains in the case of no distortions are around the level of 89.6-106.7%. First, it should be noted that the magnitude of these TFP gains does not have a clear trend. In general, for the whole period 2000-2013, by removing any distortion in each industry, the gain in 2000 is 106.7%; it slightly reduces to 94.87% in 2013. Therefore, throughout this period, the HK method gives us a gain of 97.2% on average. Furthermore, we also have an estimation of allocative efficiency improvement of 6% (206.7/194.87), or an improvement of 0.43% on average per year.

However, taking a closer look into the evolution of the TFP gain over the period, we can see a decrease in gain from 2000 to 2003, an increase from 2004 to 2011, then followed by another decrease afterwards. This fluctuation suggests that the misallocation did not always improve throughout the studied time. We also note that from 2007 to 2011, the potential gain significantly worsened, this might be the effect of the world crisis and difficulties in the Vietnamese economy which will be discussed later in the TFPR investigation.

Table 1.1: TFP gains from removing distortions within industries

Year	2000	2001	2002	2003	2004	2005	2006
Gains with no distortions	106.7	100.8	94.76	89.60	89.77	90.94	92.85
Gains relative to USA	44.65	40.52	36.29	32.68	32.8	33.62	34.95

Year	2007	2008	2009	2010	2011	2012	2013
Gains with no distortions	92.64	95.47	101.1	103.0	104.3	104.2	94.87
Gains relative to USA	34.81	36.79	40.73	42.06	42.97	42.90	36.37

Note: Entries are $100(\frac{TFP^E}{TFP} - 1)$ where $\frac{TFP}{TFP^E}$ is calculated according to Equation (1.18).

Second, even though the misallocation did not much improve, we can still evaluate the contribution of allocational efficiency to the manufacturing sector's TFP growth. We follow the analysis of Chen and Irarrazabal (2015) by using a panel regression of the log difference in TFP with an independent variable of the log difference in the allocative efficiency, $\frac{TFP_{st}}{TFP_{st}^E}$, with a constant and year dummy variable included:

$$\Delta logTFP_{st} = \alpha + \beta \Delta log\left(\frac{TFP_{st}}{TFP_{st}^E}\right) + \gamma_t + \varepsilon_{st}$$

The coefficient $\beta = 0.119$ is statistically significant at 10 percent. This result suggests that around 12% of aggregate TFP growth for the period 2000-2013 might be associated with better resource allocation. As a result, about 88% of the TFP growth can be attributed to the changes in efficient TFP, which can be represented by changes in TFPQ distribution.

Third, when applying the distribution of the US economy (using the value of TFP gains of 42.9% in 1997, which is also the largest potential gain in the US for the period studied by Hsieh and Klenow), the gains are only 32.68-44.65% for the whole period or around 38% on average.[4] These gains are slightly smaller than those of China (39.3% for the period 1998-2005) and India (46.9% for the period 1987-1994) reported by Hsieh and Klenow.

In sum, in the case of no distortions, the TFP gain for the manufacturing sector in Vietnam is fairly high, at around 97% (or around 38% if misallocation level is similar to the US in 1997).

1.4.2 Productivity and productivity dispersion

In the process of calculating the TFP gains, we have to estimate the physical productivity TFPQ and revenue productivity TFPR for each plant. Therefore, we want to investigate some characteristics of these productivity measurements and how they are related to the aggregate TFP in the sector.

[4] The gain relative to USA (in 1997) for Vietnam in 2000 is absolute gain of Vietnam in 2000 divided by the gain of the US in 1997 and equal to 206.7/142.9.

Table 1.2: Summary statistics for the distribution of productivity

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
$logTFPQ$														
S.D.	1.485	1.443	1.407	1.404	1.374	1.368	1.353	1.350	1.348	1.397	1.324	1.416	1.425	1.416
90-10	3.930	3.738	3.678	3.670	3.592	3.557	3.486	3.502	3.494	3.628	3.546	3.764	3.836	3.778
75-25	2.122	2.054	1.911	1.980	1.894	1.868	1.870	1.827	1.715	1.902	1.782	1.968	1.860	1.864
$logTFPR$														
S.D.	0.855	0.859	0.850	0.868	0.861	0.855	0.836	0.835	0.802	0.820	0.792	0.890	0.877	0.850
90-10	2.216	2.195	2.209	2.247	2.208	2.226	2.136	2.166	2.035	2.113	2.042	2.322	2.233	2.204
75-25	1.144	1.153	1.132	1.168	1.119	1.139	1.101	1.103	1.065	1.083	1.044	1.237	1.168	1.116

Note: With plant i in industry s, $TFPQ_{si} = \frac{Y_{si}}{K_{si}^{\alpha} L_{si}^{1-\alpha}}$, $TFPR_{si} = \frac{P_{si} Y_{si}}{K_{si}^{\alpha_s} L_{si}^{1-\alpha_s}}$. Industries are weighted by value-added shares. Statistics are for deviation of $log(TFPQ)$ and $log(TFPR)$ from industry mean. S.D.: standard deviation, 75−25: difference between the 75th and 25th percentiles, 90−10: difference between the 90th and 10th percentiles.

TFPQ distribution

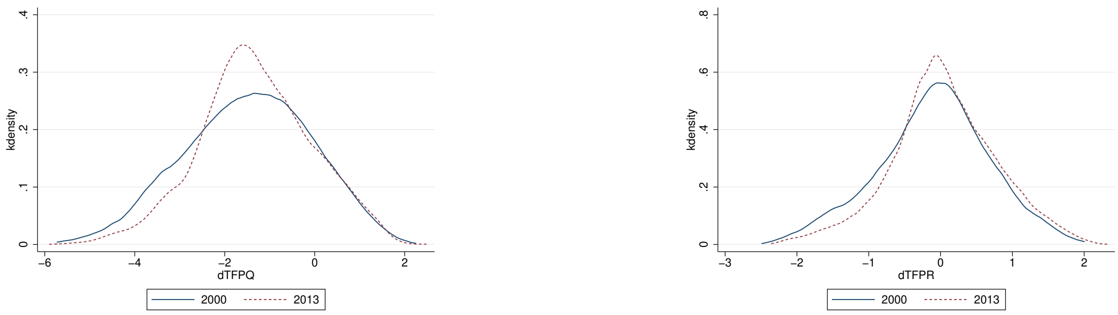
The TFPQ in the case of no distortions can vary across plants and industries and have no impact on allocational efficiency. However, we still want to investigate this productivity distribution since it affects the aggregate TFP (according to Equation (1.19)), as well as the efficient size distribution which will be discussed later in the section of actual and efficient plant size distribution).

Panel (a) of Figure 1.1 displays the distribution of TFPQ in two selected years of the period period (the beginning and the end of the period) for a quick comparison. The distribution of TFPQ is computed as $log\left(\frac{A_{si} M_s^{\frac{1}{\sigma-1}}}{\bar{A}_s}\right)$. It should be noted that the measure of TFPQ proposed by HK indicates not only plant physical productivity but also the quality and variety of its products. The weights used in the distribution's calculation are the shares of each industry value added in the total manufacturing sector.

First, the distribution of TFPQ in 2000 shows a slightly fatter left tail than the right, suggesting that economic and plant policies during the late years of the 1990's might have facilitated the continuation of lower efficient plants. Furthermore, even though the right tail of the distribution of TFPQ in two critical years (2000, 2013) does not change much, the left tail becomes thinner from 2000 to 2013. This change suggests that inefficient plants can improve themselves by increasing their physical productivity quicker than the industry average or they just simply go bankrupt and exit the market. A reason that could explain the exiting of inefficient plants is the economic and business environment reform in the 2000s. Indeed, only since 2000, have the Vietnamese plants had a better business environment because of the new Enterprise Law in 2000 that encouraged plants of various ownership status to participate in the market, thereby enhancing competition.

Second, over time, the TFPQ distribution gradually became less dispersed, as suggested by the graph of the 2013 distribution (see Figure 1.B.1 in Appendix 1.B). However, this improvement does not always hold for the entire period. Table 1.2 gives a further view of plant TFPQ according to different percentiles. In this table, several dispersion measures are reported including the standard deviation, the difference of the 90th to the 10th percentile and the difference of the 75th to the 25th percentile.

We can see that the TFPQ standard deviation dropped from 1.485 to 1.416 between 2000 and 2013; the difference of the 75th to the 25th percentile of TFPQ fell from 2.122 to 1.864; the



(a) TFPQ distribution

(b) TFPR distribution

Figure 1.1: Productivity distribution

difference of the 90th to the 10th percentiles decreased from 3.93 to 3.778. However, they always have some years of increase.

Therefore, the most important point in this analysis is that there are changes in the left tail and the dispersion of the TFPQ distribution (as the role of efficient TFP), which have an impact on aggregate TFP alongside any possible effect of allocational efficiency in the later investigation. Besides, these changes also affect the size distribution.

For further investigation of the TFPQ distribution, we choose to regress the TFPQ against several covariates (Table 1.3).

First, we try to compared the TFPQ according to types of plant ownership: state-owned which consists of principal and local-state owned plants and equitised plants with more than half of the shares held by the state, collective plants which are jointly owned by the local governments and private investors, foreign-owned which are foreign owned plants and joint ventures, private-owned plants which are domestic and privately-owned, and equitised plants with more than half of the shares held by the private shareholders.

Compared to the reference state-owned plants, the dummy coefficients are positive and significant for foreign-owned and private ones, negative and significant for the collective ones. Therefore, the most productive plants are private ones, followed by the foreign-owned, state-owned plants and the least productive are collective plants.

Second, plant TFPQ also seems to be positively correlated with plant size (proxied by the number of employees) and the regression results suggest that bigger plants tend to have higher productivity. The same observation can be made about the relationship between productivity and exporter (represented by export dummy which equals 1 in any year the plant exports) and plant's age.

To sum up, over time, TFPQ distribution is less dispersed and has a less fat left tail, meaning there is an improvement in productivity and TFPQ distribution also has an impact on aggregate TFP. Besides, this productivity seems to increase with plant age, size and export status.

TFPR distribution

According to Hsieh and Klenow's method, the TFPQ in the case of no distortion can vary across plants and industries; however, the TFPR are the same for all plants within an industry or plants manage to equalise their marginal product revenues. Therefore, the dispersion of this

Table 1.3: WLS of TFPQ on ownership status, age dummies, export status and size dummies

VARIABLES	Dependent: Deviation of log TFPQ from industry mean				
	(1)	(2)	(3)	(4)	(5)
Collective ownership	−2.084***				−1.009***
	(0.018)				(0.016)
Private ownership	1.185***				0.282***
	(0.015)				(0.014)
Foreign ownership	0.139***				0.100***
	(0.017)				(0.016)
2nd quartile of age		0.437***			0.226***
		(0.008)			(0.007)
3rd quartile of age		0.583***			0.263***
		(0.009)			(0.008)
4th quartile of age		0.570***			0.240***
		(0.009)			(0.008)
Export status				1.065***	0.276***
				(0.009)	(0.008)
2nd quartile of labour			0.371***		0.320***
			(0.007)		(0.007)
3rd quartile of labour			0.993***		0.825***
			(0.007)		(0.007)
4th quartile of labour			1.981***		1.626***
			(0.008)		(0.009)
Constant	−0.988***	−2.354***	−2.927***	−2.273***	−2.770***
	(0.024)	(0.022)	(0.019)	(0.021)	(0.022)
Observations	154,679	154,679	154,679	154,679	154,679
R-squared	0.165	0.070	0.341	0.112	0.382
Year dummies	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes

Note: The dependent variable is the deviation of log TFPQ from the industry mean. The independent variables are dummies for (i) collective-owned plants, private-owned plants, foreign-owned plants with state-owned plants as reference group, export status dummy with non-exporters as omitted group, different quartiles of age and labour with the 1st quartile as omitted group. Regressions are weighted least squares and weights are industry value-added shares. Results are pooled for all years. *** p<0.01, ** p<0.05, * p<0.1.

measure will recognise the misallocation in the economy.

First, panel (b) of Figure 1.1 shows that the distribution of TFPR, calculated as $log\left(\frac{TFPR_{si}}{TFPR_s}\right)$, is slightly more dispersed in 2000 than 2013, suggesting a small improvement in allocative efficiency since 2000 in terms of TFPR's standard deviation (SD). Moreover, in 2013, the left tail of TFPR's distribution becomes slightly thinner which might show that some less-efficient plants can catch up with the industry's average in terms of revenue productivity.

Second, the results in Table 1.2 also reveal larger distortions in Vietnam than in the United States. The TFPR standard deviation in 2013 is 0.85, much greater than the number of 0.45 of the United States in 1997. For the whole period from 2000 to 2013, the standard deviation of TFPR is around 0.85 for Vietnam, which is higher than China (0.68) and India (0.68), however it is equal to Thailand, a neighbouring developing country (0.85 in 2006) (Dheera-Aumpon, 2014), and lower than Ukraine, a transition economy (1.12 for 2002-2010) (Ryzhenkov, 2016).

A closer look at the dispersion of TFPR shows a tendency for it to reduce from 2000 until 2008, fluctuate, then start to rise until 2011 and reduce again to end up at 0.85, less than 0.855 in 2000. Thus indicating slightly better resource allocation: more plants are more productive than

Table 1.4: Standard deviation for the distribution of TFPR, by ownership

Ownership status	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
							TFPR							
State	0.793	0.768	0.765	0.780	0.864	0.886	0.869	0.856	0.803	0.795	0.890	0.900	0.878	0.887
Collective	0.819	0.832	0.800	0.840	0.822	0.844	0.793	0.818	0.771	0.745	0.636	0.697	0.697	0.663
Private	0.852	0.868	0.860	0.873	0.861	0.848	0.829	0.832	0.747	0.786	0.767	0.880	0.854	0.837
Foreign	0.926	0.885	0.897	0.913	0.867	0.869	0.863	0.829	0.832	0.843	0.833	0.830	0.843	0.830

Note: Statistics are for deviation of $log(TFPR)$ from industry means. Industries are weighted by value-added shares.

the average and fewer establishments are subject to more distortions. However, there is a period, 2008-2011, where the allocation seems to worsen. The reason might be the world crisis effect and turbulence in the Vietnamese economy during this period. At this time, GDP growth rates decreased from around 7% before 2008 to around 5.5% after this year, while inflation rate rose to reach 23.1% in 2008 and 18.7% in 2011, making borrowing from commercial banks, the main source of financing for plants in Vietnam, vastly expensive.

Third, we try to compare the development of TFPR standard deviation according to types of plant ownership in Table 1.4. For the period 2000-2013, three types of plants can reduce their TFPR dispersion: the collective, private and foreign owned and the most successful are the collective plants. The private plants can only slightly narrow down their TFPR standard deviation. It is likely that during this period of time, there is a boost in the number of private enterprises entering the market as a result of several economic policies that facilitate their investment and operations; therefore the heterogeneity in productivity, size in these type of plants make it difficult to reduce their distortions and misallocation. Besides, the TFPR dispersion of state-owned plants increases even though these enterprises undergo a gradual process of equitisation over this time. It may be the case that these plants still receive advantageous conditions from the government through protection in some industries and easy credit access (Mishra, 2011).

As in the case of TFPQ, we regress TFPR on several similar covariates. This gives insight on distortions from two points of view.

The first point is based on the observation that plants with low TFPR may have beneficial distortions while ones with high TFPR may face non-beneficial distortions.

In Column 1 of Table 1.5, we compare TFPR across plants according to their ownership status (with state-owned plants as the group of reference), export status (with non-exporters as the group of reference), quantiles of age and size (with the 1st quantile as the group of reference). The results show that state-owned and collective plants have lower TFPR than private and foreign ones. This observation suggests these plants face possibly lower distortion or they could be subsidised by the economic authorities.

Regression on quartiles of size (Column 3) shows that bigger plants are likely to have higher TFPR or face more unfavourable distortions. This result is similar to what Ha et al. (2016) find when the authors plot the kernel density of TFPR against the firms' size.

However, age is negatively related to TFPR (Column 2), which indicates that older plants may face less unfavourable distortions than younger ones.

The second point of view takes into consideration an insight from Equation (1.11). According to this formula, TFPR is inversely proportionate to the output distortions but proportionate to the capital distortions. As a result, in Figure 1.1 which plots the distribution of $log\left(\frac{TFPR_{si}}{TFPR_s}\right)$,

Table 1.5: WLS of TFPR on ownership status, age dummies, export status and size dummies

VARIABLES	Dependent: Deviation of log TFPR from industry mean				
	(1)	(2)	(3)	(4)	(5)
Collective ownership	−0.475***				−0.336***
	(0.011)				(0.012)
Private ownership	0.098***				0.016
	(0.010)				(0.010)
Foreign ownership	0.046***				0.047***
	(0.011)				(0.011)
2nd quartile of age		−0.144***			−0.117***
		(0.005)			(0.005)
3rd quartile of age		−0.160***			−0.122***
		(0.006)			(0.006)
4th quartile of age		−0.071***			−0.080***
		(0.005)			(0.006)
Export status				0.222***	0.102***
				(0.006)	(0.006)
2nd quartile of labour			0.057***		0.041***
			(0.005)		(0.005)
3rd quartile of labour			0.186***		0.130***
			(0.005)		(0.005)
4th quartile of labour			0.287***		0.191***
			(0.006)		(0.006)
Constant	−0.055***	0.342***	−0.405***	−0.321***	−0.045***
	(0.005)	(0.013)	(0.013)	(0.013)	(0.002)
Observations	154,679	154,679	154,679	154,679	154,679
R-squared	0.048	0.033	0.046	0.036	0.066
Year dummies	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes

Note: The dependent variable is the deviation of log TFPR from the industry mean. The independent variables are dummies for (i) collective-owned plants, private-owned plants, foreign-owned plants with state-owned plants as reference group, export status dummy with non-exporters as omitted group, different quartiles of age and labour with the 1st quartile as omitted group. Regressions are weighted least squares and weights are industry value-added shares. Results are pooled for all years. *** p<0.01, ** p<0.05, * p<0.1.

plants laying to the right of the mean TFPR tend to be more affected by capital distortions, while plants laying to the left are more affected by output distortions. For the studied period, the dispersion of this distribution reduces, suggesting a reduction in the impact of distortions or fewer plants subject to distortions.

First, Column 1 of Table 1.5 suggests that collective and state-owned plants tend to face output distortions, judged by the negative constant and collective plants' coefficients (the sum of constant and collective plants coefficient is also negative), possibly due to subsidies from the government. Whereas private plants tend to face more capital distortions (positive sum of constant and private plants coefficient), possibly due to difficulties in credit access, while foreign plants may have a more balanced amount of both types of distortions.

Second, related to plant size (Column 3), all plants seem to face the prevailing output distortions, indicated by a negative constant coefficient associated with the lowest quartile of size and a negative sum of constant with other plant size coefficients. However, the impact of output distortions decreases with size.

Age distribution (Column 2) show that all plants seem to face capital distortions (a positive constant coefficient associated with the lowest quartile of age and a positive sum of constant with other plant age coefficients) and the two-middle quartiles of age might face less capital distortions than the bottom and top quartile.

In brief, first, TFPR distribution became less dispersed over the period, meaning there is an improvement in allocative efficiency. However, the TFPR distribution's dispersion fluctuates and seems to increase for the period 2008-2011 during the world crisis and difficult economic conditions in Vietnam. Second, different types of plants according to ownership had different results in terms of reducing TFPR dispersion where the most successful are collective plants while the least are state-owned plants. Third, state-owned and collective plants have lower TFPR than the private and foreign ones, bigger plants tend to have higher TFPR and older plants have lower TFPR. Fourth, collective and state-owned plants tend to face output distortions, private plants tend to face more capital distortions, but plants of all sizes seem to face the prevailing output distortions and plants of all ages face capital distortions.

Actual size *vs.* efficient size

The HK model implies that change in TFPQ or physical productivity and change in misallocation can result in adjustments in the plant size distribution (in our investigation, plant size is the plant's value added).

We can represent plant size with:

$$P_{si}Y_{si} = Y_{si}^{1-\frac{1}{\sigma}} P_s Y_s^{\frac{1}{\sigma}} \tag{1.26}$$

From the above equation and Equation (1.6), we can get:

$$P_{si}Y_{si} \propto \left[\frac{A_{si}(1-\tau_{Y_{si}})}{(1+\tau_{K_{si}})^{\alpha_s}} \right]^{\sigma-1} \tag{1.27}$$

The above formula suggests that changes in TFPQ distribution and in distortions might lead to changes in the efficient plant size distribution as well as the difference between efficient and actual plant size.

If there are no distortions of any types, plants with higher productivity are usually larger and vice versa. If there is a positive correlation between A_{si} and $1+\tau_{K_{si}}$ or there is a negative correlation between A_{si} and $1-\tau_{Y_{si}}$ then plants with higher productivity are usually smaller than the efficient size. In other words, in the presence of distortions, the dispersion of actual size distribution is less than the dispersion of efficient plant size.

Figure 1.2 shows the graph of the actual and efficient plant size distribution in 2000 and 2013. Similar to the progression of the TFPQ distribution, the efficient plant size distribution in 2013 is less dispersed than in 2000 and by 2013 it has a slimmer left tail, that implies an enhancement in the efficient TFP.

For both years, the efficient plant size distribution seem to be more dispersed than their actual size one, therefore suggesting that most of the plants suffer from overproduction and hence their size should be reduced to reach their optimum.

Furthermore, the actual size distribution differs from the efficient size one and this gap rests

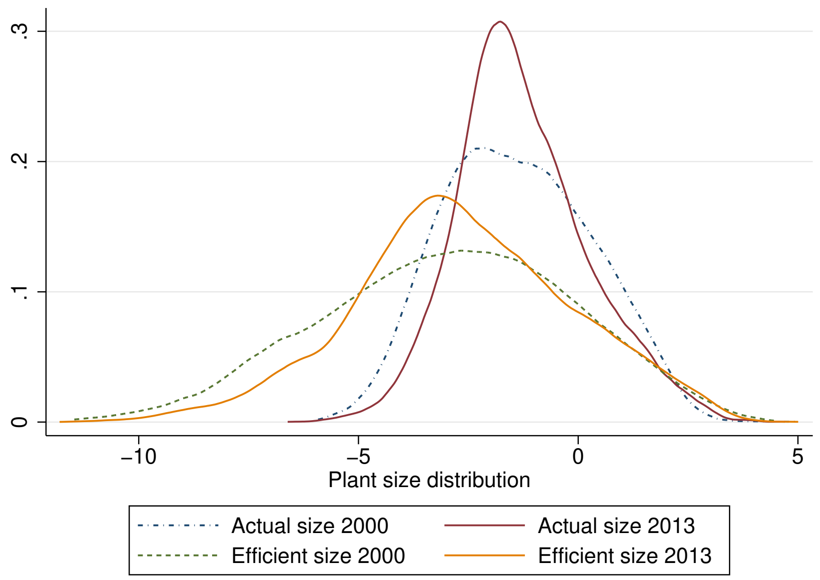


Figure 1.2: Plant size distribution

largely on the left tail. This observation indicates that the policies might implicitly favour small plants and lead to their overproduction compared to other plants that are not implicitly subsidised.

Table 1.6 displays the distribution of Vietnamese manufacturing plants by size and the deviation from the efficient size for the two years 2000 and 2013. There are four bins in this table which are represented by the rows. The proportions of efficient size to the actual size are listed as follow: the bin 0-50 where efficient size is less than half of the actual one or the plant size should be smaller by more than one-half; 50-100: actual size should be smaller by less than one-half; 100-200: actual size should be larger by less than twice; $>$ 200: actual size should be larger by more than twice.

From the table, we can find two things. First, most of the plants should be downsized more than twice in the case of no distortion. This is implied by the number of most plants which lies in the first column across any size quartile for both 2000 and 2013. In this light, most of the plants might implicitly receive subsidies and therefore they produce more than the output they should produce in the case there is no distortion. Besides, due to the misallocation of resources in the economy, many plants cannot attain their potential development.

Second, throughout the period, there is evidence that less efficient plants are downscaled, while more efficient plants increase their size with more production. In 2013, the segment of small plants that should be downscaled by at least 50% decreases to 18.86% from 21.15% in 2000. This consolidates what we observe from Figure 1.2 that the actual plant-size distribution in 2000 is farther from its efficient distribution than in 2013, notably when we observe the left tail.

In brief, the actual plant size distribution is less dispersed than the efficient size one; compared to the efficient size distribution, many plants overproduced in this period and should be reduced to reach their best size and the efficient plant size distribution becomes less dispersed over time.

Table 1.6: Actual size vs. efficient size

2000	0-50	50-100	100-200	200+
1st quartile	7.43	5.37	4.83	7.36
2nd quartile	10.35	5.10	4.11	5.43
3rd quartile	13.70	5.18	3.30	2.82
Bottom quartile	21.15	2.18	1.04	0.64
Share of total	52.63	17.84	13.28	16.25

2013	0-50	50-100	100-200	200+
1st quartile	7.13	4.95	5.19	7.73
2nd quartile	10.28	6.27	3.62	4.83
3rd quartile	12.91	6.09	3.08	2.93
Bottom quartile	18.86	3.08	1.64	1.43
Share of total	49.18	20.38	13.54	16.92

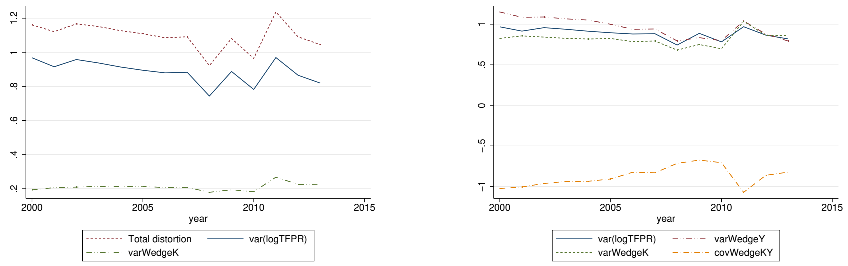
1.4.3 Decomposition of resource misallocation

A question that naturally follows is if there is a change in allocative efficiency, then what is the contribution of different components. Therefore, in order to catch the main forces which affect the change and to know to what extent that this change can be explained by variation in different components, we rely on the assumption of a joint log normal distribution between distortions and physical productivity and follow the decomposition method suggested by Chen and Irarrazabal (2015) according to Equation (1.20).

In Panel (a) of Figure 1.3, we graph the evolution in total gain (or total distortions) from eliminating idiosyncratic distortions and its various components. We can observe that the total distortions can be more strongly tracked by the dispersion in TFPR than the capital distortions. Compared to TFPR variation, capital distortion has a much smaller and more stable effect. In other words, the TFPR variation is the main driving force of misallocation in the case of Vietnam.

For that reason, we look deeper into the TFPR variation by further decomposing the variance of TFPR using Equation (1.21) and the results are depicted in Figure 1.3 Panel (b) which shows the evolution in var(logTFPR) and its various components. The TFPR variation comes mainly from the output distortions, however, the effect of capital distortions is not far behind. While TFPR variation declines slowly from 2000 until 2008, by contrast, the capital distortion witnesses a small change in this period. However, both seem to fluctuate after 2008 with higher variation still belonging to the output measure.

Further, the contribution of covariance between capital and output distortion is positive for the whole period, which suggests that high output distortion frequently come with high capital distortion. The magnitude of covariance effect decreases during the analysed period, which may indicate that policies resulting in capital and output distortions develop more independently over time.



(a) Total distortion and components

(b) var(logTFPR) and components

Note: (a) Total misallocation is decomposed following Equation (1.20). (b) Variance of TFPR is decomposed following Equation (1.21).

Figure 1.3: Decomposition of resource misallocation

Which type of distortions has deeper impact on misallocation?

From the decomposition method by Chen and Irarrazabal (2015), we find that the main source of misallocation comes from TFPR variation while capital distortions have a smaller effect. This decomposition does not help us to compare the impact between output and capital distortions. Bearing in mind the question to know which distortions have the higher impact, we modify this method to decompose the total gain into the variation in each of those two distortions and their interaction.

In order to compare the effect of each type of distortions, we then modify the above approach to directly decompose the total distortions (or the total gain) in to three components, which consist of capital distortions, output distortions and the covariance between those two. The following formula can be obtained when we substitute Equation (1.21) into Equation (1.20):

$$\begin{aligned} logTFP^E - logTFP =& \frac{\sigma}{2} var[log(1-\tau_{Ysi})] + \frac{\sigma\alpha_s^2 - \alpha_s^2 + \alpha_s}{2} var[log(1+\tau_{Ksi})] \\ & - \sigma\alpha_s cov[log(1-\tau_{Y_{si}}), log(1+\tau_{Ksi})] \end{aligned} \tag{1.28}$$

Results from the above method are presented in Figure 1.4, where the output capital distortions might have a slightly bigger impact on total gain than the output one for the period 2000-2004. However, this pattern changes from 2005 onward, with a higher variation of output variance. The covariance term can be interpreted in the same manner as in the last section.

We later found out that this approach was already used in Ryzhenkov (2016). Furthermore, the above result may not get rid of the interaction between the two distortions due to the presence of the covariance between them. Therefore, we also try another approach by setting each of the distortion ”wedge” equal to 0 then recomputing the gains. These gains will be compared to each other and give us a picture of the impact of each type of distortions on the TFP gain in the absence of the other.

Figure 1.5 suggests that the gain from eliminating capital distortion ($\tau_{K_{si}} = 0$) or only output distortions in presence is clearly higher than the other one. The pattern of gain from this exercise seems to track closely the evolution of total gain: a decrease, a long steady increase and then a final decrease. In contrast, the gain from only capital distortions (eliminating the output distortions) is

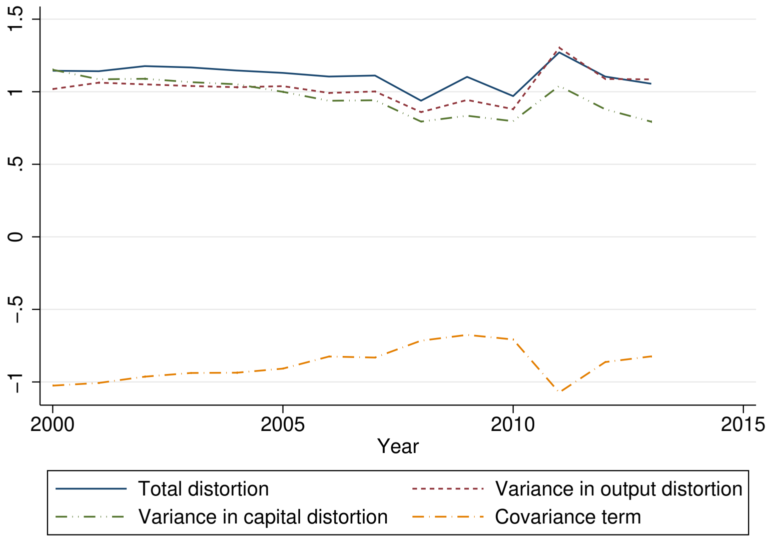


Figure 1.4: TFP gains decomposition by distortions

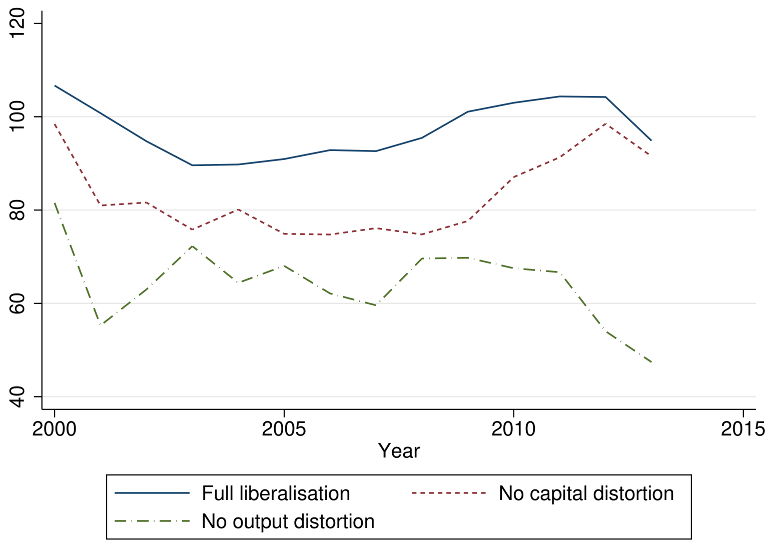
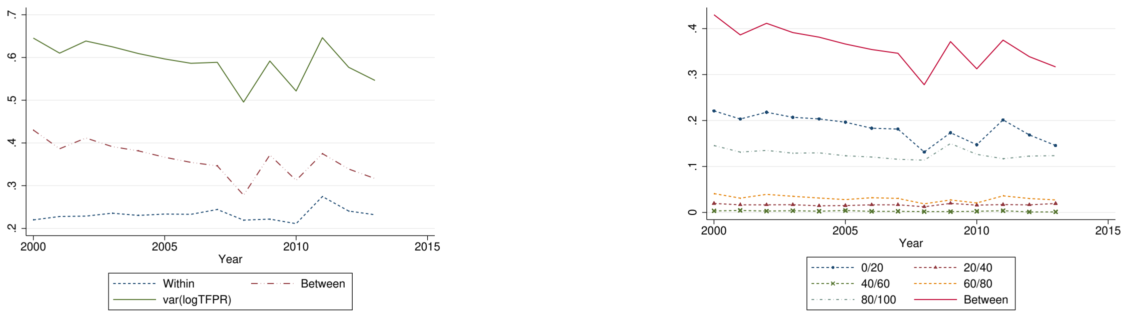


Figure 1.5: Potential TFP gains by type of plant-specific distortion

lower and more fluctuating, with an overall decreasing trend through the whole period.

Hence, the output distortion seems to have a higher impact and this impact might increase, while the capital distortion is lower and decreasing impact. That might result from economic policies that eased the capital-related issues in plants such as banking reforms beginning in the 1990s, the establishment of the Vietnam Stock Exchange in 2000, etc.

To sum up, total gain from allocational efficiency can be explained mainly by the variation in revenue productivity (TFPR) and in small part by the capital distortions in the sense of Chen and Irarrazabal (2015). More decomposition by modifying the above method or setting each of the distortion ”wedges” equal to zero reveal that output distortions seems to have deeper impact on misallocation than the capital distortions.



(a) Components of var(logTFPR)

(b) Quintiles of var(logTFPR)

Figure 1.6: Decomposition of TFPR

Misallocation across plants of different productivity levels

From the investigation of gain and the dispersion of TFPR and TPFQ, we find an improvement in allocational efficiency that comes mainly from variation in TFPR. We want to know where this improvement happens across plants and different levels of productivity among plants in the sample. For that reason, we rely on another decomposition by Chen and Irarrazabal (2015) which is represented by Equation (1.22) to break down the change in resource allocation across plants according to several degrees of productivity. For each year, we put plants into 5 bins (quintiles) according to their physical productivity (TFPQ), then decompose the TFPR variation (the log TFPR variance) into within and between-group components.

Chen and Irarrazabal (2015) argue that in their method, the "between" component, which catches the dispersion of TFPR between different groups of productivity level, by its own nature, can remove the idiosyncratic factors impacting the TFPR variance and therefore can be used to evaluate the degree of misallocation among different quintiles of TFPQ. At the same time, the within-group component is still affected by other idiosyncratic factors and is used to catch the misallocation within each quintile.

In Panel (a) of Figure 1.6, we plot the results of this decomposition which show that a significant source of TFPR variance comes from the between-group component. For the whole period, the between-group part was considered to be responsible for 69.3% of the decrease in TFPR variance.[5]

Related to the within-group component, this part also has a large impact, however the magnitude is mostly around half of the between one.

Panel (b) of Figure 1.6 gives us a deeper understanding of the main productivity bins which are accountable for misallocation with an additional decomposition of the between-group component. While the lowest quintile contributes 66.4% of the between-group component, the highest

[5]The contribution to the total TFPR variance of Between-group component is calculated as follow:

$$\frac{\Delta \frac{1}{N} \sum_q^Q N_q \left(\overline{TFPR}_q - \overline{TFPR}\right)^2}{\Delta var(logTFPR)}$$

where Δ is the change between the last and the first year.

quintile is responsible for 22.5%.[6] Therefore, the estimated decomposition implies that reallocation happen mostly in the lowest and the highest productivity plants. Similar to the result obtained by Chen and Irarrazabal (2015) for Chile and Ryzhenkov (2016) for Ukraine, this decomposition shows that the convergence to the mean by average TFPR from the lowest and highest quintiles constitutes the main cause for the change in the between-group variance of TFPR.

1.4.4 Robustness check

The gain and misallocation might depend on measurement error in the data and the chosen parameter. Therefore, we try to check the robustness of our main results with several estimations.

First, to check the sensitivity of the result to alternative values of parameters such as the elasticity of substitution σ or the capital share α, we set $\sigma = 2$ and 5 which corresponds to the markup of 2 (=2/(2-1)) and 1.25 (=5/(5-1)) and also vary the technology parameter or the capital share with different values: 1/3, the industry average capital share calculated from the data and plant-specific share.

Results from Table 1.7 suggest that the gain and standard deviation of TFPR are sensitive to changes in these parameters to some extent. For example, with $\sigma = 2$ the gain is 70.3%, with σ=5 the gain is 143.4%, compared to the benchmark case of 97.2%. The standard deviation of TFPR with $\sigma = 2$ is 0.86, with $\sigma = 5$ is 0.89, which are fairly similar to 0.85 of the benchmark case. Hsieh and Klenow (2009) argue this characteristic by slow reallocation among plants with diverse productivity, so in the case of higher of elasticity of substitution, the gain is higher.

Second, in the main results, unity wage are used. In a robustness check, we replace the employment or the number of workers in each plant by wage bills to account for the difference in hours worked and worker quality. Robustness check 6 show that both the gain and the standard deviation of TFPR are smaller than the benchmark case.

Third, with *R*=0.12 instead of 0.1, the result are considerably close with gains are 97 and 97.2% and standard deviation of TFPR are both 0.85.

The important observation we get from these results with several types of robustness check is that their gains share a similar pattern as the benchmark case, although with some small differences: a decrease in gain from 2000 to 2003, an increase from 2004 to 2011, then another decrease afterwards.

Lastly, there is a concern about the contribution of entering and exiting plants to the gain and the variation of TFPR. Therefore, we should control this effect and examine the role of the extensive margin and the intensive margin by using a balanced panel. In this check, we find a reduction in the gains since only the most productive plants can last through the whole period with a gain of 64.3% and standard deviation of TFPR is 0.76, compared to 97.2% and 0.85 from the benchmark. There are two points that should be noted: (i) these results take into account TFP gains only from the most well-performing plants which survive throughout the period 2000-2013, thus underestimating the gain, (ii) compared to the benchmark with a gain in allocative efficiency of 6.1% (=206.7/194.87)

[6] Each quintile's contribution to between-group components is:

$$\frac{\Delta \frac{N_q}{N}(\overline{TFPR}_q - \overline{TFPR})^2}{\Delta between - groupcomponent}$$

for the whole period, this gain from the intensive margin is 23.2% (=189.6/153.8), or the intensive margin dominates the extensive one in this area.

To sum up, several robustness checks in Table 1.7 suggest that the TFP gains and the standard deviation of TFPR in different scenarios are fairly similar to the benchmark case. Therefore, we can conclude that the misallocation in the Vietnamese manufacturing sector for the period 2000-2013 is large, judged by the gains of TFP and SD of TFRP compared to other countries.

1.5 Conclusion

In this paper, in order to investigate the effect of resource misallocation on manufacturing productivity in Vietnam for the period 2000-2013, we adopt the Hsieh and Klenow (2009) model and the decomposition method by Chen and Irarrazabal (2015). The study is based on the data from the annual Enterprises Survey collected by the General Statistics Office of Vietnam. The study has some key findings.

First, we find significant resource misallocation in Vietnamese manufacturing. Our results show that the indicator of allocative efficiency, which is the variation of revenue productivity, obviously surpasses the US economy. In the case of no distortion or all output and capital distortions removed, potential TFP gains could be equal to 89.6-106.7%. Furthermore, if the Vietnamese manufacturing sector can move to the same level of misallocation as the US, the aggregate TFP could increase by 32.68-44.65%. Besides this, about 88% of the TFP growth can be attributed to the changes in efficient TFP and allocational efficiency account for 12%.

Second, there is an improvement in allocational efficiency judged by the fact that TFPR distribution became less dispersed. In addition, state-owned and collective plants have lower TFPR than the private and foreign ones, bigger plants have higher TFPR and older plants have lower TFPR, collective and state-owned plants tend to face output distortions, private plants tend to face more capital distortions, plants of all sizes seem to face the prevailing output distortions and plants of all ages face capital distortions.

Third, the actual plant size distribution is less dispersed than efficient size ones and the efficient plant size distribution becomes less dispersed over time; compared to efficient size distribution, many plants overproduced in this period and should be reduced to reach their best size.

Fourth, the decomposition following Chen and Irarrazabal (2015) shows that the total allocation gain is primarily led by TFPR variation. In a modified decomposition and an investigation of TFP gain with the absence of each type of distortion, we find that the TFPR variation is driven mostly by output distortions, with capital distortions playing a smaller role. On the other hand, the TFPR variation is determined mainly by the between-group component of different groups of TFPQ. This between-group TPFR variation happens mostly in the lowest and the highest productivity plants.

Table 1.7: Robustness analysis

	Benchmark	Robust							
Elasticity (σ)	3	2	5	3	3	3	3	3	3
Technology (α)	US	US	US	Vietnam	Plant specific	1/3	US	US	US
Wage	Unity	Unity	Unity	Unity	Unity	Unity	Wage bills	Unity	Unity
TFPR S.D.	0.85	0.86	0.89	0.81	0.78	0.73	0.75	0.76	0.85
TFP gain (%)	97.2	70.3	143.4	84.3	71.7	82.1	73.1	64.3	97.0
Panel	Unbalanced	Unbalanced	Unbalanced	Unbalanced	Unbalanced	Unbalanced	Unbalanced	Balanced	Unbalanced
R	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.12

Appendix

1.A Table

Table 1.A.1: Structure of manufacturing by industry (average in 2000-2013)

Industry	VSIC 1993	Value added share	Employment share
Food products and beverage	15	18.71	14.32
Tobacco products	16	1.257	0.430
Textiles	17	4.474	4.965
Wearing products	18	12.57	18.79
Leather products	19	9.589	16.04
Wood and cork manufacturing	20	2.187	4.549
Paper products	21	1.773	1.773
Publishing, priting, recording media	22	1.404	1.054
Manufacture of coke, refined petroleum products and nuclear fuel	23	0.347	0.078
Chemical manufacturing	24	7.336	2.882
Rubber and plastics products	25	4.779	3.823
Other non-metallic mineral products	26	7.500	7.567
Basic metals manufacturing	27	1.385	1.064
Fabricated metal products, except machinery and equipment	28	4.913	4.693
Machinery and equipment	29	2.711	2.286
Office, accounting and computing machinery	30	0.444	0.389
Electrical machinery and apparatus	31	2.435	1.608
Radio, television and communication equipment	32	3.468	2.343
Medical, precision and optical instruments, watches and clocks	33	0.568	0.403
Motor vehicles, trailers and semi-trailers	34	2.867	1.353
Other transport equipment	35	4.324	2.701
Furniture	36	5.124	7.042

Note: Industry share is computed as mean of value-added share and mean of employment share over 2000-2013. Entries are in percent.

Table 1.A.2: Variation of TFPR by manufacturing industries defined as divisions of VSIC 1993

Industry name	2000	2007	2013
Food products and beverage	0.839	0.813	0.876
Tobacco products	1.143	0.735	0.790
Textiles	0.760	0.736	0.735
Wearing products	0.747	0.787	0.555
Leather products	0.790	0.753	0.714
Wood and cork manufacturing	0.677	0.790	0.734
Paper products	0.752	0.779	0.672
Publishing, priting, recording media	0.700	0.659	0.476
Manufacture of coke, refined petroleum products and nuclear fuel	0.545	0.820	0.975
Chemical manufacturing	0.878	0.866	0.839
Rubber and plastics products	0.797	0.720	0.649
Other non-metallic mineral products	0.739	0.760	0.751
Basic metals manufacturing	0.731	0.744	0.700
Fabricated metal products, except machinery and equipment	0.791	0.732	0.602
Machinery and equipment	0.794	0.675	0.717
Office, accounting and computing machinery	.	0.452	0.880
Electrical machinery and apparatus	0.773	0.704	0.711
Radio, television and communication equipment	0.757	0.844	0.837
Medical, precision and optical instruments, watches and clocks	0.516	0.794	0.697
Motor vehicles, trailers and semi-trailers	0.667	0.710	0.743
Other transport equipment	0.801	0.702	0.744
Furniture	0.790	0.727	0.699

Note: The variation of TFPR for each manufacturing sector is calculated as the standard deviation of $log(TFPR_{si})$, where $TFPR_{si}$ is $TFPR$ of plant i in the two-digit industry s.

1.B Figures

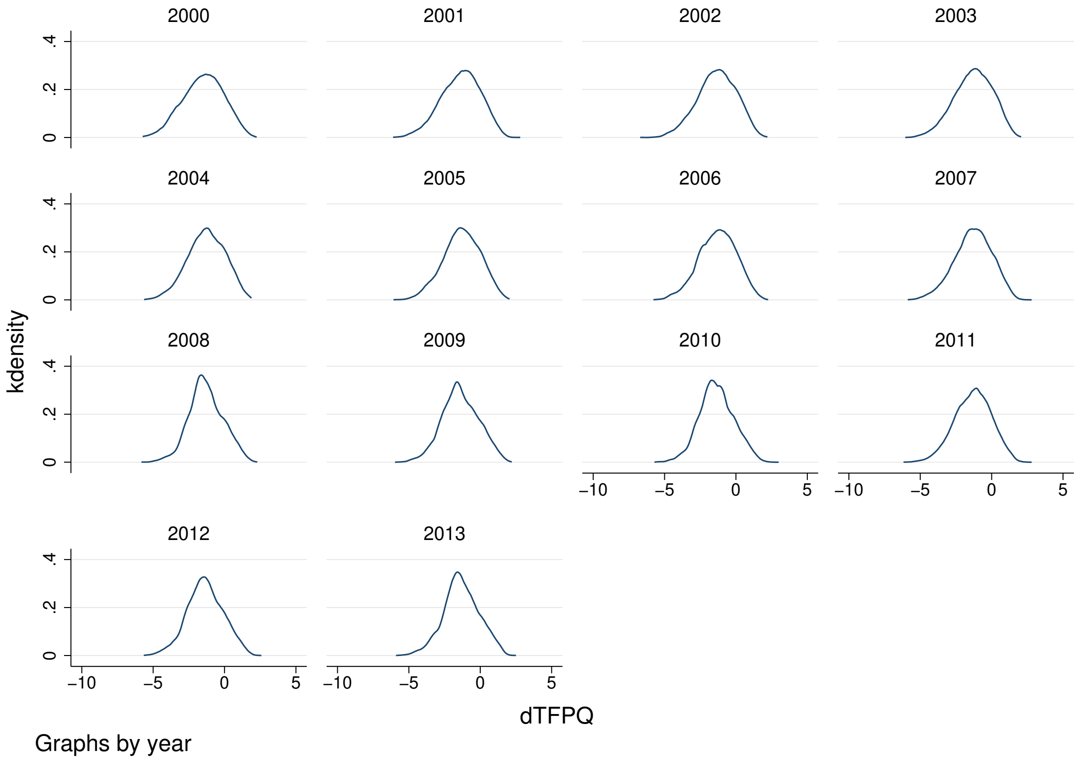


Figure 1.B.1: TFPQ distribution (all years)

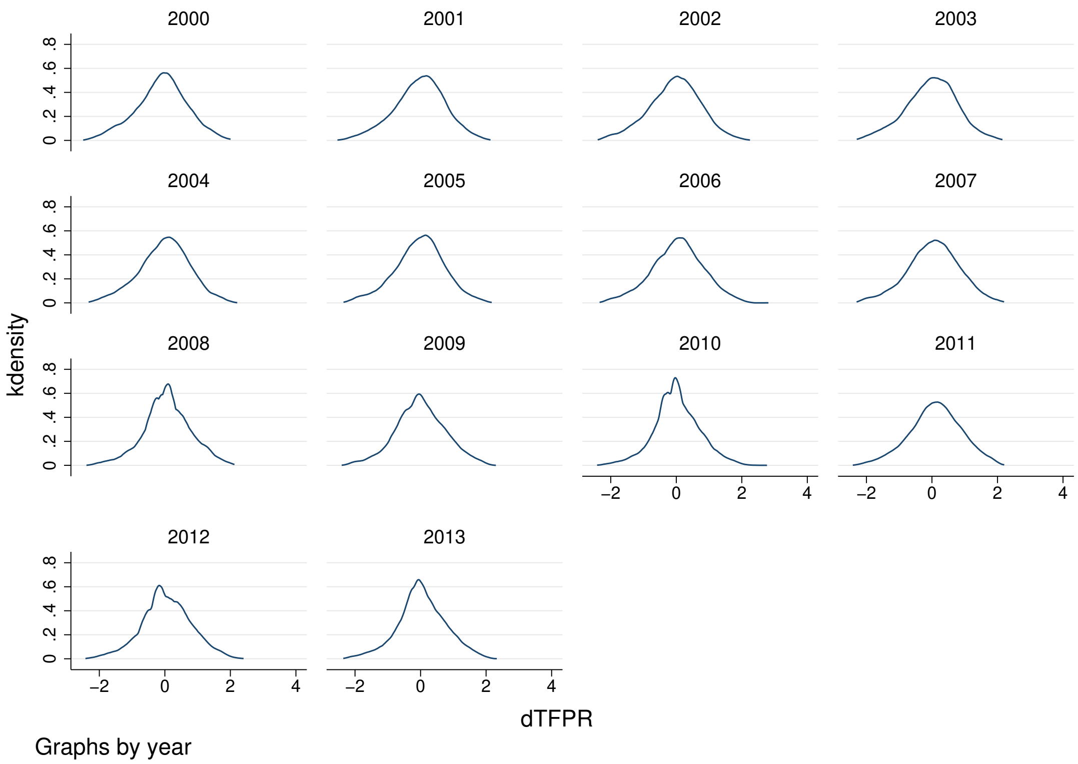


Figure 1.B.2: TFPR distribution (all years)

Chapter 2

Productivity Dynamics in Vietnam: an Application of Dynamic Olley-Pakes Productivity Decomposition

2.1 Introduction

The growing availability of micro data in recent decades provides a better understanding of productivity growth. Using rich longitudinal microdata, while a strand of literature attempts to examine the factors underlying productivity growth (Bartelsman, Leeuwen, and Nieuwenhuijsen, 1998; Bernard and Jensen, 1999; Lichtenberg and Siegel, 1990, 1991), another group of papers, studying productivity dispersion across firms, focuses more on how productivity evolves (Baily, Hulten, and Campbell, 1992; Baldwin and Gorecki, 1998; Bartelsman and Dhrymes, 1998; Pakes and Ericson, 1998).[7] These studies point out the higher degree of persistence of continuing firms as well as the pattern of productivity for entering and exiting firms. These findings on micro productivity growth are important to delve deeper into macro productivity growth. Of papers that use the microdata to describe the productivity evolution, recent studies develop the computations used in empirical analysis to examine the source of aggregate productivity growth by decomposing it into the contribution of different components. Firstly, aggregate productivity growth could be a result of the change in producers' productivity. Secondly, it could vary due to composition change between producers (the change in market share of continuing firms or firm dynamics). Applying the most rigorous decomposition method suggested by Melitz and Polanec (2015), the present paper contributes empirically to this strand of literature by determining the source of aggregate productivity growth in a developing country.

As a small economy, Vietnam has taken measures to enhance aggregate productivity. During the period 2000-2013, Vietnamese firms had to face substantial changes in the business environment starting with the new Enterprise Law in 2000. For the first time, many market barriers were abandoned in favour of private firms. Furthermore, the new law eliminated many business licenses which were considerable entry deterrents to private enterprises. This induced a growing number of firms to enter the market. In this context, using the plant-level data in Vietnamese man-

[7] see Bartelsman and Doms (2000) for further discussion on these two groups of papers.

ufacturing, our main contribution is threefold. First, applying the decomposition method of Melitz and Polanec (2015), we investigate the contributions of incumbents, new entrants and exiters to aggregate productivity. At the same time, we explore plants' life-cycles by examining the post-performance of new entrants and the pre-exit performance of exiting plants. Second, since there is a change in the structure of ownership status due to the new Enterprise Law (a decreasing number of state-owners and a growing number of either private- or foreign-owners), we decompose aggregate productivity into the contributions of groups of plants according to their ownership status and look at the reallocations between these groups. For this purpose, we apply the method augmented by Hashiguchi et al. (2015) from that of Melitz and Polanec (2015), that allows us to examine, at the same time, the reallocation between groups of plants according to their ownership status (State, Foreign, Private) and the reallocations between plants within each of the above groups. Third, the exports of Vietnam considerably increased and like many other developing countries, exporting is often considered a way to increase economic growth. Therefore, using the same methodology as before, we look at the reallocation between exporting and non-exporting plants as the source of aggregate productivity growth, which seems to be rarely considered in the literature.

This chapter has some key findings to report. Firstly, using plant-level data from the Vietnamese Enterprise Survey collected annually by the General Statistics Office of Vietnam (VGSO) from 2000 to 2013, we apply the decomposition method of Melitz and Polanec (2015), called dynamic Olley-Pakes (DOP) decomposition, to explore the source of aggregate productivity growth for Vietnamese manufacturing sector. Considering three groups of plants (survivors, entrants and exiters), we find that the increase in aggregate productivity relative to 2000 is mostly due to the growing contribution of surviving plants, in which the improvement in productivity within plants plays a large part. The market share reallocation between surviving plants is not as good as in the base year 2000 because of its negative values, but at least it becomes less severe as the time span increases.

To identify the contribution of entrants and exiters, this decomposition compares their productivity with that of survivors in the corresponding period. In the case of Vietnam, we find a positive contribution of exiters and a negative contribution of entering plants. This result seems to contrast with that of Ha and Kiyota (2014) who find a positive contribution from new entrants and a negative one from exiting plants. The crucial reason for this difference stems from the productivity of reference used to compute the groups of plants' contributions. Indeed, these authors use the same productivity of reference, that is the (unweighed) average productivity in the first period, while we use the productivity of survivors in the corresponding period as reference.

Most interestingly, we explore the plant's productivity life cycle and find that new plants can learn from market functioning and behave as shown in the model of Ericson and Pakes (1995). Accordingly, new entrants begin with a relatively low level of investment. The majority of new entrants cannot overturn their position of bad initial draws and stay behind the rivals in the competition and eventually have to leave the market. While other entrants with a decent start can extend their earnings and subsequently have more investment, raising the likelihood that they will grow further. Indeed, we follow Bellone, Musso, Quéré, and Nesta (2006) in examining the post-entry performance of new plants and the pre-exit performance of exiting plants. We compare the productivity of new entrants and exiters in the entry (or exit) cohorts to that of plants surviving throughout the 14 years. We find that the group of incumbents has the highest productivity, and the new entrants could not catch up these plants during their present time in the sample. In an alternative way, we modify the method of Bellone et al. (2006) by allowing a more relaxed definition of incumbents at time t as plants being at both time $t-1$ and t. These are plants having been in the market at time t. In this case, we find that newly entrant plants can catch up these incumbents 3 to 5 years after entering the market. Exiting plants are less productive than the incumbents and the difference in productivity is more severe when they are close to their exit year.

Secondly, we find that the positive contribution of surviving plants and the positive contribution of exiting plants are robust when we use the two alternative methodologies of Foster et al. (2001) (hereafter, FHK) and Griliches and Regev (1995) (hereafter, GR). The main difference among these three decompositions is the contrasting contribution of entering plants: positive for both the FHK and GR decompositions in later years and negative for the DOP decomposition in all years. The reason is that these decompositions use different productivities of reference while computing the contribution of the three groups of plants. Indeed, while the DOP decomposition uses the productivity of survivors in the corresponding year (productivity of survivors at $t = 1$ while computing the contribution of exiters, and that at $t = 2$ for the contribution of entrants), FHK and GR use the same productivity of reference in both years: the weighted aggregate productivity in the first period for FHK and the average of weighted aggregate productivity of two periods for GR. This leads to the contribution of entrants being overmeasured for FHK and GR, and much more severely for FHK. Although the contribution of exiters is positive in all cases, it is overestimated for GR and underestimated for FHK.

Thirdly, we apply the methodology of Hashiguchi et al. (2015) who extend that of Melitz and Polanec (2015) to investigate the source of aggregate productivity growth in specific groups of plants. We divide all plants in the sample according to three alternative criteria: (i) the two-digit industry, (ii) ownership status (state-owned, private, foreign-owned) and (iii) export status. We then study the reallocation between groups of plants (State, Foreign, Private in the case of plant ownership status, or Export and Non-export in the case of plant export status, or merely among small industries in the whole sector) and the reallocation between plants within each of the above groups. Considering the two-digit industries, we find that the improvement in industries' productivity (the within-industry component) plays a large part in aggregate productivity growth (about three quarters). The reallocation between industries is positive and improves over time, meaning that the reallocation in the activities (value added) between industries enhances the aggregate productivity growth. To investigate the source of the large within-industry effect, we decompose this component into the within-, between-plant and plant dynamics. We observe that the within-plant component contributes mostly to the change in aggregate productivity growth. The reallocation between plants is negative but improves over time. Furthermore, we find that the three components vary differently according to the output share by ownership status. Indeed, industries with relatively higher foreign owners' output share experience an increase in within, between and net entry components.

To examine the source of the growing aggregate productivity over time according to plants' ownership status, we divide the whole sample into three groups: State, Private and Foreign. We find that the improvement in productivity within each group enhances the aggregate productivity. On the reallocation, Private is the group that experiences a better market share reallocation. This could be the result of the encouragement induced by the policies of the Vietnam government in favour of private plants in the late 1990s.

Regarding the plants' export status, we select our period into two subperiods, 2002-2004 and 2010-2013, in which the export statuses are well reported. In each period, we decompose the aggregate productivity into within- and between-group components. In both periods, the contribution of within-group component is moderate (it is even negative during 2002-2004). However, the market share reallocation (between groups), that improves and plays a large part in the aggregate productivity growth, contributes most to stimulating the aggregate productivity. When decomposing the within-group component into within- and between-plants components for each group of exporters and non-exporters, we find that exporting plants have higher within-plant component and a better market share reallocation than non-exporting ones.

Analysing the sources of aggregate productivity is not a new topic, it is embedded in the Schumpeterian growth theory with the notion of creative destruction. Accordingly, this destruction implicitly entails the creation of new products and new processes in the demise of old products and processes. Inspired by idea of Schumpeter (1942), Aghion and Howitt (1992) developed an endogenous growth model with creative destruction. The quality of the products is improved thanks to continuous endogenous innovation and hence creative destruction is formed. In this process, fierce competition among research and development firms creates various innovations which turn into technological progress and this progress is the main source for growth. As a result of these innovations, new intermediate goods are discovered and can be used more efficiently than the old ones in the production of the final output. The research firms can seek for monopoly rents through patented innovations and this is their foremost motivation in research investment. However, through time, new innovations will discourage the innovations of the previous period because of the monopoly rents and the appearance of new innovations brings death to the old ones. Alternatively, Caballero and Hammour (1994) developed a vintage model of creative destruction by emphasizing the role of firm dynamics. They considered the demand fluctuations as a source of economic changes. In their model, the change in output is a function of the rate of creation and the rate of destruction. Therefore, industries can offset a demand reduction by either reducing the rate of creation of firms which adopt new processes or increasing the rate of destruction of firm with obsolete techniques. At any given creation rate, a demand decrease makes the obsolete technology unprofitable and these firms are forced out of the market.

Empirically, some methods are developed to identify the source of aggregate productivity growth between two points in time, noted as $t = 1$ and $t = 2$. They decompose the changes in aggregate productivity into the within component (change in productivity of surviving plants regardless the change in their market share), the between component (the market share reallocation across surviving production units) and the producer dynamics component (the entry and exit contribution). However, the contributions of those components to aggregate productivity are sensitive to the decomposition methodologies used, the business cycle and the horizon over which the productivity growth is measured (Foster et al., 2001). This leads to differences not only in cross-country comparisons but also in the same country under study. Therefore, the factors which lead to changes in the contributions of these components to productivity growth can be understood if we try to investigate the variations along these dimensions. In this line, we firstly decompose the aggregate productivity in the Vietnamese manufacturing sector by applying three widely used methodologies, which take entry and exit as well as continuing plants into consideration: Foster et al. (2001); Griliches and Regev (1995); Melitz and Polanec (2015). At the same time, we shed light on the differences between our results and those of Ha and Kiyota (2014), who used the methodology of Foster et al. (2001), with the reference productivity set as the unweighted average plant productivity at time $t = 1$.

Each decomposition method used in this chapter has its own pros and cons that are useful to be considered. Foster et al. (2001) developed the method of Baily et al. (1992) and distinguish the change in productivity of survivors (the within effect) and the change in their market share (the between effect) from the joint change of their market share and productivity (the cross/ covariance effect).[8] This makes these contributions easier to interpret. However, the authors also point out that this method could suffer from the random measurement error in the market share, which can produce a positive covariance between total factor productivity (TFP) changes and output share and a falsely low within-plant effect, or a negative covariance between variation in labour productivity and employment share and a falsely high within-plant effect.[9] The method of Griliches and Regev (1995) could mitigate the sensitivity to this type of random measurement error because it uses averaging across time of the shares. However, at the same time, using the average market shares to measure the within-plant effect does not allow the separation of the within-plant effect from the covariance effect. In the third method, Melitz and Polanec (2015) augmented the decomposition of Olley and Pakes (1996) by introducing the effect of entering and exiting plants. This approach mitigates the disadvantage of the two previous methods. Unlike the method of Griliches and Regev (1995), the within effect is simply the change of unweighted plant productivity, so it is not related to the market share reallocation effect. Moreover, since the between component is measured as the cross-sectional (unweighted) mean across all plants, it is more persistent and less dominated by measurement error.[10] In view of this interesting feature, we rely mostly on the method of Melitz and Polanec (2015) to examine the aggregate productivity growth.

This chapter is structured as follows: we present the decomposition methodologies in Section 2.2. Section 2.3 presents various results. Finally, we make some conclusions in Section 2.4.

2.2 Methodology

In this section, we describe two main decomposition methods used in this chapter, that of Melitz and Polanec (2015) (hereafter, MP) and Hashiguchi et al. (2015). The latter is an augmented version of the former, that allows us to decompose aggregate productivity into several subgroups. Both methodologies rely on the well-known method proposed by Olley and Pakes (1996) (hereafter OP). Let's consider aggregate productivity at t, Φ_t, that is defined as a share-weighted sum of plant productivity ϕ_{it} with its market share as weights, s_{it} ($\sum_i s_{it} = 1$):

$$\Phi_t = \sum_i s_{it}\phi_{it} \tag{2.1}$$

[8]The within effect is measured as the change in plant productivity weighted by it market share in the initial period. The between effect is the weighted-sum of change in surviving plant market share with the deviation of plant-level productivity from aggregate productivity in the initial period as weights. See Appendix 2.A for a brief review of this method.

[9]For example, for labour productivity measured as output per number of employees, in the initial period, if plants have spuriously high number of employees, they will have falsely low labour productivity. This implies a negative covariance term and a falsely high within-plant effect. If there is measurement error in ouputs such as plants with spuriously higher output in the initial period will have higher TFP, we will have a spuriously low within-plant effect.

[10]see Foster et al. (2001) for more discussion about these decomposition methods.

According to OP, the aggregate productivity could be decomposed into two components as follows:

$$\Phi_t = \mu_t + \sum_i (s_{it} - \bar{s}_t) \left(\phi_{it} - \bar{\phi}_t\right) \tag{2.2}$$

where $\mu_t = \frac{1}{n_t} \sum_{i=1}^{n_t} \phi_{it}$ is the unweighted mean productivity and $\bar{s}_t = \frac{1}{n_t}$ is the unweighted mean share (n_t is the number of producers in year t). The second component is the covariance between productivity and market share, denoted as cov_t.[11] The positive covariance indicates that plant with higher market share having higher productivity. The increasing covariance implies that the higher market share goes to more productive plants. In that case, we have a better market share reallocation between plants. Based on OP, the change in productivity from $t = 1$ to 2, that means, $\Delta\Phi = \Phi_2 - \Phi_1$, is decomposed in a natural way as:

$$\Delta\Phi = \Delta\mu_t + \Delta cov_t \tag{2.3}$$

One feature of the OP decomposition is that unfortunately, it does not account for the entering and exiting plants, so it does not provide any information on the contribution of these two groups of plants. Therefore, basing on the OP decomposition, MP develop a new method that allows for plant dynamics, called Dynamic OP (DOP) decomposition. They define surviving plants as plants which are present in both $t = 1$ and $t = 2$. Exiting plants designate plants which are present at $t = 1$ but not at $t = 2$, entering plants are plants being present only at $t = 2$. Therefore, in the initial period $t = 1$, we have surviving plants and exiting plants, and in $t = 2$ surviving plants and new entrants.

The productivity of each group of plants is computed as the average plant productivity weighted by its market share in the group. Indeed, MP define $s_t^G = \sum_{i \in G} s_{it}$ as market share of a group G. The market share of plant i in its group is then s_{it}/s_t^G. The (weighted) average productivity of group G is measured as:

$$\Phi_t^G = \sum_{i \in G} \left(\frac{s_{it}}{s_t^G}\right) \phi_{it}$$

For each period, aggregate productivity is calculated from (weighted) average productivity of survivors (S), entrants (E) and exiters (X):

$$\begin{aligned} \Phi_1 &= s_1^S \Phi_1^S + s_1^X \Phi_1^X = \Phi_1^S + s_1^X \left(\Phi_1^X - \Phi_1^S\right) \\ \Phi_2 &= s_2^S \Phi_2^S + s_2^E \Phi_2^E = \Phi_2^S + s_2^E \left(\Phi_2^E - \Phi_2^S\right) \end{aligned} \tag{2.4}$$

The productivity change is then decomposed as follows:

$$\Delta\Phi = \left(\Phi_2^S - \Phi_1^S\right) + s_2^E \left(\Phi_2^E - \Phi_2^S\right) + s_1^X \left(\Phi_1^S - \Phi_1^X\right) \tag{2.5}$$

Applying the original OP method as in Equation (2.3), the change in productivity of surviving group (the first term) can be decomposed as the change in the unweighted mean productivity

[11] The covariance operator would typically be multipled by $\frac{1}{n_t}$. However, according to MP, $\frac{1}{n_t}$ is essentially incorporated already in market share s_{it}.

of survivors and the covariance change between market share and incumbent productivity. That means, $\Phi_2^S - \Phi_1^S = \Delta\mu^S + \Delta cov^S$ Equation (2.5) is rewritten as follows:

$$\Delta\Phi = \left(\Delta\mu^S + \Delta cov^S\right) + s_2^E\left(\Phi_2^E - \Phi_2^S\right) + s_1^X\left(\Phi_1^S - \Phi_1^X\right) \tag{2.6}$$

Equation (2.5) decomposes the aggregate productivity growth into components for the three groups of plants: survivors, entrants and exiters. Different from other decomposition methods (Foster et al., 2001; Griliches and Regev, 1995), Melitz and Polanec (2015) compare productivity of entrants and exiters to that of survivors in the same period.[12] Therefore, each group contribution can be related to a specific counterfactual scenario: the contribution of surviving plants, $(\Phi_2^S - \Phi_1^S)$ is simply the aggregate productivity that would have been obtained in case of no entry and exit. The contribution of entry, $s_2^E\left(\Phi_2^E - \Phi_2^S\right)$, is the change in aggregate productivity generated by adding or removing the group of entrants. In the same logic, the contribution of exit, $s_1^X\left(\Phi_1^S - \Phi_1^X\right)$, is the change in aggregate productivity generated by adding/removing the group of exiting plants. From this decomposition, entrants positively contribute to productivity growth if (and only if) they have higher productivity Φ_2^E than the remaining (surviving) plants Φ_2^S in the same time period when they enter the market ($t = 2$). Exiters positively contribute to productivity growth if (and only if) they have lower productivity Φ_2^X than the remaining (surviving) plants Φ_1^S in the same time period when they exit ($t = 1$).

The Augmented Dynamic Olley-Pakes method

Hashiguchi et al. (2015) propose a method which extend that of Melitz and Polanec (2015) in order to investigate the allocation efficiency as well as the reallocation among several groups. The authors label this method as the Augmented dynamic OP (ADOP) decomposition.

We start from the OP method and suppose that there are J groups (J sectors), the aggregate productivity is:

$$\Phi_t = \textstyle\sum_{j=1}^{J} w_{jt}\tilde{\mu}_{jt} \tag{2.7}$$

where w_{jt} is group j's share at time t, $\tilde{\mu}_{jt}$ denotes the weighted average productivity of this group. Using the OP method for this equation gives:

$$\Phi_t = \frac{1}{J}\sum_{j=1}^{J}\tilde{\mu}_{jt} + \tilde{\text{cov}}_t \tag{2.8}$$

The "between effect" $\tilde{\text{cov}}_t$ denotes the reallocation effect among different J groups, or inter-group allocation efficiency.

The "within effect" $\tilde{\mu}_{jt}$ can be further decomposed using the OP method:

$$\tilde{\mu}_{jt} = \mu_{jt} + \text{cov}_{jt} \tag{2.9}$$

Using the two above equations, we can get the augmented OP (AOP) decomposition:

[12] see Appendix 2.A for a brief review of Foster et al. (2001)'s and Griliches and Regev (1995)'s decomposition method.

$$\Phi_t = \frac{1}{J}\sum_{j=1}^{J}\mu_{jt} + \frac{1}{J}\sum_{j=1}^{J}\text{cov}_{jt} + \tilde{\text{cov}}_t \tag{2.10}$$

where the first term is the unweighted mean of productivity, the second denotes the reallocation within the j-th group and the last one represents the reallocation between groups.

In this line of reasoning, the aggregate productivity decomposition for the first period can be rewritten as:

$$\Phi_1 = \frac{1}{j}\sum_{j=1}^{J}\tilde{\mu}_{j1} + \tilde{\text{cov}}_1 \tag{2.11}$$

where $\tilde{\mu}_{j1} = \sum_{i\in\Omega_{j1}}(s_{i1}/w_{j1})\phi_{i1}$ and $\tilde{\text{cov}}_1 = \sum_{j=1}^{J}(w_{j1} - w_1^*)(\tilde{\mu}_{j1} - \tilde{\mu}_1^*)$. w_1^* and $\tilde{\mu}_1^*$ are the averages of w_{j1} and $\tilde{\mu}_{j1}$. We rewrite the weight $z_{ij1} = s_{i1}/w_{j1}$ as:

$$\sum_{i\in\Omega_{j1}} z_{ij1} = \sum_{i\in\Omega_{j1}^S} z_{ij1} + \sum_{i\in\Omega_{j1}^X} z_{ij1} = z_{j1}^S + z_{j1}^X = 1$$

with Ω_j^S and Ω_j^X denoting the sets of surviving and exiting plants from group j. Further decomposing these terms into the weighted average productivity of surviving plants and the contribution of exiting plants gives:

$$\tilde{\mu}_{j1} = \sum_{i\in\Omega_j^S}\frac{z_{ij1}}{z_{j1}^S}\phi_{i1} + z_{j1}^X\Big(\sum_{i\in\Omega_j^X}\frac{z_{ij1}}{z_{j1}^X}\phi_{i1} - \sum_{i\in\Omega_j^S}\frac{z_{ij1}}{z_{j1}^S}\phi_{i1}\Big) = \Phi_{j1}^S + z_{j1}^X(\Phi_{j1}^X - \Phi_{j1}^S) = \Phi_{j1}^S - ext_j \tag{2.12}$$

where the weighted mean productivity of surviving and exiting plants for group j are Φ_{j1}^S and Φ_{j1}^X, and ext_j is the contribution of exiting plants to group j's aggregate productivity $\tilde{\mu}_{j1}$. Using the spirit of the OP decomposition method, we can decompose the term Φ_{j1}^S as:

$$\Phi_{j1}^S = \frac{1}{N_{j1}^S}\sum_{i\in\Omega_j^S}\phi_{i1} + \sum_{i\in\Omega_j^S}\Big(\frac{z_{ij1}}{z_{j1}^S} - \frac{1}{N_{j1}^S}\sum_{i\in\Omega_j^S}\frac{z_{ij1}}{z_{j1}^S}\Big)\Big(\phi_{i1} - \frac{1}{N_{j1}^S}\sum_{i\in\Omega_j^S}\phi_{i1}\Big) = \mu_{j1}^S + \text{cov}_{j1}^S \tag{2.13}$$

where μ_{j1}^S is the simple mean productivity of surviving plants at time 1 and cov_{j1}^S is the reallocation within group j at time 1. Therefore, the aggregate productivity decomposition for the first period Φ_1 is rewritten as:

$$\Phi_1 = \frac{1}{J}\sum_{j=1}^{J}(\mu_{j1}^S + \text{cov}_{j1}^S - ext_j) + \tilde{\text{cov}}_1 \tag{2.14}$$

where the first term is the "within effect" and the second is the "between effect" among groups.

Similarly, the aggregate productivity at time 2 is:

$$\Phi_2 = \frac{1}{J}\sum_{j=1}^{J}\tilde{\mu}_{j2} + \tilde{\text{cov}}_2 = \frac{1}{J}\sum_{j=1}^{J}(\Phi^S_{j2} + z^E_{j2}(\Phi^E_{j2} - \Phi^S_{j2})) + \tilde{\text{cov}}_2 = \frac{1}{J}\sum_{j=1}^{J}(\mu^S_{j2} + \text{cov}^S_{j2} + ent_j) + \tilde{\text{cov}}_2 \tag{2.15}$$

where the first and second terms are the within and between effects while $ent_j = z^E_{j2}(\Phi^E_{j2} - \Phi^S_{j2})$ is the contribution of entering plants to aggregate productivity $\tilde{\mu}_{j2}$. At the end, the augmented dynamic OP (ADOP) decomposition is the difference between Φ_1 and Φ_2:

$$\Delta\Phi = \frac{1}{J}\sum_{j=1}^{J}(\Delta\mu^S_j + \Delta\text{cov}^S_j + ent_j + ext_j) + \Delta\tilde{\text{cov}} \tag{2.16}$$

where the first term is the change in within group reallocation and the second is the change in intergroup reallocation. Furthermore, cov^S_j is the change among surviving plants within group j.

2.3 Results

2.3.1 Data and variables

Data

In this research, we use the data from the Vietnamese Enterprise Survey collected annually by the General Statistics Office of Vietnam (VGSO) from 2000 to 2013. The rich dataset includes registered firms in various industries such as agriculture, manufacturing, construction, transport and services. However, in this research, we focus on the manufacturing industry. Firms are distinguished by their own tax identifiers. These surveys report abundant information of various types of firm (state-owned enterprises (SOEs, hereafter), foreign-owned firms, domestic private firms). There is general information (tax identifiers, firm code, plant code, type of ownership, year of establishment, export/import status, major industry, city), worker related information (number of workers, wages), and accounting information (assets, liabilities, sales,...). In order to make this paper comparable with previous studies, we only take into account firms with more than 10 workers. Since we are willing to use the data at plant-level, we combine firm identifiers with plant code and city to distinguish each plant.

However, the dataset had some features which create difficulties. Firstly, export status and export sales are not available for every year. So, we use other information such as export taxes to determine whether plants exported or not. Even so, the number of trading plants is still under-represented for some years. Secondly, since intermediate input data is not available in the dataset, we use an indirect way to calculate intermediate input. In fact, it can be calculated as follows:[13]

$$\textit{Intermediate Input}_{it} = \textit{Sales}_{it} - (\textit{Wages}_{it} + \textit{Depreciation}_{it} + \textit{PretaxProfit}_{it}) \tag{2.17}$$

[13]Unfortunately, we do not have data on interest paid on credit and loans, which is part of the payments to capital. Therefore, as in Ha and Kiyota (2014), we use the more relaxed definition of value added, that is the sum of wages, depreciation and pre-tax profit.

Thirdly, in the year of 2001, as the data on foreign-owned plants' equity shares are not available, we interpolate the data in 2001 as follows: if plants are foreign-owned in 2000 and 2002, then they are so in 2001. Additionally, some plants changed industry during this period. To deal with this problem, we follow Ha et al. (2016) in defining the industry to which the plants belong.[14] Then, we also identify entry, exit and continuing plants and only account for plants without re-entry. Thereby, we obtain unbalanced panel data for 59,388 plants over the period 2000-2013.

During the period studied, there was change in Vietnam Standard Industry Classification (VSIC, hereafter). Before 2007, VSIC was based on ISIC Rev.3, which was called VSIC 1993. From 2007, the old classification was replaced by VSIC 2007 which was developed on the basis of ISIS Rev.4 and ASIAN Common Industrial Classification (ACIC). For comparison with previous studies, we convert VSIC 2007 to VSIC 1993.

Variables

The main variable involving aggregate productivity that we use in our analysis is total factor productivity (TFP) with value added share as weights. TFP at plant-level is measured as the residuals in the regression of the plant output and its inputs. The productivity of plant i in a given 2-digit industry is given by:

$$tfp_{it} = y_{it} - \hat{\beta}_k k_{it} - \hat{\beta}_l l_{it} \tag{2.18}$$

where tfp_{it} is plant-level TFP, y_{it}, k_{it} and l_{it} denote real value added, real capital and number of employees of plant i. All variables are taken in logs. $\hat{\beta}_k$ and $\hat{\beta}_l$ are output elasticities of capital and labour for each 2-digit industry. These are estimated following the Levinsohn and Petrin (2003)'s approach that uses the intermediate input demand to control for productivity shocks.[15] As pointed out by Foster et al. (2001), the decomposition results could be sensitive to the choice of productivity measurement and weights. Therefore, we also use labour productivity and employment shares as weights for comparison.

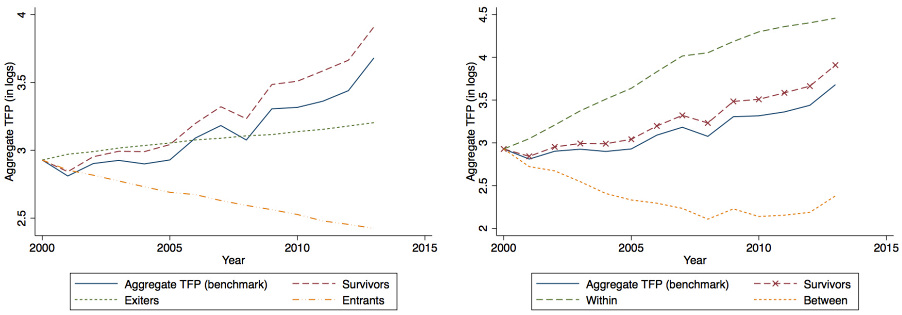
Using the aggregate productivity decomposition methods that account for plant dynamics, we define surviving plants as plants which are present in both $t = 1$ and $t = 2$, exiting plants designate those which are present in $t = 1$ but not in $t = 2$, entering plants are plants being present only in $t = 2$.

2.3.2 DOP decomposition for Vietnamese manufacturing sector

One important feature of DOP decomposition is that the decomposition uses survivors' productivity as a benchmark to determine the contribution of entrants and exiters. Melitz and Polanec (2015) argue that because we cannot observe either the productivity of entrants in period 1, or the productivity of exiters in period 2, this decomposition interprets more accurately the contribution of those two groups of plants in a specific counterfactual scenario.

[14] If a plant has changed industries, the plant's industry is the one in which this plant stays most of the time. If a plant stays in two or more industries for the same length of time, we use the most recent industry as the plant's industry.

[15] See Appendix 2.B for more details on the production function estimation.



(a) Benchmark *vs.* counterfactual TFP evolution

(b) Within *vs.* between component (in logs)

Figure 2.1: Aggregate TFP evolution

Figure 2.1 shows the evolution of aggregate TFP, that is estimated from the data (benchmark) and the counterfactual ones corresponding to the contribution of each group of plants (survivors, entrants and exiters) (Figure 2.1a) and the contribution of the within and between effect (Figure 2.1b). To obtain the evolution of counterfactual aggregate TFP, we begin with the aggregate productivity in the first year of the period, 2000, then add the year-over-year contribution of each group of plants.[16] Those contributions are estimated following Equation (2.5) and Equation (2.6). In this exercise, we use value-added share as weight.[17]

As shown in Figure 2.1a, the benchmark aggregate productivity (the solid blue line) seems to be stable in the first five years, and then steadily grows from 2005. The dashed red line corresponds to an experiment in which the aggregate productivity is only due to the change in productivity of survivors and the contributions of other groups are zero. This line has a similar pattern to the benchmark . Moreover, it stays above the aggregate TFP line (the solid blue line), suggesting that, in the absence of entrants and exiters, the aggregate productivity would have increased with a higher growth rate than the benchmark. Further, as indicated in the DOP methodology, the contribution of surviving plants could be decomposed into the change in the unweighted average plant productivities (the within-plant component) and the covariance change between incumbent productivity and its market share (the between-plant component or the market share reallocation). By a similar process, we represent the experimental aggregate productivity corresponding to these two components in Figure 2.1b. If the productivity growth was due only to the change in the unweighted average surviving plant productivities (the dashed green line in Figure 2.1b), the aggregate TFP would have increased sharply from 2.9 to 4.5, against 3.7 for the benchmark line (the solid blue line) and 3.9 for the counterfactual one corresponding to the contribution of survivors (the connected dashed red line). The increasing pattern of this counterfactual line implies, firstly, the positive contribution of the within component, and secondly, that the experimental productivity growth when there was only the increase in within-plant component would have been higher than the estimated productivity growth (benchmark).

By contrast, the counterfactual aggregate TFP generated by the between component (the dashed yellow line) would have reduced over time, especially before 2008. This suggests negative (counterfactual) productivity growth, or in other words, negative contribution of the between com-

[16] In other words, if $t = 2$ is the current year, $t = 1$ indicates the previous year.

[17] As mentioned in Section 2.3.1 we use a more relaxed definition of value added, that is the sum of wages, depreciation and pre-tax profit, because we do not have data on credit and loans for all years.

ponent. After 2008, this experimental line would have slightly increased, implying the positive contribution of this component. Moreover, since the between component (or the market share reallocation) is measured as change in the covariance between plant productivity and output share, its positive value indicates that the higher share of output goes to more productive plants. Therefore, we have a better reallocation between surviving plants in laters years.

We now return to Figure 2.1a. The second experiment (plotted in dashed green) represents the evolution of aggregate productivity while considering only the contribution of exiters. This counterfactual aggregate productivity slightly increases, suggesting that the contribution of exiters is positive. In other words, plants exiting the market have relatively lower productivity than that of incumbents. The short-dashed yellow line shows the third experiment of adding the group of entrants while the contribution of other groups is zero. In this experiment, the couterfactual aggregate TFP would have been downward slopping, which indicates that the new entrants have lower productivity than the incumbents. It could be the case that it takes time for them to learn. When we take longer period differences, such as 3-year differences and 5-year differences (the dashed red line and the dashed-dotted green line respectively in Figure 2.2), the contribution of new entrants is still negative. But for the longer period differences, the counterfactual line become higher, which means that productivity gap between new entrants and the incumbents is closer.

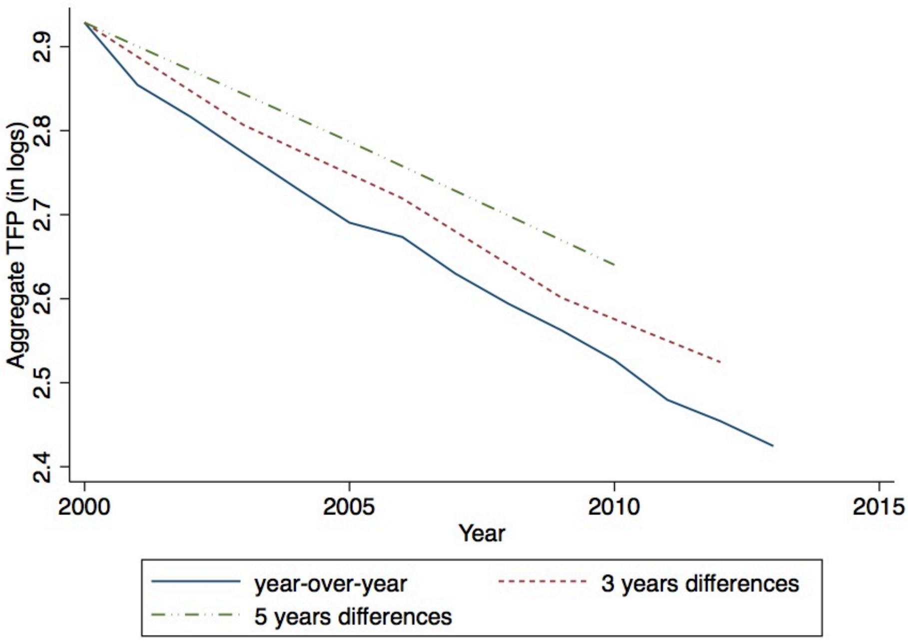


Figure 2.2: Counterfactual aggregate TFP with the contribution of new entrants

Table 2.1 decomposes the aggregate productivity change relative to 2000 (the first period (t = 1) is fixed at 2000). In this way, we can examine the source of aggregate productivity when the time span increases. As defined above, the incumbents consist of plants being in both 2000 and t = 2, exiting plants are those being in 2000 but leaving from the market in t = 2, new entrants are plants not present in 2000 but present in t = 2. This setting would help us to investigate the long term contribution of each group of plants to aggregate productivity growth. Following Melitz and Polanec (2015), we use alternatively two measures of productivity: TFP (the top panel) and labour productivity (the bottom panel) measured by value-added share per number of employees, with their corresponding market share measurement (value-added share and employment

Table 2.1: Decomposition of aggregate productivity change relative to 2000

Year	Within	Between	Surviving plants	Plant dynamics Entering plants	 Exiting plants	Aggregate TFP growth
			TFP (in log percent) - Value-added share as weight			
2001	12.07	−20.67	−8.60	−7.42	4.24	−11.78
2002	18.74	−17.01	1.73	−10.35	5.98	−2.64
2003	25.29	−21.41	3.88	−12.19	8.02	−0.29
2004	27.79	−29.69	−1.90	−11.04	10.04	−2.90
2005	31.58	−30.56	1.03	−14.16	13.16	0.03
2006	40.83	−13.77	27.06	−19.58	8.67	16.15
2007	52.35	−13.02	39.33	−21.98	8.04	25.39
2008	50.05	−26.83	23.22	−17.12	8.64	14.74
2009	62.75	−4.55	58.20	−30.28	9.78	37.70
2010	69.07	−7.14	61.92	−31.93	8.77	38.76
2011	70.64	−17.15	53.49	−20.50	10.32	43.31
2012	72.39	−2.19	70.20	−30.40	11.29	51.09
2013	76.43	4.56	80.99	−19.58	13.75	75.16
			Labour productivity (in log percent) - Employment share as weight			
2001	13.88	−5.42	8.46	−12.08	2.27	−1.34
2002	21.08	−9.36	11.72	−14.46	5.15	2.42
2003	28.17	−7.48	20.69	−16.72	6.44	10.41
2004	31.90	−6.23	25.68	−18.43	7.98	15.23
2005	36.63	−9.90	26.73	−17.70	10.57	19.60
2006	41.92	−3.98	37.94	−17.14	10.38	31.18
2007	54.76	−8.10	46.66	−17.15	12.72	42.23
2008	50.88	−11.05	39.83	−20.44	14.33	33.72
2009	63.52	5.34	68.85	−30.76	15.29	53.38
2010	70.43	1.47	71.90	−24.49	16.31	63.72
2011	70.19	2.93	73.12	−26.20	18.73	65.65
2012	74.97	10.85	85.82	−25.30	18.02	78.55
2013	79.84	9.09	88.93	−20.35	19.52	88.10

share, respectively.) Let's consider the top panel of Table 2.1 in which we use the aggregate TFP changes. The within and between components sum to the contribution of survivors. The contribution of survivors, entrants and exiters sum to the aggregate productivity growth, that is reported in the last column. Firstly, in the case of Vietnamese manufacturing, the positive aggregate TFP growth is due mostly to the contribution of survivors, in which the within component is positive and plays a large part. However, the between component is negative, meaning that the market share reallocation between incumbents is not as good as the base year 2000. Its negative values could be explained by the reduction of the covariance term in the investigated year relative to 2000 (Table 2.C.1 in Appendix 2.C).[18] Secondly, the group of exiters contribute positively to the aggregate TFP changes. Looking at Table 2.5 which shows the (aggregate) TFP of each group of plants in t, we observe that in the first period ($t = 1$), the productivity of exiters is lower than that of incumbents. So, plants with lower productivity than that of those remaining in 2000 ($t = 1$) will decide to exit the market in the second period ($t = 2$). Thirdly, the entrants generate negative productivity growth. That is because the productivity of entrants is lower than that of survivors in the second period, although it is higher than aggregate TFP in $t = 1$ (Φ_1) (Table 2.5).

[18] As shown in Table 2.C.1 in Appendix 2.C, the covariance terms in both periods $t = 1$ and $t = 2$ are positive, so, we always have that plants with a higher market share are more productive.

It seems that using different productivity measures and weights with DOP decomposition does not change our results. Indeed, besides the TFP and value-added share, we also decompose labour productivity and consider employment share as weight. As shown in the bottom panel of Table 2.1, this productivity measure and weight reflects the same pattern of contribution of the three groups when the time span increases: survivors contribute mostly to aggregate productivity, exiters positively contribute and entrants make a negative contribution because their productivity is lower than that of survivors in the same period. Moreover, when the time span increases, it seems that the negative contribution of entrants is more severe. This is because, on the one hand, the weighted average of the entrants group is lower than that of the incumbents. On the other hand, as shown in Table 2.5, the market share of new entrants grows sharply as the time span increases, from 0.17 in 2001 to 0.77 in 2013. This amplifies the negative contribution of this group, that is measured as the difference between productivity of entrant group and that of incumbent group, weighted by the market share of entrant group.

The decomposition method gives information about the productivity of new entrants in the year they enter the market, but does not indicate their performance afterward. For the latter purpose, we attempt to assess the post-entry performance of new plants. To do so, we follow and then modify Bellone et al. (2006) by following the average productivity of different entry cohorts relative to incumbent plants over time. Firstly, we consider incumbents as surviving plants throughout 14 years, from 2000 to 2013 (as in Bellone et al. (2006)). Then, in an alternative way, we allow a more relaxed definition of incumbents at time t, that includes, plants being at both time $t-1$ and t. The set of incumbents in this case is bigger than that in the previous case. Indeed, it is obvious that the incumbent group includes plants being present in all of the 14 years. Other than that, under this definition, this group consists of plants entering at a time before $t-1$ and being in the market at least until t. For example, incumbents in 2003 are plants being present at 2002 and 2003, including: (i) surviving plants throughout the period from 2000-2013 (as in the previous case), (ii) plants entering in 2000, 2001 and 2002, and staying in the market at least until the end of 2003.[19] In other words, we account for all plants (excluded entrants in the entry cohort year) having been in the market in the investigated year. For instance, we attempt to compare the average productivity of entrants in the cohort year, such as 2002, in 2002, 2003, and so on, to that of surviving plants in the corresponding years (2002, 2003, and so on). In both cases, the relative productivity of the entry cohort is defined as the unweighted average of the relative individual TFP. The latter is defined as:

$$lnTFP_i^r t = lnTFP_i t - ln\bar{TFP}_t^{inc} \tag{2.19}$$

where $ln\bar{TFP}_t^{inc}$ denotes the unweighted average TFP of incumbent plants at t.

Table 2.2 shows the TFP level of entrants relative to incumbent plants by entry cohort. The top panel considers incumbents as surviving plants throughout the period, the bottom panel is related to plants being in both the considered year and the previous one. In comparison to the same group of incumbents (in the top panel), entrants can hardly catch up the incumbent plants in the sample, even after surviving several years. However, the productivity gap reduces over time, which implies that a learning process, even a slow one, could help entrant plants ameliorate their productivity. When compare to plants present in the investigated year (in the bottom panel), the

[19]It should be noted that we leave out plants re-entering the market.

Table 2.2: Relative productivity of the entry cohort to incumbent plants

(a) Fix incumbent plants

	Cohort year												
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
2001	−0.72												
2002	−0.39	−0.81											
2003	−0.36	−0.40	−0.85										
2004	−0.32	−0.33	−0.42	−0.78									
2005	−0.27	−0.34	−0.37	−0.46	−0.79								
2006	−0.28	−0.29	−0.31	−0.38	−0.42	−0.78							
2007	−0.25	−0.27	−0.30	−0.36	−0.36	−0.44	−0.70						
2008	−0.25	−0.20	−0.27	−0.36	−0.33	−0.36	−0.44	−0.79					
2009	−0.21	−0.25	−0.25	−0.34	−0.35	−0.34	−0.41	−0.86	−0.74				
2010	−0.24	−0.15	−0.25	−0.30	−0.28	−0.31	−0.37	−0.83	−0.54	−0.69			
2011	−0.20	−0.13	−0.20	−0.31	−0.25	−0.25	−0.29	−0.81	−0.47	−0.49	−0.64		
2012	−0.20	−0.10	−0.16	−0.26	−0.19	−0.24	−0.25	−0.76	−0.36	−0.39	−0.45	−0.56	
2013	−0.18	−0.06	−0.10	−0.20	−0.16	−0.15	−0.20	−0.66	−0.32	−0.30	−0.37	−0.41	−0.55

(b) Dynamic incumbent plants

	Cohort year												
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
2001	−0.48												
2002	−0.21	−0.54											
2003	−0.11	−0.15	−0.54										
2004	−0.05	−0.06	−0.14	−0.45									
2005	0.02	−0.04	−0.06	−0.16	−0.47								
2006	−0.00	0.01	0.00	−0.06	−0.15	−0.46							
2007	0.03	0.02	0.01	−0.05	−0.07	−0.13	−0.37						
2008	0.03	0.10	0.04	−0.05	−0.03	−0.05	−0.14	−0.44					
2009	0.26	0.23	0.25	0.16	0.13	0.17	0.10	−0.54	−0.24				
2010	0.19	0.29	0.22	0.17	0.16	0.17	0.12	−0.49	−0.07	−0.21			
2011	0.19	0.26	0.23	0.11	0.14	0.18	0.15	−0.46	−0.03	−0.07	−0.21		
2012	0.14	0.26	0.24	0.11	0.15	0.13	0.15	−0.42	0.04	−0.01	−0.08	−0.16	
2013	0.09	0.25	0.24	0.12	0.12	0.15	0.14	−0.38	0.01	0.02	−0.05	−0.08	−0.19

entrants could catch up with these plants in three to five years after their entry. This is because these incumbents consist of the highest productive plants (plants surviving throughout the period) and the relatively lower productive ones (plants entering before the investigated year). The latter diminishes the productivity of the group of incumbent plants in this case.

Using the same approach, we investigate the pre-exit performance of exiters in Table 2.3. As in the case for post-entry performance, we consider two different cases corresponding to different definitions of incumbent plants: (i) incumbents consist of plants being in the market throughout 14 years (from 2000-2013) (the top panel), and (ii) incumbents include plants staying in the market in the investigated year and the year before (the bottom panel). For example, we compare the productivity of exiting plants in the cohort year, such as 2003, in 2000, 2001 and 2002 to that of incumbents in the corresponding years.[20] According to Table 2.3, the negative productivity of the exit cohort relative to incumbent plants implies that the productivity of exiters is lower than that of incumbents, whether the number of incumbents is fixed or changes over time. The productivity gap becomes wider when plants are closer to their exit year. Moreover, the smaller productivity gap in the case of dynamic incumbents (Table 2.3a) compared to the case of fixed incumbents (Table 2.3b) is due to the fact that the former consists of not only the latter, the most productive

[20]It should be noted that, in this specification: (i) exiters in year t are defined as plants not being in the market in this year but the previous one $t-1$, in respect of the definition of exiters we use throughout this study; (ii) the corresponding dynamic incumbents are plants present in both $t-1$ and t but do not include plants entering in $t-1$. According to our definition, for the corhort year 2001, exiting plants are those being in the market in 2000, but not 2001. Dynamic incumbents in 2001 are plants being in both 2000 and 2001 but do not include plants entering in 2000. However, since our period studied begin in 2000, all plants are considered as new entrants in 2000. So, we would not have dynamic incumbents in 2001. Therefore, in Table 2.3b we do not consider plants exiting in the cohort year 2001.

Table 2.3: Relative productivity of the exit cohort to incumbent plants

(a) Fix incumbent plants

	Cohort year												
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
2000	−0.53	−0.33	−0.32	−0.38	−0.37	−0.34	−0.32	−0.52	−0.33	−0.21	−0.16	−0.19	1.21e-09
2001	−0.72	−0.50	−0.54	−0.52	−0.45	−0.44	−0.45	−0.51	−0.48	−0.51	−0.35	−0.40	−0.20
2002		−0.66	−0.60	−0.56	−0.42	−0.44	−0.49	−0.56	−0.52	−0.39	−0.32	−0.33	−0.21
2003			−0.69	−0.60	−0.45	−0.49	−0.55	−0.48	−0.51	−0.43	−0.36	−0.35	−0.26
2004				−0.64	−0.55	−0.53	−0.54	−0.51	−0.49	−0.50	−0.39	−0.35	−0.23
2005					−0.66	−0.55	−0.54	−0.57	−0.47	−0.47	−0.47	−0.37	−0.25
2006						−0.63	−0.55	−0.53	−0.53	−0.50	−0.48	−0.39	−0.24
2007							−0.62	−0.57	−0.52	−0.52	−0.49	−0.39	−0.26
2008								−0.57	−0.67	−0.63	−0.62	−0.63	−0.40
2009									−0.76	−0.70	−0.68	−0.66	−0.42
2010										−0.68	−0.68	−0.62	−0.41
2011											−0.74	−0.62	−0.40
2012												−0.63	−0.35
2013													−0.37

(b) Dynamic incumbent plants

	Cohort year												
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
2000	-	−0.24	−0.01	0.05	−0.05	−0.09	−0.00	0.14	−0.22	−0.02	0.08	0.12	0.18
2001	-	−0.56	−0.28	−0.27	−0.26	−0.23	−0.20	−0.11	−0.28	−0.26	−0.30	−0.15	−0.12
2002			−0.45	−0.33	−0.31	−0.22	−0.17	−0.18	−0.30	−0.27	−0.16	−0.10	−0.07
2003				−0.42	−0.33	−0.25	−0.20	−0.22	−0.21	−0.23	−0.18	−0.14	−0.05
2004					−0.36	−0.31	−0.23	−0.22	−0.23	−0.20	−0.23	−0.15	−0.06
2005						−0.41	−0.24	−0.21	−0.26	−0.16	−0.18	−0.22	−0.05
2006							−0.38	−0.25	−0.23	−0.23	−0.22	−0.23	−0.08
2007								−0.34	−0.26	−0.22	−0.24	−0.23	−0.08
2008									−0.24	−0.38	−0.33	−0.33	−0.31
2009										−0.33	−0.23	−0.22	−0.17
2010											−0.26	−0.26	−0.16
2011												−0.37	−0.21
2012													−0.30

plants, but also the less productive one. Those less productive plants slowdown the productivity of the incumbent groups, and hence diminish the productivity gap between exiters and the dynamic incumbents.

In sum, on aggregate productivity decomposition, we find that aggregate productivity growth is mostly due to the growing contribution of surviving plants, in which the change in unweighted plant productivities (the within component) plays a major role. The exiting plants contribute positively to aggregate productivity growth because they have relative lower productivity than survivors. By contrast, we point out a negative contribution of new entrants since their productivity is lower than that of incumbents.

Looking at the post-entry and pre-exit performance of plants, we have some major findings: (i) plants surviving throughout the period 2000-2013 have the highest productivity, (ii) entrant plants could not catch up this group of plants, but 3 to 5 years after their entrance, they can catch up other plants which have already been in the market, (iii) exiters are less productive than either incumbents throughout the period 2000-2013 or plants having been in the market, especially when they are closer to their exit year. So, plant's productivity life-cycle behaves as we would expect.

2.3.3 DOP decomposition in comparison with other decomposition methods

In Table 2.4, following Melitz and Polanec (2015), we decompose aggregate productivity growth to estimate the contributions of all three groups of plants (survivors, entrants and exiters)

Table 2.4: Comparison of decompositions of aggregate productivity change relative to 2000

Year	Surviving plants			Entering plants			Exiting plants			All plants
	GR	FHK	DOP	GR	FHK	DOP	GR	FHK	DOP	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
2001	−7.47	−7.29	−8.60	−7.16	−8.15	−7.42	2.85	3.66	4.24	−11.78
2002	0.73	0.83	1.73	−8.13	−8.46	−10.35	4.77	4.99	5.98	−2.64
2003	1.50	1.52	3.88	−8.15	−8.20	−12.19	6.36	6.39	8.02	−0.29
2004	−3.29	−3.02	−1.90	−6.95	−7.57	−11.04	7.34	7.68	10.04	−2.90
2005	−2.30	−2.31	1.03	−7.28	−7.28	−14.16	9.62	9.61	13.16	0.03
2006	12.14	10.34	27.06	−4.47	−0.05	−19.58	8.48	5.87	8.67	16.15
2007	16.90	13.90	39.33	−1.23	6.36	−21.98	9.72	5.13	8.04	25.39
2008	7.83	5.94	23.22	−1.23	3.55	−17.12	8.14	5.25	8.64	14.74
2009	22.23	17.44	58.20	2.03	14.43	−30.28	13.44	5.83	9.78	37.70
2010	22.11	16.98	61.92	3.39	16.73	−31.93	13.26	5.05	8.77	38.76
2011	17.82	11.62	53.49	10.17	25.93	−20.50	15.32	5.77	10.32	43.31
2012	22.12	14.85	70.20	11.15	30.13	−30.40	17.82	6.12	11.29	51.09
2013	25.84	14.79	80.99	24.29	53.13	−19.58	25.03	7.24	13.75	75.16

Note: This table is based on Table 3 in Melitz and Polanec (2015). Entries are TFP change in log percent (can also be interpreted as percentage point changes) with value-added share as weight.

and compare these contributions across different methodologies reviewed in Section 2.2 and Appendix 2.A. We choose 2000 as the base year and report the decomposition during the period 2001-2013 to examine the effect of the time span on measurement biases.

Overall, we observe that the aggregate productivity of the manufacturing sector increased gradually from 2001, despite of a decrease in 2008 that could be due to the global financial crisis (***Column 10***).

At first, we observe that all three decompositions have some common features: (i) the group of surviros contribute mostly positively to the change in aggregate productivity, (ii) similarly, the group of exiters has positive contribution. The most obvious difference among the three decompositions is the contrasting contributions of entrants: the positive contributions of entrants for both GR and FHK decompositions in later years, and the all negative ones for DOP decomposition. This difference is due to the mis-measured contribution of this group of plants, that will be discussed later.

In a deeper investigation, these decomposition methods reveal some more differences. Firstly, while difference in the contribution of survivors, and also entrants and exiters in three cases is moderate in the short-run, it expands over time. Indeed, at the end of the period, the contribution of survivors by either the GR or FHK method is only a third (for GR) or a fifth (for FHK) of that calculated from DOP decomposition (***Columns 1-3***). This undermeasurement of the contribution of survivors in the GR and FHK decomposition is related to the difference in measurement of the contributions of entrants and exiters. In the case of Vietnam, the contribution of entrants for either FHK or GR decomposition is over-measured, and the bias for FHK is the most severe because it uses the lowest reference productivity level $\left(\Phi_1 < \bar{\Phi} < \Phi_2^S\right)$ (***Columns 4-6***). Another reason explaining the lowest contribution of entrants for the DOP decomposition relatively to other methods could be the evolution of aggregate productivity of entrants and survivors. According to Table 2.5, aggregate productivity of both surviving and entering plants grows over time. Therefore, the contribution of entrants, measured as the difference between the productivity of those plants and that of survivors, become smaller. For the FHK decomposition, since the reference productivity is fixed at $t = 1$ (*i.e.*, 2000), the contribution of entrants to aggregate TFP increases when the productivity

Table 2.5: Aggregate TFP and value-added share

Year		In $t=1$					In $t=2$				
		All plants	Surviving plants		Exiting plants		All plants	Surviving plants		Entering plants	
$t=1$	$t=2$	Φ_1 (1)	Φ_{S1} (2)	s_{S1} (3)	Φ_{X1} (4)	s_{X1} (5)	Φ_2 (6)	Φ_{S2} (7)	s_{S2} (8)	Φ_{E2} (9)	s_{E2} (10)
2000	2001	2.93	2.97	0.86	2.66	0.14	2.81	2.89	0.83	2.44	0.17
2000	2002	2.93	2.99	0.84	2.63	0.17	2.90	3.01	0.76	2.58	0.25
2000	2003	2.93	3.01	0.80	2.61	0.20	2.93	3.05	0.67	2.68	0.34
2000	2004	2.93	3.03	0.77	2.60	0.24	2.90	3.01	0.57	2.75	0.43
2000	2005	2.93	3.06	0.73	2.57	0.27	2.93	3.07	0.52	2.78	0.49
2000	2006	2.93	3.02	0.68	2.75	0.32	3.09	3.29	0.45	2.93	0.55
2000	2007	2.93	3.01	0.64	2.79	0.36	3.18	3.40	0.40	3.04	0.60
2000	2008	2.93	3.02	0.61	2.80	0.39	3.08	3.25	0.35	2.98	0.65
2000	2009	2.93	3.03	0.60	2.78	0.40	3.31	3.61	0.34	3.15	0.66
2000	2010	2.93	3.02	0.58	2.81	0.42	3.32	3.64	0.31	3.17	0.69
2000	2011	2.93	3.03	0.56	2.80	0.44	3.36	3.57	0.27	3.29	0.73
2000	2012	2.93	3.04	0.54	2.80	0.46	3.44	3.74	0.26	3.33	0.74
2000	2013	2.93	3.07	0.53	2.78	0.47	3.68	3.88	0.23	3.62	0.77

Note: The DOP decomposition is applied.

of this group rises.

By the same logic, for exiting plants, the contribution for GR decomposition is highly overestimated as its productivity level of reference is $\bar{\Phi}$, which is higher than aggregate productivity in the base year Φ_1 (for FHK) and that of survivors in the base year Φ_1^S (for DOP) (***Column 7-9***). Moreover, in the case of Vietnam, the contribution of exiters for FHK decomposition is underestimated because aggregate productivity of survivors is even higher than that of all plants in the same period.

Secondly, when the time span increases, while entering plants contribute positively to aggregate TFP growth in the case using the GR and FHK decompositions, their contribution is negative for the DOP decomposition. This implies that new plants might decide to enter the market while they have productivity higher than either the aggregate productivity in the past (FHK) or the mean aggregate productivity (GR). However, the DOP decomposition results suggest that their productivity is still lower than that of the incumbents. In other words, the incumbents remain the most productive plants (Table 2.5).

So, in the case of Vietnam, while comparing the three decompositions, we show that: (i) the overmeasurement of the contribution of entrants for FHK decomposition is the most serious due to its lowest productivity of reference; (ii) the contribution of exiters is overestimated for GR decomposition but underestimated for FHK; (iii) the contribution of survivors is thus mismeasured for those two alternative decompositions, especially for FHK decomposition. Our findings are similar to that of Melitz and Polanec (2015), who apply those three methods for Slovenian manufacturing. Despite those differences, all decomposition methods show the largest contribution of surviving plants and the positive contribution of exiting plants. More interestingly, although new entrants may have lower productivity than incumbents, their relative productivity improves over time. This evidence shows that new plants can learn from market functioning and behave as shown in the model of Ericson and Pakes (1995). Accordingly, new entrants begin with a relatively low level of investment. The majority of new entrants cannot overturn their position of bad initial draws and stay behind the rivals in the competition and eventually have to leave the market. While other entrants with a decent start can extend their earnings and subsequently have more investment, raising the likelihood that they will grow further.

Comparision with Ha and Kiyota (2014)

In the previous section, using the DOP decomposition, we show the negative contribution of entering plants and the positive contribution of exiting plants. Moreover, we find a positive and relatively large improvement in plant productivity while a negative market share reallocation between surviving plants. Our results seem to be in contrast to Ha and Kiyota (2014), who found a positive and relatively large market share reallocation effect, a positive contribution of entrants and a negative contribution of exiters. It should be noted that in this research, we investigate the productivity decomposition in different time spans, while Ha and Kiyota (2014) conducted a one-year interval analysis.[21] However, when we perform a one-year interval analysis (Table 2.C.3 in Appendix 2.C), the difference still remains. The main reason is that we use the productivity of the survivor group as reference, whereas Ha and Kiyota (2014) use the same reference productivity to compute the contribution of all three groups of plants.

Indeed, they use the methodology of Foster et al. (2001) with unweighted mean productivity as reference without any stated reason.[22] It should be noted that in their original paper, Foster et al. (2001) suggest using the weighted average productivity in the first period as reference (Equation (2.24)). The unweighted average might have the advantage of the lowest variance among unbiased estimators of the population mean when the sample observation is drawn with equal probability, like our data. However, recent papers find that differences in the marketshare covariance play an important role in aggregate productivity changes (Collard-Wexler and De Loecker, 2015; Melitz and Polanec, 2015; Nishida, Petrin, Rotemberg, and White, 2017; Olley and Pakes, 1996). Therefore, the market share reallocation should be accounted for when computing aggregate productivity. Recall that according to OP, the (weighted) aggregate productivity is decomposed into the unweighted average plant productivity, $\bar{\phi}_t$, and the covariance term, $\Delta s_{it}\Delta\phi_{it}$. Since plant productivity and market share are positively correlated, the unweighted average productivity is lower than its weighted value. This implies that the contribution of entrants in the research of Ha and Kiyota (2014) might be over-estimated in comparison to the ”weighted” approach of FHK, and it is even higher in comparison than our results: the contribution of exiters in this unweighted FHK decomposition is under-estimated in comparison to the original weighted method, then it becomes more pronounced to the DOP decomposition.

The second reason, that could explain the difference in the market share reallocation, is that, as discussed in Section 2.2, the market share reallocation effect in DOP decomposition reflects the change in the covariance term, the joint cross-sectional distribution of market share and plant productivity. Meanwhile, in the FHK decomposition, the market share reallocation effect consists of the between component and the cross-term. The cross-term reflects the covariance of market share and productivity changes for each plant.

2.3.4 Productivity decomposition and different groups of plants

In the previous section, we divided the entire sample of plants into three groups (survivors, entrants and exiters) to investigate the contribution of each group to aggregate productivity, as well as the within-plant component and market share reallocation among incumbent plants.

[21] If $t = 2$ is the current year, $t = 1$ is the previous one.
[22] see Equation (4) in Ha and Kiyota (2014).

In this section, we include another dimension by differentiating all plants according to three alternative criteria: (i) the two-digit industry, (ii) the ownership status (state-owned, private, foreign-owned) and (iii) the export status. In this way, we aim to investigate the reallocation among groups, the within and between component as well as the contribution of exiters and entrants *in* each group of plants. It should be noted that the definition of exiting and entering plants is slightly different among the criteria. Indeed, in the first criterion (two-digit industry), due to our cleaning method, a plant belongs to *one* industry throughout its existence. Thus, the fact that plants exit industry j in $t = 1$ means that they will not be in the market in $t = 2$; plants entering an industry in $t = 2$ implies that they were not in the market in $t = 1$. In the second criterion, however, plants that exit, for example, the state-owned group in $t = 1$ could either disappear in the market or enter in another groups such as the private group in $t = 2$; by the same logic, plants that exit the export market in $t = 1$ could either exit the market or become a non-exporter.

Because in $t = 1$ and $t = 2$ we always have the same groups for each criterion, that means, there are no "group" dynamics (no group in and no group out), the OP decomposition augmented by Hashiguchi et al. (2015) is sufficient to investigate the reallocation efficiency *among* groups. Nevertheless, since the plant dynamics occur in each group of plants (*i.e.* each industry, or each ownership status group, or each export status group), we apply the augmented dynamic OP (ADOP) decomposition to shed light to the reallocation of plants *in* this group.

Productivity decomposition and two-digit industry

Table 2.6 records different components explaining the aggregate TFP growth (relative to 2000). Firstly, using the ADOP decomposition for 21 industries ($J = 21$), we decompose the aggregate TFP growth (**Column 1**) into the within (**Column 2**) and the between component (**Column 7**) among industries. The unweighted mean changes in the (aggregate) industry productivity (within) plays a large part in the aggregate TFP growth. As the time span increases, this contribution is more remarkable. Interestingly, we observe that the covariance between market share and productivity for industries increases relative to 2000 (the between component is mostly positive). This implies that the market share reallocation among industries improves relative to 2000, in other words, an industry with a higher market share tends to have better productivity.

Secondly, because there are plant dynamics within each industry, we identify surviving, exiting and entering plants for each industry. Then, we use the ADOP decomposition to decompose the within-industry component in **Column 2** into four subcomponents: average of the within-plant component for survivors, average of between-plant component for survivors, average of contribution of exiters, average of contribution of entrants. These subcomponents are recorded in **Columns 3-6**, respectively. Some major observations are drawn from these results: the average within-plant component for survivors explains most of the within-industry component; the average market share reallocation is negative but improves in the last two years; on average, entrants contribute negatively to the change in aggregate TFP (their productivity is lower than that of the incumbents), exiters contribute positively to aggregate TFP growth (plants with lower productivity than that of the survivors tends to exit the market).

In a deeper investigation, Figure 2.3 shows the within- and between-plant components across industries over the 13-year interval from 2000-2013. Nearly all industries (except the basic metals manufacturers) experience an increase in the unweighted mean productivity of surviving

Table 2.6: Aggregate TFP decomposition relative to 2000: two-digit industrial code

Year	Aggregate TFP growth	Within	Within decomposition				Between
	$\Delta\Phi$	$J^{-1}\sum_j \Delta\tilde{\mu}_j$	$J^{-1}\sum_j \Delta\mu_j^S$	$J^{-1}\sum_j \Delta\text{cov}_j^S$	$J^{-1}\sum_j \text{ent}_j$	$J^{-1}\sum_j \text{ext}_j$	Δcov
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
2001	−11.78	−9.66	12.22	−20.08	−6.78	4.99	−2.12
2002	−2.64	−4.13	18.72	−19.89	−8.55	5.59	1.49
2003	−0.29	−1.98	25.55	−23.74	−10.99	7.20	1.69
2004	−2.90	−4.73	28.50	−30.70	−12.28	9.75	1.83
2005	0.05	−3.97	31.16	−34.69	−12.62	12.18	4.01
2006	14.97	11.00	35.20	−20.84	−11.02	7.66	3.98
2007	25.16	21.33	52.65	−23.49	−15.66	7.83	3.84
2008	13.84	7.93	46.05	−29.13	−13.17	4.18	5.91
2009	37.41	35.60	67.16	−20.12	−16.38	4.94	1.82
2010	37.63	38.67	77.45	−27.93	−16.36	5.51	−1.05
2011	43.23	38.92	74.38	−33.30	−8.77	6.62	4.30
2012	50.69	44.50	72.49	−22.97	−12.11	7.09	6.19
2013	75.40	56.43	77.53	−18.96	−10.76	8.62	18.97

Note: We follow the methodology ofhashiguchi2015allocation to decompose aggregate TFP growth, (2)=(3)+(4)+(5)+(6). In this exercise, we rule out the Office, accounting and computing machinery industry because it is represented only since 2005. Although the aggregate TFP growth from 2005 recorded in this table is slightly different from those above, its trend still remains. Entries are TFP growth in log percent.

plants, reflected by positive within-plant components. A better market share reallocation (positive between-plant component) is recorded in some industries with relative higher capital-output ratios in 2000 (Textiles; Medical, precision and optical instruments, watches and clocks; Rubber and plastics products; Food products and beverages).[23]

To better understand how different components of (aggregate) industry productivity growth are related to the industrial characteristics, we regress these components on the value added share by ownership. The reason for concentrating on ownership status is the encouragement policy of the Vietnamese government from the late 1990s in favour of private- and foreign-owned plants. In order to promote economic development, authorities introduced the Law on Private Enterprises and Companies Law in 1990. In 1999, the Law on Enterprises replaced the ealier law and introduce Partnerships in addition to Limited Liability Companies and Shareholding Companies. Moreover, in 1998, the Vietnamese government officially decided to implement the equitisation program following which the state-owned enterprises could be transformed into joint stock companies. Specifically, we perform the regression model as follows:

$$Y_j = \alpha_0 + \beta_1 VAshare_{j0}^d + \mathbf{X}_{j0}\gamma + \varepsilon_j \tag{2.20}$$

where Y_j is the within-, between-plant components, the contribution of entrants and exiters for the 13-year interval from 2000-2013; $VAshare_{j0}^d$ is the value added share by ownership d in the base year 2000, with $d = \{\text{state, private, foreign}\}$. Equation (2.20) is performed on three components for each ownership status d. $\mathbf{X}_{j0}$ is a vector of control at industry-level in 2000, which consists of the capital intensity (measured as the capital-labour ratio, in logs), the industrial tariff rate and the industrial agglomeration degree to control for market structure.[24] The industrial agglomeration degree is the index of industry geographical concentration developed by Ellison and Glaeser (1997) (hereafter, EG index) and defined as

[23] Those industries have capital-output ratio no less than median, see Table 2.C.2 in Appendix 2.C.

[24] We use the Most Favoured Nation (MFN) average tariff from the database of the World Integrated Trade Solution (WITS) developed by the World Bank.

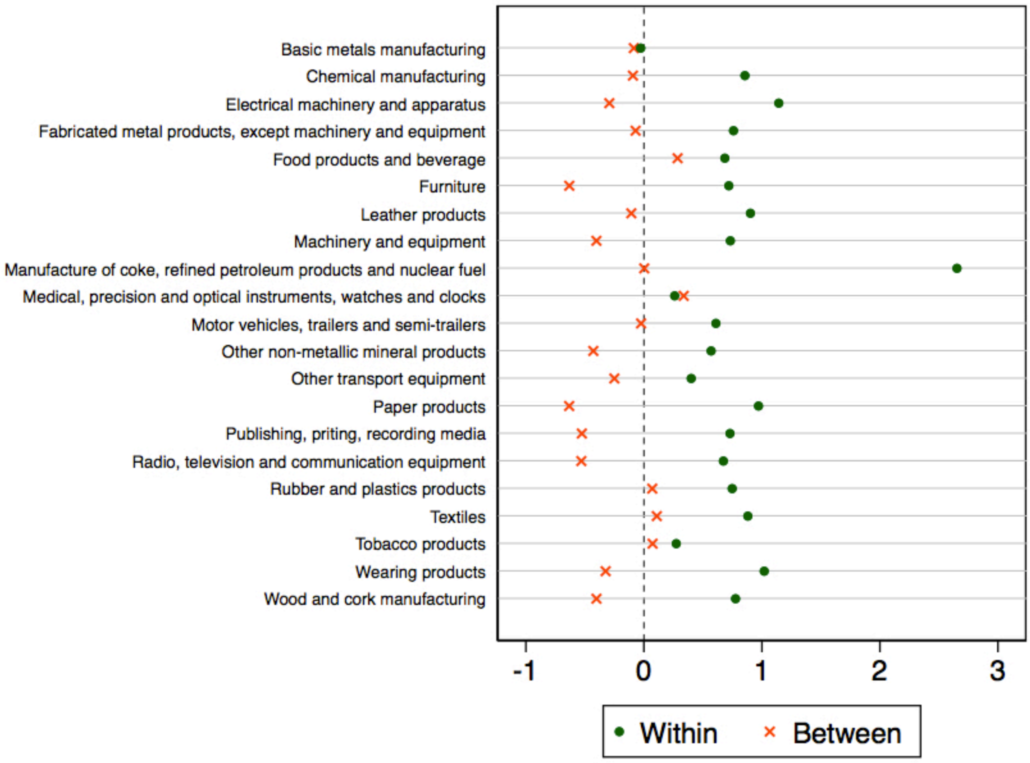


Figure 2.3: Within *vs.* between components for each industry (in logs), 2000-2013

$$EG_i \equiv \frac{G - \left(1 - \sum_j x_j^2\right) H}{\left(1 - \sum_j x_j^2\right)(1 - H)} = \frac{\sum_{j=1}^{M} (s_j - x_j)^2 - \left(1 - \sum_j^M x_j^2\right)^2 \sum_{k=1}^{N} z_k^2}{\left(1 - \sum_j^M x_j^2\right)\left(1 - \sum_k^N z_k^2\right)} \tag{2.21}$$

where EG_i is the index of geographical concentration of industry i, s_j (with $j = 1, ..., M$), denote shares of industry i's employment in each of M geographic areas, x_j are shares of total employment in each of those areas, so, $G \equiv \sum_{j=1}^{M} (s_j - x_j)^2$ is defined as a measure of "raw" geographic concentration of industry i. z_k ,where $k = 1, ..., N$, is the kth plant's share of the industry's employment, $H \equiv \sum_k^N z_k^2$ is the Herfindahl index of the industry plant size distribution. As noted in Ellison and Glaeser (1994), EG_i is the probability that any pair of plants are jointly set in the same location or is the measurement of the advantage in location choice. A higher EG_i indicates that the industry i is more geographically concentrated.

At industry-level, we regress the within, between and net-entry component to the market share for each ownership status (state, private, foreign owners) and other covariates. Table 2.7 reports the regression results. We find that the three components (within, between, net entry) vary differently according to the output share by ownership status. The contribution of within component would increase for industries with a higher share of foreign owners or lower share of private owners. The between-plant component is positively correlated with the output share of foreign-owned plants, meaning that the inter-plant reallocation would be improved in an industry

Table 2.7: Industry productivity decomposition and output share by ownership status

	Within (1)	Between (2)	Net entry (3)	Within (4)	Between (5)	Net entry (6)	Within (7)	Between (8)	Net entry (9)
State owned output share	−0.035	−0.245	−0.102						
	(0.051)	(0.154)	(0.070)						
Private owned output share				−0.190**	−0.160	0.013			
				(0.086)	(0.210)	(0.103)			
Foreign owned output share							0.158**	0.378***	0.108*
							(0.062)	(0.105)	(0.061)
$Tariff_{jt}$	0.002	0.003	0.000	0.001	0.001	−0.001	0.002	0.003	0.000
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Capital-labour ratio (in logs)	−0.012	0.036	0.009	−0.008	0.038	0.008	−0.007	0.047	0.012
	(0.010)	(0.032)	(0.027)	(0.012)	(0.034)	(0.026)	(0.009)	(0.035)	(0.027)
EG index	0.167*	0.338**	−0.189	0.084	0.190	−0.221	0.152*	0.246	−0.227
	(0.096)	(0.147)	(0.252)	(0.082)	(0.140)	(0.262)	(0.076)	(0.154)	(0.226)
Constant	0.082*	−0.115	−0.017	0.130**	−0.067	−0.016	−0.042	−0.406**	−0.099
	(0.041)	(0.162)	(0.143)	(0.053)	(0.168)	(0.170)	(0.069)	(0.148)	(0.160)
Observations	260	260	260	260	260	260	260	260	260
R-squared	0.308	0.221	0.123	0.322	0.214	0.109	0.324	0.231	0.123
Industry fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry-year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

with a growing share of foreign owners. On the plant dynamics, industries with growing foreign owners' output share or decreasing state owners' share have a growing contribution of net entry. Besides, trade liberalisation seems not to affect the contribution of these three components. A better contribution of within component is presented in the industry with a higher degree of geographic concentration.

In sum, at the two-digit industry level, the within-industry component contributes mostly to the change in aggregate TFP, the market share reallocation between industries is better relative to 2000. For each industry, the within subcomponent for surviving plants seems to drive the industry TFP growth. The market share reallocation between incumbents contributes negatively to aggregate TFP growth. However, when accounting for the ownership status, we find that the reallocation could improve in an industry with a higher market share of foreign-owned plants.

Productivity decomposition and ownership status

In the previous subsection, we show how different subcomponents in the (aggregate) industry productivity growth are related to the market share by ownership status. In this subsection, we investigate the productivity decomposition and ownership status in a different scope. We divide the whole sample of plants into three groups (state-, private- and foreign-owned plants), and then examine the reallocation between these groups and the between-plant component inside each group. In this way, we can compare different components among these three groups. As noted before, in each group, exiters indicates plants that exit this group at $t = 1$, they can eventually either exit the market or remain in the market with other ownership status at $t = 2$, the same logic is applied for entrants.

Table 2.8 represents the (aggregate) TFP and the market share of State, Private and Foreign groups for the first and last year of our sample (2000 and 2013). At both points in time, Foreign group has the highest productivity and the Private group the lowest. There is a shift in market share from State to other ownership groups. While the value added share of the State group de-

Table 2.8: Market share and productivity of different groups according to their ownership status

Group	Variable	$t_1 = 2000$	$t_2 = 2013$
	Aggregate TFP (in logs)	2.929	3.680
State	TFP (in logs)	2.852	3.594
	Value added share	*0.441*	*0.0691*
Private	TFP (in logs)	2.206	3.135
	Value added share	*0.122*	*0.283*
Foreign	TFP (in logs)	3.207	3.928
	Value added share	*0.437*	*0.648*

Table 2.9: Within and between ownership groups (relative to 2000)

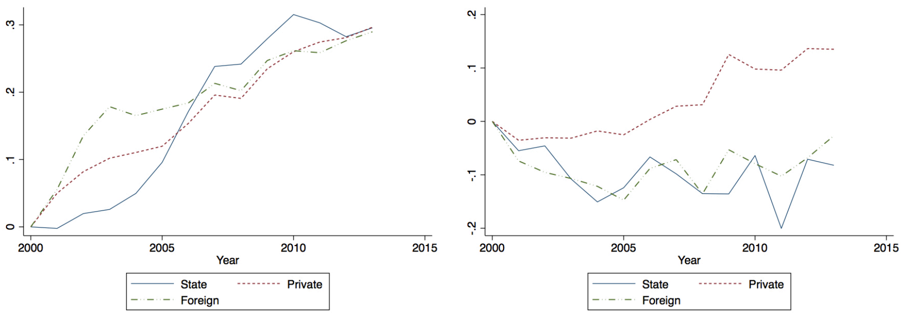
Year	Within	Between	ΔTFP
2001	−9.35	−2.43	−11.78
2002	4.21	−6.85	−2.64
2003	3.83	−4.12	−0.29
2004	3.18	−6.08	−2.90
2005	8.92	−8.89	0.03
2006	26.13	−9.98	16.15
2007	37.38	−11.99	25.39
2008	30.07	−15.33	14.74
2009	48.09	−10.39	37.70
2010	54.86	−16.10	38.76
2011	54.77	−11.46	43.31
2012	63.86	−12.78	51.09
2013	79.71	−4.56	75.16

Note: Entries are in log percent.

creases significantly from 0.441 to 0.069, that of the Private is doubled and the Foreign 1.5 times greater after 13 years. This makes the contribution of the State group in the aggregate productivity much smaller than the Private and Foreign groups in 2013 (it contributes only 6% to aggregate productivity in 2013 against 43% in 2000).

Considering the entire manufacturing sector with these three groups by ownership status, we apply the augmented OP decomposition (AOPD) to investigate the market share reallocation among these groups, since there are no dynamics at group-level. According to Table 2.9, we find that the within-group component rises significantly when the time spans increases. It dominates the aggregate TFP growth and offsets the negative contribution of the between-group.

For each group, we use the ADOP decomposition to compare the components of these three groups. Figure 2.4 shows the within- and between-plant components, the contribution of entrants and exiters for each ownership group. Following Harrison, Martin, and Nataraj (2013); Pavcnik (2002), we normalize productivity values to 0 in the base year 2000. Hence, the contribution represented in this figure can be interpreted as change since 2000. In all groups of plants by own-



(a) Within component (b) Between component

Figure 2.4: Different components for each ownership group (relative to 2000)

ership status, we observe an improvement in productivity within plants for each group. Moreover, these within-plant components contribute most to the change in productivity (Table 2.C.4 in Appendix 2.C). In the second half of the period, the within component for State is higher than that for other groups. On the reallocation efficiency, Private is the group with the best market share reallocation, its between component is positive and improve gradually when the time span increases. By contrast, the market share reallocation for either State or Foreign is negative (despite the slight raising in the late period for Foreign). This fact suggests that the encouragement policies in favour of private plants in the late of 1990s have worked.

Productivity decomposition and export status

In this section, we address another aspect of aggregate productivity growth. One of the widely publicised recent debates focuses on the relationship between international trade and long run output and productivity progress. In this line, this subsection aims to compare the contribution of reallocation in productivity change between two groups: exporters *vs.* non-exporters. We apply the same methodology as in the previous subsection. As before, exiters mean that plants exit the export market in $t = 1$, so they could either have left the market of still remain as non-exporters in $t = 2$; entrants in $t = 2$ could be either new plants entering the market or plants switching from non-exporters to exporters. We consider two periods 2002-2004 and 2010-2013 when the export status is mostly well reported in the census.

Table 2.10: Aggregate productivity decomposition and export status. Within- *vs.* between-group

Period	Within	Between	Aggregate TFP growth
2002-2004	−15.97	15.71	−0.27
2010-2013	6.97	29.42	36.40

Note: The augmented OP decomposition is applied for two groups: exporters and non-exporters. Entries are in log percent.

As shown in Table 2.10, we observe a positive market share reallocation between two groups exporters and non-exporters. During 2002-2004, the high market share reallocation offsets the neg-

Table 2.11: Within-group decomposition and export status.

Period	Export				Non-export			
	Within	Between	Entrants	Exiters	Within	Between	Entrants	Exiters
2002-2004	14.99	−10.10	−2.98	1.81	5.04	−23.96	−4.42	3.65
2010-2013	8.07	13.54	−3.88	1.23	0.23	−13.26	−3.91	4.96

Note: The ADOP decomposition is applied for two groups: exporters and non-exporters. These contributions sum to the within component in Table 2.10. Entries are in log percent.

ative within-group component. In the second period, the between component contributes mostly to the change in productivity, it contributes 80% to the TFP growth, while the contribution of the within component is moderate. This finding supports the fact that even when the within-group component is moderate, a better reallocation between exporters and non-exporters could enhance productivity growth.

We then decompose the within-group component to four components for each group of plants (Table 2.11). The within- and between-plant components for exporters are relatively higher than non-exporters for both periods. Therefore, exporting plants have better improvement in productivity within plants and a better market share reallocation than non-exporting ones. While the within-plant component decreases for either exporters or non-exporters, the market share reallocation improves from one period to another. Especially it becomes positive in 2010-2013 for exporting plants, meaning that exporters with high productivity have higher market share.

2.4 Conclusion

In this chapter, we focus on the different sources of aggregate productivity growth in the Vietnamese manufacturing industries and pointed out the differences in the results between three decomposition methods. Our first finding is that aggregate productivity growth is mostly due to the within-plant effect. The higher contribution of the within component is consistent even when we perform various exercises by dividing the whole sample into different groups of plants according to their two-digit industries, their ownership status. However, we find that the market share reallocation between exporters and non-exporters improves and plays a large part in the aggregate productivity growth. The between-plant component contributes mostly to the change in aggregate productivity for exporters when we consider the difference between export and non-export groups.

We do not observe a better market share reallocation (the between component) either across survivors' plant in manufacturing or across survivors' plant in each of 2-digit industries, but at least there is improvement over time. Moreover, the between-plant component could improve when the industries either increase their foreign owners' share or decrease their state owners' share. A better market share reallocation is particularly recorded for private ownership plants and for exporting plants during the period 2010-2013.

While investigating plant dynamics, we find that the contribution of exiters is positive because plants with relatively lower productivity than incumbents tend to exit the market. The productivity of new plants is lower than that of survivors in the year they enter the market. Therefore, they contribute negatively to aggregate productivity growth. Besides, we find that the group of plants surviving throughout the 14 years from 2000 to 2013 have the highest productivity. Thus, new entrants could not catch up with this group of plants. But interestingly, they can be more productive than dynamic incumbents 3 to 5 years after entering the market. This is the evidence that new plants can learn from market functioning in the sense of Ericson and Pakes (1995). Exiting plants are always less productive than either incumbents or plants having been in the market for a few years. The productivity difference between these groups is greater when low productivity plants are closer to their exit year. This means that a plant's productivity life-cycle behaves as we would expect.

Appendix

2.A Other decomposition methods

In this section, we review briefly two other decomposition methods used for comparison to the MP in Section 2.3.3: that of Foster et al. (2001) and Griliches and Regev (1995).

As before, let's denote Φ_t aggregate productivity at t, that is the weighted-average plant productivities ϕ_{it} with s_{it} plant market share ($\sum_i s_{it} = 1$):

$$\Phi_t = \sum_i s_{it}\phi_{it} \tag{2.22}$$

As shown by Melitz and Polanec (2015), those two methods rely on the same reference productivity level in both periods, Φ_{ref} to evaluate the contributions of different groups of plants. Accordingly, aggregate productivity change, $\Delta\Phi$ could be written as follows:

$$\begin{aligned}
\Delta\Phi &= \Phi_2 - \Phi_1 \\
&= \sum_i (s_{i2}\phi_{i2} - s_{i1}\phi_{i1}) + \sum_i (s_{i2} - s_{i1})\,\Phi_{ref} \\
&= \sum_i \left[s_{i2}\left(\phi_{i2} - \Phi_{ref}\right) - s_{i1}\left(\phi_{i1} - \Phi_{ref}\right)\right]
\end{aligned}$$

Griliches and Regev (1995) consider the average aggregate productivity between the two periods as the reference value, that means, $\bar{\Phi} = \frac{\Phi_1+\Phi_2}{2}$. The aggregate productivity change is then decomposed as:

$$\begin{aligned}
\Delta\Phi &= \sum_{i\in S}\left[s_{i2}\left(\phi_{i2} - \bar{\Phi}\right) - s_{i1}\left(\phi_{i1} - \bar{\Phi}\right)\right] + \sum_{i\in E} s_{i2}\left(\phi_{i2} - \bar{\Phi}\right) - \sum_{i\in X} s_{i1}\left(\phi_{i1} - \bar{\Phi}\right) \\
&= \sum_{i\in S}\bar{s}_i\Delta\phi_i + \sum_{i\in S}\Delta s_i\left(\bar{\phi}_i - \bar{\Phi}\right) + \sum_{i\in E} s_{i2}\left(\phi_{i2} - \bar{\Phi}\right) - \sum_{i\in X} s_{i1}\left(\phi_{i1} - \bar{\Phi}\right) \qquad (2.23)
\end{aligned}$$

where $\bar{s}_i$ and $\bar{\phi}_i$ are average market share of plant i and average productivity of plant i between two periods, respectively ($\bar{s}_i = \frac{s_1+s_2}{2}, \bar{\phi}_i = \frac{\phi_1+\phi_2}{2}$); Δs_i and $\Delta\phi_i$ denote the change in market share and plant productivity, respectively, $\Delta s_i = s_{i2} - s_{i1}, \Delta\phi_i = \phi_{i2} - \phi_{i1}$. The first line divides aggregate productivity growth into the contribution of surviving, entering and exiting plants. In the second line, aggregate productivity growth is decomposed into four components: the first two terms indicate the contribution of surviving plants, the third term is the contribution of new entrants and the last is that of exiting plants. The first component is interpreted as a within effect, that is the sum of change in incumbent productivity weighted by the average market shares. The second component (between effect) is the change in plant market share indexed by the deviation of the average plant productivity from the average aggregate productivity. The last two components compare plant-level productivity of new entrants (the third component) and exiters (the fourth component) to the reference productivity level. If the new entrants productivity is higher than the average aggregate productivity, this group of plants contributes positively to aggregate productivity growth. The contribution of exiters is positive only if exiting plant productivity is lower than the average aggregate productivity.

Different from Griliches and Regev (1995), Foster et al. (2001) use the aggregate productivity of the intial period, Φ_1, as the reference productivity level. Aggregate productivity decomposition is given by:

$$\begin{aligned}\Delta\Phi &= \sum_{i\in S}\left[s_{i2}\left(\phi_{i2}-\Phi_1\right)-s_{i1}\left(\phi_{i1}-\Phi_1\right)\right]+\sum_{i\in E}s_{i2}\left(\phi_{i2}-\Phi_1\right)-\sum_{i\in X}s_{i1}\left(\phi_{i1}-\Phi_1\right)\\ &=\sum_{i\in S}s_1\Delta\phi_i+\sum_{i\in S}\Delta s_i\left(\phi_{i1}-\Phi_1\right)+\sum_{i\in S}\Delta s_i\Delta\phi_i+\sum_{i\in E}s_{i2}\left(\phi_{i2}-\Phi_1\right)-\sum_{i\in X}s_{i1}\left(\phi_{i1}-\Phi_1\right)\end{aligned} \tag{2.24}$$

In the first line, aggregate productivity growth is decomposed into the contribution of suriving, entering and exiting plants. In the second line, the contribution of survivors is separated into within-plant effect (the first component), between-plant effect (the second) and the cross term (the third). The within-plant component is the part of change in aggregate productivity that is due to the change in survivors' productivity while the market share remains in $t = 1$. The between-plant term indicates the sum of change in survivors' market share weighted by the deviation of plant-level productivity of survivors from aggregate productivity in the initial period $t = 1$. The change in plant's market share contributes positively to the between component only if its productivity is higher than aggregate productivity in the initial period. The cross term is the joint change of survivors' market share and their productivity. As in the method of Griliches and Regev (1995), the contribution of entering and exiting plants is determined by comparing their productivity with the reference productivity level Φ_1. The group of entrants contributes positively to the aggregate productivity growth if new entrants' productivity is higher than aggregate productivity in $t = 1$. Similarly, the contribution of exiting plants is positive only when plants with lower productivity than Φ_1 exit the market.

2.B Production function estimation

To estimate the total factor productivity (TFP), we follow the methodology suggested by Levinsohn and Petrin (2003) (herafter, LP). This methodology relies on the restricted profit value added production function, that provides, as argued by Petrin and Levinsohn (2012), a much more direct welfare interpretation, and therefore can be useful in further investigation. [25]

[25] see Gandhi, Navarro, Rivers, et al. (2017) for further discussion on different specifications of production function.

The value added Cobb-Douglas production function in logs is given by:

$$y_{it} = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \omega_{it} + \epsilon_{it} \tag{2.25}$$

where y_{it} is value added, l_{it}, k_{it} denotes labour, capital. Two unobservable terms to the econometricians are ω_{it} and ϵ_{it}. The latter, ϵ_{it}, represents measurement error in the output and is unobserved by firms when they make input decisions, so it is uncorrelated with k_{it} and l_{it}; the former, ω_{it}, called the productivity shock, is predictable by firms during their decision making process and hence could be correlated with input variables. The correlation between ω_{it} and a firm's input choices raises the endogeneity problem when estimating the production function. In that case, estimation by OLS will bias the coefficients of the observed inputs. Therefore, the methodology attempts to eliminate the endogeneity issue by measuring ω_{it} through an equation.

Firstly, following Olley and Pakes (1996), the authors assume that the productivity ω_{it} follows a first-order Markov process, *i.e.,*

$$p(\omega_{it+1}|I_{it}) = p(\omega_{it+1}|\omega_{it}) \tag{2.26}$$

where I_{it} is firm i's information set at t. This assumption states that the future productivity, ω_{it+1} is expected to be based on the current set of information, I_{it}, that consists of current and past productivity and not the future one ω_{it+1}. This assumption also implies that a firm's decisions at t are uncorrelated with the unanticipated innovation in ω_{it+1} between t and $t+1$, denoted ξ_{it+1} with $\xi_{it+1} = \omega_{it+1} - E[\omega_{it+1}|I_{it}] = \omega_{it+1} - E[\omega_{it+1}|\omega_{it}]$ and $E[\xi_{it+1}|I_{it} = 0]$.

Then, capital is assumed to be accumulated following the law of motion:

$$k_{it} = (1-\delta)\, k_{it-1} + i_{it-1} \tag{2.27}$$

where i_{it-1} is firm i's investment at $t-1$. The law of motion implies that firm's capital stock at t is actually decided at $t-1$, so belongs to I_{it-1}. This implication is crucial to identify the output elasticity of capital, β_k.

In the LP model, the non-dynamic variables l_{it} and m_{it} are assumed to be chosen simultaneously at time t, after firms have information about ω_{it}. Therefore, l_{it} might be correlated with the current value of ω_{it}. So, ω_{it} needs to be controlled to solve the endogeneity problem with respect to l_{it}.

Thirdly, to deal with the endogeneity problem, LP identify the control function in which ω_{it} is the only econometric unobservable and it must be strictly monotonic in ω_{it}. Different from OP, LP propose the intermediate input, m_{it}, as proxy for ω_{it}. This is because for developing countries, the OP's proxy variable, investment, may not be available for all observations, it can be 0 for many plants in some years. [26] The intermediate input demand function is given by $m_{it} = f_t^*(k_{it}, \omega_{it})$. [27]

The monotonicity assumption allows the control function to be inverted to obtain ω_{it} as a

[26]In our data, only 15% of total firms per year have the information about investment. The positive intermediate inputs are reported for over 80% of firm per year.

[27]Note that l_{it} does not appear in the intermediate input demand function. This implies that l_{it} does not affect a firm's optimal choice of m_{it}. It could be said that l_{it} and m_{it} are chosen at the same time. See Ackerberg, Caves, and Frazer (2015) for further discussion.

function of intermediate input and capital:

$$\omega_{it} = f^{*-1}(k_{it}, m_{it}) \tag{2.28}$$

The equation above implies that ω_{it} could "be measured" through the variables observed by the econometricians. Although the functional form is unknown, f^{*-1} could be approximated by the polynomial in k_{it} and m_{it}. Substituting Equation (2.28) in Equation (2.25), we have:

$$y_{it} = \beta_l l_{it} + \tau_t(k_{it}, m_{it}) + \epsilon_{it} \tag{2.29}$$

where

$$\begin{aligned} \tau_t(k_{it}, m_{it}) &= \beta_0 + \beta_k k_{it} + \omega_{it} \\ &= \beta_0 + \beta_k k_{it} + f^{*-1}(k_{it}, m_{it}) \end{aligned} \tag{2.30}$$

To estimate the labour coefficient, LP suggest using the semiparametric estimation of the linear part of a regression model proposed by Pagan and Ullah (1999). We will describe this methodology while representing the procedure to obtain the coefficients of interest. [28] This step is called the first step.

The second step involves estimating the output elasticity of capital, β_k, from Equation (2.30). We rely on the timing of a firm's input choice in Equation (2.27) to identify the moment condition and then estimate β. Recalling that this approach supposes that ω_{it} follows the first order of Markov process, i.e., $\omega_{it} = E[\omega_{it}|I_{it-1}] + \xi_{it}$ where $E[\xi_{it}|I_{it-1}] = 0$. So, Equation (2.30) could be written as:

$$\tau_t = \beta_0 + \beta_k k_{it} + E[\omega_{it}|I_{it-1}] + \xi_{it} \tag{2.31}$$

As mentionned above, since k_{it} is actually decided at $t-1$, $k_{it} \in I_{it-1}$, this implies that $E[\xi_{it}|k_{it}] = 0$. We obtain the uncorrelated relation between ξ_{it} and k_{it} that comes from the mean independence:

$$E[\xi_{it} k_{it}] = 0 \tag{2.32}$$

This is the moment that we use to estimate the capital coefficient. As suggested by LP, we can also include an over-identification condition, that is: $E[\xi_{it} l_{it-1}] = 0$. Indeed, as mentioned above, since l_{it} is chosen at time t, l_{it-1} belongs to I_{it-1}. Therefore, the moment condition we use in the second stage is given by:

$$E\left(\xi_{it}(\beta)\begin{pmatrix} l_{it-1} \\ k_{it} \end{pmatrix}\right) = 0 \tag{2.33}$$

So, the procedure to obtain input coefficients can be executed as follows. In the first stage, following Pagan and Ullah (1999), we regress y_{it} and l_{it} on k_{it} and m_{it} to obtain the new regressors net of intermediate input and capital. That means, $y(m_t, k_t) = y_{it} - E(y_{it}|m_{it}, k_{it})$ and $l(m_t, k_t) =$

[28] One could regress y_{it} on l_{it} and the high order polynomial in (k_{it}, m_{it}) to obtain the estimator of β_l.

$l_{it} - E(l_{it}|m_{it}, k_{it})$. $\hat{\beta}_l$ is obtained simply by regressing $y(m_t, k_t)$ on $l(m_t, k_t)$. From this stage, we also have the estimate of τ_t,$\hat{\tau}_t$.

We then perform the second stage in a GMM context. We choose the candidate values of β, denoted as (β_0^*, β_k^*) or β^*. From Equation (2.30), we have the estimate of productivity ω_{it} given β^*:

$$\omega_{it}^*(\beta^*) = \hat{\tau}_{it} - \beta_0^* + \beta_k^* k_{it} \tag{2.34}$$

Since ω follows the first order Markov process, it implies that:

$$\omega_{it} = g(\omega_{it-1}) + \xi_{it} \tag{2.35}$$

Regressing $\omega_{it}^*(\beta^*)$ on the 3rd-order of its lags, ω_{it-1}^*, we obtain $\xi_{it}(\beta^*)$. Applying the moment conditions in Equation (2.33) and using the standard GMM techniques and the block bootstrapping for the standard errors, we obtain the remaining estimated parameters of the production function.

2.C Tables

Table 2.C.1: Joint distribution of market share and productivity, 2000 as base year

Year	cov_{S1}	cov_{S2}	Reallocation
	TFP - Value-added share as weight		
2001	1.29	1.08	−20.67
2002	1.26	1.09	−17.01
2003	1.26	1.04	−21.41
2004	1.26	0.97	−29.69
2005	1.27	0.96	−30.56
2006	1.19	1.05	−13.77
2007	1.15	1.02	−13.02
2008	1.15	0.88	−26.83
2009	1.12	1.07	−4.55
2010	1.08	1.01	−7.14
2011	1.07	0.90	−17.15
2012	1.06	1.04	−2.19
2013	1.07	1.11	4.56
	Labour productivity - Employment share as weight		
2001	0.29	0.24	−5.42
2002	0.27	0.18	−9.36
2003	0.26	0.18	−7.48
2004	0.26	0.19	−6.23
2005	0.25	0.16	−9.90
2006	0.22	0.18	−3.98
2007	0.20	0.12	−8.10
2008	0.19	0.08	−11.05
2009	0.15	0.21	5.34
2010	0.13	0.15	1.47
2011	0.14	0.17	2.93
2012	0.11	0.22	10.85
2013	0.10	0.19	9.09

Note: cov_{S1} and cov_{S2} indicate the joint distribution of market share and productivity of surviving plants at $t = 1$ (2000) and $t = 2$ respectively. $cov_{St} = \sum_{i \in S}(s_{it} - \bar{s}_t)(\phi_{it} - \bar{\phi}_t)$, where $\bar{\phi}_t$ is the unweighted average productivity and $\bar{s}_t$ is the average market share. $Reallocation = cov_{S2} - cov_{S1}$.

Table 2.C.2: Capital-output ratio, two-digit industries, 2000

Industry code	Industry name	Capital-output ratio
26	Other non-metallic mineral products	1.412
28	Fabricated metal products, except machinery and equipment	1.005
34	Motor vehicles, trailers and semi-trailers	0.973
23	Manufacture of coke, refined petroleum products and nuclear fuel	0.954
17	Textiles	0.947
29	Machinery and equipment	0.863
32	Radio, television and communication equipment	0.846
25	Rubber and plastics products	0.812
15	Food products and beverage	0.783
31	Electrical machinery and apparatus	0.734
33	Medical, precision and optical instruments, watches and clocks	**0.699**
19	Leather products	0.598
16	Tobacco products	0.377
24	Chemical manufacturing	0.336
21	Paper products	0.311
20	Wood and cork manufacturing	0.266
22	Publishing, printing, recording media	0.182
18	Wearing products	0.179
36	Furniture	0.090
35	Other transport equipment	0.049
27	Basic metals manufacturing	−0.027

Note: This ratio equals capital/value added. Median value is in bold.

Table 2.C.3: Decomposition of aggregate productivity change (year-by-year)

Year	Within	Between	Surviving plants	plant dynamics Entering plants	 Exiting plants	Aggregate TFP growth
	Labour productivity (in log percent) - Employment share as weight					
2001	13.88	−5.42	8.46	−12.08	2.27	−1.34
2002	16.33	−7.44	8.89	−10.17	5.04	3.76
2003	17.64	−3.13	14.51	−9.83	3.31	7.99
2004	14.25	−4.02	10.23	−9.10	3.69	4.82
2005	13.72	−4.63	9.09	−7.17	2.45	4.37
2006	15.96	−1.68	14.27	−4.87	2.18	11.58
2007	19.11	−2.68	16.42	−7.83	2.45	11.05
2008	4.12	−4.92	−0.80	−9.96	2.25	−8.51
2009	14.26	8.55	22.81	−5.93	2.78	19.66
2010	12.99	−0.95	12.04	−4.75	3.05	10.34
2011	1.42	6.25	7.66	−8.21	2.48	1.93
2012	11.25	3.92	15.17	−4.64	2.36	12.90
2013	6.55	3.74	10.29	−4.60	3.86	9.55
	TFP (in log percent) - Value-added share as weight					
2001	12.54	−26.47	−13.93	−4.99	3.68	−15.24
2002	15.74	−9.54	6.20	−4.49	2.64	4.34
2003	16.34	0.88	17.22	−2.24	2.17	17.15
2004	13.72	−12.34	1.38	−2.60	1.43	0.21
2005	12.66	−8.43	4.23	−2.97	−0.23	1.03
2006	19.34	−0.23	19.11	−3.23	1.76	17.64
2007	18.41	−7.94	10.48	−3.61	1.74	8.61
2008	3.72	−11.57	−7.85	−3.86	0.42	−11.29
2009	12.97	6.38	19.35	−3.02	0.29	16.62
2010	11.19	−4.40	6.79	−4.89	2.19	4.10
2011	6.43	2.74	9.17	−4.50	2.46	7.12
2012	3.87	5.72	9.60	−3.24	1.93	8.29
2013	5.33	9.40	14.73	−2.71	1.36	13.39

Table 2.C.4: Within-group decomposition (relative to 2000)

Year	State				Private				Foreign			
	Within	Between	Entrants	Exiters	Within	Between	Entrants	Exiters	Within	Between	Entrants	Exiters
2001	−0.23	−5.49	−1.40	1.61	4.99	−3.54	−2.60	0.58	5.57	−7.40	−2.45	1.00
2002	1.98	−4.58	−2.23	1.93	8.20	−3.07	−1.01	0.89	13.53	−9.54	−3.19	1.31
2003	2.59	−10.79	−1.21	2.55	10.20	−3.14	−1.81	1.20	17.85	−10.69	−4.57	1.64
2004	4.97	−15.08	−1.49	3.42	11.05	−1.78	−1.60	1.30	16.52	−12.15	−3.78	1.79
2005	9.58	−12.42	−3.00	5.14	11.97	−2.51	−2.82	2.10	17.49	−14.74	−3.71	1.83
2006	17.16	−6.66	−4.03	0.81	15.38	0.37	−4.34	2.53	18.41	−8.80	−6.34	1.65
2007	23.82	−9.82	−5.27	1.29	19.58	2.84	−6.07	3.10	21.32	−7.15	−6.90	0.63
2008	24.17	−13.53	−3.50	1.86	19.08	3.13	−6.48	2.60	20.22	−13.56	−4.14	0.21
2009	27.93	−13.57	−4.78	2.30	23.48	12.54	−14.67	3.01	24.71	−5.31	−7.82	0.27
2010	31.53	−6.39	−6.78	0.43	26.01	9.80	−13.68	3.60	26.16	−7.92	−8.44	0.55
2011	30.30	−20.03	3.89	0.92	27.45	9.60	−13.56	4.02	25.86	−10.25	−4.41	0.98
2012	28.24	−7.09	−1.84	1.13	28.10	13.63	−16.09	3.86	27.64	−6.82	−8.14	1.25
2013	29.56	−8.22	1.11	2.29	29.64	13.51	−16.08	3.88	28.97	−2.68	−3.82	1.55

Note: These contributions sum to the within component in Table 2.9. Entries are in log percent.

Chapter 3

Markups and Trade Liberalisation: the Case of Vietnamese Manufacturing Sector

3.1 Introduction

Trying to explain the cross-country differences in total factor productivity (TFP) and differences in GDP per capita across countries, the traditional theory of growth considers sluggish acceptance of new technology and the uneconomical use of technology of a country as the main cause of these differences. Another strand of literature focuses on the allocation of resources to heterogeneous production units in an economy. Temporarily ignoring the technology problems, they argue that misallocation of inputs across production units, or firms, can slowdown aggregate TFP growth. In poorer countries, a low aggregate TFP can be the result of several types of market frictions which prevent marginal revenue products from being equalised across firms. These frictions could be various (policy distortions, adjustment costs, credit market frictions, trade barriers, etc.) but when these constraints are binding, resources are prevented from channelling from low to high productivity firms. Instead of considering restriction of input choices as a source of misallocation, new studies move toward the investigation of imperfections in the output markets. One of the focal points is the role of markup variation as a source of misallocation. Peters (2013) proposes a model to investigate the relationship between distribution of markups and misallocation. If firms have monopoly power and set firm-specific markups, generally considered as the amount added to the cost price of goods, the imperfect market can create misallocation. Epifani and Gancia (2011) point out that there exists an inter-sectoral misallocation in the presence of heterogeneity in markups regardless of the reasons: trade, regulation or difference in ability to collaborate across sectors. The authors also argue that there is under-production in terms of firm output or product variety in industries with above-average markups; however, an appropriate intervention can improve the equilibrium. Such an intervention which has received a lot of attention recently is trade liberalisation. Indeed, international trade liberalisation makes producers face greater competition, therefore it can reduce markup dispersion, mitigate misallocation and increase aggregate productivity (Edmond, Midrigan, and Xu, 2015). However, it seems to be the case that reducing trade barriers is important for markups, but the effect on aggregate TFP is rather small (Peters, 2013).

In this light, instead of studying directly the effect of trade liberalisation on misallocation, our research aims to examine empirically the effect of trade liberalisation on industries' markup dispersion, one of the sources of inefficient allocation. In addition, we attempt to explore the role of trade on firm pricing by comparing the markups between exporters and non-exporters. Before doing this exercise, we investigate the evolution and characteristics of firm markups and their distribution to overview markups in Vietnamese manufacturing. For these purposes, we use rich plant-level data in manufacturing in Vietnam for the period 2000-2013. To measure the plant markups, defined as price-marginal cost ratio, we follow the methodology of De Loecker and Warzynski (2012) which is based on the production function framework. Accordingly, two main elements need to be identified in order to estimate markups: the expenditure share of labour in total sales, which is directly observed from the data, and the output elasticity of labour input, which is obtained when estimating the production function. Different from the previous one suggested by Hall (1988) that is mostly used on industry-level data, this methodology is more accurate when applied to micro data. Indeed, it is related to the approach of Olley and Pakes (1996) on estimating production function with firm-level data, that are more structural in nature in dealing with the well-known endogeneity problem (Ackerberg et al., 2015).[29] Besides, this markup measurement technique is more appropriate when examining the impact of variations in the business environment, such as trade liberalisation as in our case, on markups. Since trade liberalisation affects plant productivity, its impact on markup change will be biased when not controlling for the productivity shock (De Loecker and Warzynski, 2012). To estimate the production function, we apply the approach proposed by Ackerberg et al. (2015), who augment the Olley and Pakes (1996)'s and Levinsohn and Petrin (2003)'s approaches, to the value added production function in both the Cobb-Douglas specification in our main results and the more flexible one, *i.e.,* translog model for the robustness check.

Several key findings merge. First, investigating the markup characteristics, we find that the aggregate markup in Vietnamese manufacturing decreases gradually over time. This trajectory even holds when we introduce different input and output weights, that are used to account for the plant size, to compute the aggregate markup. More interestingly, we observe that the output weighted markups are always higher than the input weighted markups. This implies that plants use few inputs but still generate more sales. Together with the diminishing aggregate markup, plant markups were less dispersed in 2013 in comparison to the beginning of the period 2000. The downward trend of the aggregate markup suggests higher competition between plants, that make plants change their prices to better reflect their costs, therefore benefiting the consumers. At the same time, the lower markup dispersion at the end of the period implies a bettering market share reallocation between Vietnamese manufacturing plants. These observations lead us to study further the reason behind the evolution of the aggregate markup and the markup dispersion.

Second, we investigate the source of the reducing aggregate markup by decomposing it at industry-level and at plant-level. We follow De Loecker, Eeckhout, and Unger (2020) in decomposing the year-by-year change in aggregate markup into the change of markup within industry, the change in the composition between industries and the joint change in markup and the industry composition. We find that the change within industry contributes mostly to the fall of the ag-

[29] Two other "classic" methodologies solving the endogeneity problem that firm's optimal choice of inputs are correlated with productivity are fixed-effects and instrumental variables estimation approaches. They might not, however, work well in practice, see Ackerberg et al. (2015) for further discussion.

gregate markup. The contribution of the joint effect is relatively minor, but it is mostly positive. This implies the positive correlation between industries' markup and their market share (measured by value added share). We then base on the insight of Melitz and Polanec (2015) to re-perform this exercise at plant-level. The difference between this decomposition methodology and that proposed by De Loecker et al. (2020) derives from the markup of reference used to compare with the markups of surviving, entering and exiting plants. Indeed, in the De Loecker et al. (2020)'s work, the aggregate markups in the previous period is used as the benchmark. In our research, we use the markup of incumbents as the benchmark. We divide plants into three groups: continuers, entrants and exiters, then compute the (weighted) aggregate markup for each group. The change in the aggregate markup is computed through the difference between the groups' markups and the average markups of continuers. The markup change of the group of continuers is then decomposed by the change in markup within plants and between plants (reallocation term), as the insight of Olley and Pakes (1996). An advantage of the reallocation term of the Olley and Pakes (1996)'s type is that it helps better understanding whether the reallocation increases the aggregate markup over time. As in the case of industry-level, we find that the decline of markups within plants drives the aggregate markup change. Moreover, the positive reallocation term (even it is small) implies a positive correlation between market share and markups. In other words, the reallocation of activities (value added) could increase markup.

Third, to analyse the role of trade on plant pricing, we compare markups between exporters and non-exporters. To do so, we examine the causality between markups and export status. We aim to answer the question whether plants with higher markups seek to become exporters, or exporting activities help plants raise their markups, or both. Using the model of decision to export by the plant, as Bernard and Jensen (1999), we confirm that plants' with prior success (larger size, higher productivity, wages and markups) are likely to choose to export in the case of Vietnamese manufacturer. Given this causality, we use IV regression to compare markups between exporters and non-exporters. We confirm the finding of De Loecker and Warzynski (2012) that exporters charge higher markup than non-exporters. Besides, after entering into export markets, plants continue to charge higher markups. This implies that higher markups for exporting plants is the result of not only the selection process but also the learning-by-exporting process.

Fourth, since the period from 2000 to 2013 saw an important trade reform: Vietnam's accessed into the WTO in early 2007, this study, strongly related to Lu and Yu (2015), sheds light on how trade liberalisation could affect markup dispersion for Vietnamese manufacturing industries by using the difference-in-differences methodology. We define industry at the two-digit of Vietnam Standard Industry Classification 1993. For each two-digit industry, the markup dispersion in a given year is calculated by Theil index. In our difference-in-differences design, we use the Most Favoured Nation (MFN) ad-valorem average tariff in 2006, the year before WTO's accession, as the measure of industry's exposure to trade liberalisation. We classify industries into two groups: (i) the treatment group consists of industries which experienced high tariff before WTO's accession, then a larger tariff reduction after WTO's accession; (ii) industries in the control groups were more open before 2007, so their tariff reduction is less significant after 2007. Then we compare the markup dispersion between these two groups before and after 2007. Since the period covers the financial crisis in 2008, we also check whether there is a joint impact of WTO's accession and the crisis on markup dispersion. We find evidence that trade liberalisation reduces industry markup dispersion. An important assumption for the validity of the difference-in-differences technique is the parallel trends. This means, without intervention in 2007 (or before 2007), markup dispersion trends of the treatment and the control group must be identical. Therefore, we perform various falsification tests to check validity. We also use different measurements of markup dispersion and industry's exposure to trade liberalisation for the robustness checks. Moreover, to examine the heterogeneous effect of trade liberalisation on markup dispersion according to different industry characteristics (such as industry labour intensity, share of number of state-owned plants within industry, share of number of foreign-owned plants and share of number of inland plants), we use the triple differences rather than divide our sample and then estimate separately the effect for each sub-sample as Lu and Yu (2015) because it is straightforward for such comparison. For example, dividing the sample using state-owned status and separate re-estimation might render the difference-in-differences estimates inaccurate because it ignores the interaction between plants, or creates a selection problem. We find that trade liberalisation reduces markup dispersion more in the industries with relatively more state-owned plants.

Our findings relate to several papers studying the link between trade and markups using plant-level data. Regarding the relationship between markups and plant size, Edmond et al. (2015) propose a model of imperfect competition and show that plant markups are positively correlated with plant's share of sectoral revenue. The positive correlation between markups and plant size is supported by several empirical studies. In their paper, the authors point out that larger producers set higher markups for the Taiwanese manufacturing industry. Atkin, Chaudhry, Chaudhry, Khandelwal, and Verhoogen (2015) observe a positive correlation between plant markups and employment in the case of Pakistan. De Loecker et al. (2020) confirm the positive relationship between markups and various plant size measurements (employment, cost of goods sold, sales) in the US economy.

Few studies investigate the source of change in the aggregate markup over time. Using the plant-level data of the US, De Loecker et al. (2020) decompose the economy-wide markup change at the industry- and plant-level. They find that the aggregate markup change is driven mostly by the change of markup at the industry-level (two thirds of the change in the aggregate markup) rather than the change in the composition of the industries (one third), measured as the sum of change in industry market share weighted by industry productivity in the initial period. Applying the decomposition method for a developing country, we find that the decrease in aggregate markup of manufacturing industry is due mostly to the change of markup within the industry (within). While the change of the within term slows down the aggregate markups, the change in the composition of the industries (between) slightly increases the aggregate markup. The latter implies the positive correlation between the industries' markup and their market share, as in the theoretical literature mentioned above, and indicates that the reallocation between industries could increase the aggregate markup. At the plant-level, different from De Loecker et al. (2020), our decomposition is constructed based on the comparison in markups between different groups of plants (continuers, entrants and exiters) with that of continuing plants in the corresponding period. The change of the within factor (*i.e.,* change of (unweighted) average markup within plant) drives the decline of the aggregate markup.

On the link between markups and export status, Atkeson and Burstein (2008) build a model in which exporting firms have a tendency to set a higher markup of price over marginal cost. They argue that firms with the lowest marginal costs of production are more willing to pay the costs to export. In equilibrium, these firms are likely to set lower prices in the domestic market, have higher sales and have a higher market share in their local sector than non-exporting firms. The model combines the tendency of exporters to have a higher market share in their inland sector and the assumption that a firm's markup is an increasing function of the firm's market share within its sector to predict that exporters tend to charge a higher markup. Bernard, Eaton, Jensen, and Kortum (2003) develop a trade model which provides an insight into markups and a firm's export status. Their model shows that more efficient firms are more likely to export. Besides, they also tend to be in a leading position compared to their rivals, and thus can set a higher markup. Empirically, De Loecker and Warzynski (2012) confirm the link between firm markup and export status in the case of Slovenian manufacturing. Further, these authors explore the pattern of markups for export entering firms. They find that considerably higher markups accompany export entry. We contribute to the literature by examining the causality between markups and export status: whether plants with higher markups tend to become exporters, or exporting activities help plants to raise their markups, or both. Estimating the model of decision to export by the plant, as Bernard and Jensen (1999), we confirm that plants with prior success (larger size, higher productivity, wages and markups) are more likely to export in the case of Vietnamese manufacturers and they continue to have higher markups after entering export markets.

Another strand of literature that relates to our research concerns the relationship between trade and markups. In early studies, Levinsohn (1993) examined the "imports-as-market-discipline" hypothesis using firm level data from the Turkish manufacturing sector with the assumption of invariant markups among firms. The hypothesis implies that trade liberalisation can reduce markups in industries with imperfectly competitive and import-competing markets. Specifically, the author finds evidence of lowering markups within a group of imperfectly competitive industries when trade was liberalised or the level of protection was reduced. Harrison (1994) uses plant-level data to investigate changes in market power following trade reform in Cote d'Ivoire. The author found that the price-marginal cost ratio fell following the trade reform for sectors with higher tariffs and quota protection, such as food and textiles. Recently, Edmond et al. (2015) proposed a theoretical framework in which they use a quantitative model with endogenous variable markups to investigate whether trade liberalisation reduces markups and markup dispersion. They argue that trade can reduce markup dispersion if there is misallocation, that international trade makes producers face greater competition and that domestic producers and foreign producers seem to have fairly comparable levels of productivity in the same sector. In that case, trade liberalisation opens up the heat of competition within sectors and constrains market power thereby lowering markups and markup dispersion. Investigating the impact of Indian trade liberalisation in the form of tariff reduction on both final goods and inputs, De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) show that the net effect is a rise in markups since the costs of final good producers decrease due to the input tariffs fall. Lu and Yu (2015) studied the impact of trade liberalisation under China's accession to WTO in 2001 by using the methodology built by De Loecker and Warzynski (2012) with Chinese firm-level data. The authors tried to compare the markup dispersion experienced by industries with larger tariff cut (the treatment group) to industries with lower tariff reductions (the control group) upon the accession to WTO, meaning before and after the 2001 WTO accession. Theirs difference-in-differences coefficients show that trade liberalisation considerably reduces the dispersion of firm markups in China.

The chapter is structured as follows. In Section 3.2, we describe our database and De Loecker and Warzynski (2012)'s methodology in estimating plant level variable markups. In Section 3.3, we present the empirical results. The conclusion is in Section 3.4.

3.2 Empirical framework

3.2.1 Data

Main data

In this research, we use two different datasets. The first one is the plant-data from the Vietnamese Enterprise Survey collected annually by the General Statistics Office of Vietnam (VGSO) from 2000. The data includes registered firms in various industries such as agriculture, manufacturing, construction, transport and services. However, in this research, we focus on manufacturing industry. Firms are distinguished by their own tax identifiers. These surveys report abundant information on various types of firm (state-owned enterprises (SOEs, thereafter), foreign-owned firms, domestic private firms). There is general information (tax identifiers, firm code, plant code, type of ownership, year of establishment, export/import status, major industry, city), worker related information (number of workers, wages), accounting information (assets, liabilities, sales,...). Since we are using the data at plant-level, we combine firm identifiers with plant code and city to distinguish plants. In order to make this paper comparable with previous studies, we only take into account plants with more than 10 workers.

However, the dataset had some features which presented problems. Firstly, export status and export sales are not available for every year. So, we use other information such as export taxes to define exporters. Even so, the number of trading plant is still under-represented for some years. Secondly, since intermediate input data is not available in the dataset, we use an indirect way to calculate intermediate input. In fact, it can be calculated as follows:[30]

$$Intermediate\ Input_{it} = Sales_{it} - (Wages_{it} + Depreciation_{it} + PretaxProfit_{it}) \tag{3.1}$$

Thirdly, for the year of 2001, the data on foreign-owned plants' equity share is not available, so we interpolate the data in 2001 as follows: if plants are foreign-owned in 2000 and 2002, then they are so in 2001. Besides, some plants changed industry during this period. To deal with this problem, we follow Ha et al. (2016) to define the industry to which plants belong.[31] Then, we also identify entry, exit and continuing plants and only account for plants without re-entry. Finally, we obtain an unbalanced panel data with 59,388 plants over the period 2000-2013.

During the period studied, there was change in Vietnam Standard Industry Classification (VSIC, thereafter). Before 2007, VSIC was based on ISIC Rev.3, which was called VSIC 1993. From 2007, the old classification was replaced by VSIC 2007 which was developed on the base of ISIS Rev.4 and ASIAN Common Industrial Classification (ACIC). For the convenience of correspondence with tariff dataset and comparison with previous studies, we convert VSIC 2007 to VSIC 1993.

[30]Unfortunately, we do not have data on interest paid on credit and loans, which is part of the payments to capital. Therefore, as in Ha and Kiyota (2014), we use the more relaxe definition of value added, that is the sum of wages, depreciation and pre-tax profit.

[31]If a plant has changed industries, the plant's industry is the one in which this plant stays most of the time. If a plant stays in more than one industry for the same length of time, we use the most recent industry as the plant's industry.

Other data

The second dataset is about Vietnamese tariffs available on the World Integrated Trade Solution (WITS) developed by the World Bank. In this research, we use the two-digit of VSIC industry to mitigate the measurement error due to the fact that plants switch industry during their operation period. Therefore, we use the nomenclature ISIC Rev.3 at 2-digit as product classification. For each product, tariff data provide information on different duty types (Most Favoured Nation - MFN, Bound Tariff - BND and Effectively Applied-AHS), average, minimum and maximum ad-valorem tariffs. Since tariff rates are not available for 2011, we interpolate these values in 2011 as the average of the ones for 2010 and 2012. We use the MFN average tariff at 2-digit industry.

3.2.2 Estimating markup from producer behavior

In this paper, we follow De Loecker and Warzynski (2012) to estimate markups at plant-level. This approach relies on the optimal choice of variable input from the cost-minimization problem. Under the first-order condition of the variable input, markups, which is defined as price-marginal cost ratio, is related to the output elasticity of variable input and its expenditure share in total sales. For the purpose at hand, the former is obtained from the production function estimation while the latter is directly observed from the data.

This approach begins with the production function of plant i at time t:

$$Q_{it} = F_{it}(L_{it}, K_{it}, \omega_{it}) \tag{3.2}$$

where L_{it}, K_{it} denote plant's inputs of labour and capital; ω_{it} denotes plant's productivity. The derivation of markup expression requires that function $F(.)$ is continuous and twice differentiable with respects to its arguments.

The Lagrange function for cost-minimization problem of plant i is then:

$$\mathcal{L}(L_{it}, K_{it}, \lambda_{it}) = w_{it}L_{it} + r_{it}K_{it} + \lambda_{it}\left(Q_{it} - F_{it}(.)\right) \tag{3.3}$$

where w_{it}, r_{it} denotes wage rate and rental price for capital, respectively. Since capital is considered as a dynamic input of plant, we assume that plants seek the optimal choice of labour for the cost-minimization problem. The first-order condition of labour input is:

$$\frac{\partial \mathcal{L}}{\partial L_{it}} = w_{it} - \lambda_{it}\frac{\partial F_{it}(.)}{\partial L_{it}} = 0 \tag{3.4}$$

where λ_{it} is the marginal cost of production at a given level of output and $\lambda_{it} = \frac{\partial \mathcal{L}}{\partial Q_{it}}$. By rearranging Equation (3.4) and multiplying both sides by $\frac{L_{it}}{Q_{it}}$, the following expression is obtained:

$$\frac{\partial F_{it}(.)}{\partial L_{it}}\frac{L_{it}}{Q_{it}} = \frac{1}{\lambda_{it}}\frac{w_{it}L_{it}}{Q_{it}} \tag{3.5}$$

The right hand side represents the output elasticity of labour input, denoted by θ^L_{it}:

$$\theta^L_{it} = \frac{1}{\lambda_{it}}\frac{w_{it}L_{it}}{Q_{it}} \tag{3.6}$$

Since markup μ_{it} is defined as the price-marginal cost fraction, $\mu_{it} \equiv \frac{P_{it}}{\lambda_{it}}$, Equation (3.6) can be rewritten as:

$$\theta_{it}^{L} = \mu_{it} \frac{w_{it} L_{it}}{P_{it} Q_{it}} \tag{3.7}$$

Let's denote α_{it}^{L} the share of expenditure on L_{it} in total sales ($P_{it}Q_{it}$), i.e.

$$\alpha_{it}^{L} = \frac{w_{it} L_{it}}{P_{it} Q_{it}} \tag{3.8}$$

Replacing α_{it}^{L} in Equation (3.7), we have:

$$\theta_{it}^{L} = \mu_{it} \alpha_{it}^{L} \tag{3.9}$$

So, the plant-level markups is calculated as:

$$\mu_{it} = \frac{\theta_{it}^{L}}{\alpha_{it}^{L}} \tag{3.10}$$

Two main elements need to be identified in order to estimate markups: the expenditure share of labour in total sales, α_{it}^{L}, and the output elasticity of labour input, θ_{it}^{L}. While the former could be directly observed in the data, the latter will be obtained from the production function estimation, which is described in the following section.

Production function estimation

The estimation of production functions is the crucial step in meansuring plant markups. There are two major empirical approaches to estimating the production function which are gross output and value added production function. In the first approach, labour, capital and intermediate inputs are used to produce output. In the second approach, the output of a plant is a function of capital and labour only. In regards to the value added production function, Gandhi et al. (2017) distinguish two specifications to relate gross output to value added: the restricted profit value added and the structural value added. In the restricted profit value added specification, the dependent variable is value added and the intermediate input is not included in the estimated production function. The second specification is to follow Leontief in the intermediate inputs, i.e. the intermediate inputs are a perfect complement to the combination of labour and capital, the dependent variable is thus still gross output. In this paper, we use the restricted profit value added production function, as Petrin and Levinsohn (2012) argue that the estimates from this approach have a much more direct welfare interpretation and therefore can be useful in further investigations.

We focus on the value added Cobb-Douglas production function in logs for a given two-digit industry, i.e.

$$y_{it} = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \omega_{it} + \epsilon_{it} \tag{3.11}$$

where y_{it} is value added, l_{it}, k_{it} denotes labour, capital. Two unobservable terms to the econometricians are ω_{it} and ϵ_{it}. The difference between those two is that ϵ_{it} represents measurement error in the output and is unobserved by plants when they make input decisions, so, it is uncorrelated with k_{it} and l_{it}; by contrast, ω_{it}, called the productivity shock, is predictable by plants during their decision making process and hence it could be correlated with input variables. The correlation between ω_{it} and plant's input choices raises the endogeneity problem when estimating the production function. In that case, estimation by OLS will bias the coefficients of the observed inputs.

An well-known approach dealing with this issue is pioneered by Olley and Pakes (1996) (hereafter OP), extended by Levinsohn and Petrin (2003) (hereafter LP). These two methodologies are then augmented by Ackerberg et al. (2015) (hereafter ACF). The set of methodologies try to eliminate the endogeneity issue by measuring ω_{it} through an equation. To do so, it is based on several fundamental assumptions.

First, the productivity ω_{it} follows a first-order Markov process, *i.e.,*

$$p\left(\omega_{it+1}|I_{it}\right) = p\left(\omega_{it+1}|\omega_{it}\right) \tag{3.12}$$

where I_{it} is plant i's information set at t. This assumption states that the future productivity, ω_{it+1} is expected based on the current set of information, I_{it}, that consists of current and past productivity and not the future one ω_{it+1}. This assumption also implies that plant's decisions at t are uncorrelated with the unanticipated innovation in ω_{it+1} between t and $t+1$, denoted ξ_{it+1} with $\xi_{it+1} = \omega_{it+1} - E[\omega_{it+1}|I_{it}] = \omega_{it+1} - E[\omega_{it+1}|\omega_{it}]$ and $E[\xi_{it+1}|I_{it} = 0]$.

Second, all three approaches make the same assumption on the timing of the plant's capital choice. Capital is accumulated following the usual equation:

$$k_{it} = (1-\delta)\,k_{it-1} + i_{it-1} \tag{3.13}$$

where i_{it-1} is plant i's investment at $t-1$. The law of motion implies that plant's capital stock at t is actually decided at $t-1$, so belongs to I_{it-1}. This implication is crucial in identifying the output elasticity of capital.

About the labour input, OP assume that l_{it} is a non-dynamic variable and chosen at t. In the LP model, the non-dynamic variables l_{it} and m_{it} are assumed to be chosen simultaneously at time t, after plants have information about ω_{it}. Therefore, l_{it} might be correlated with the current value of ω_{it}. So, ω_{it} needs to be controlled for to solve the endogeneity problem with respect to l_{it}.

Third, these approaches identify the control function in which ω_{it} is the only econometric unobservable and it must be strictly monotonic in ω_{it}. The function choice differs between OP and LP. Indeed, OP consider plant's investment demand function, *i.e.,* $i_{it} = f_t\left(k_{it}, \omega_{it}\right)$. However, LP argue that in many datasets in developing countries, investment can be 0 for many plants in some years. The OP proxy is then unavailable. LP propose intermediate inputs, m_{it}, as a new proxy that is not often 0. The intermediate input demand function is given by $m_{it} = f_t^*(k_{it}, \omega_{it})$.

The monotonicity assumption allows the control function to be inverted to obtain ω_{it} as a function of intermediate input and capital:[32]

[32] In our research, we follow LP by using intermediate input rather than investment as a proxy variable for productivity because there is only 15% of total plants per year in our raw data having information about investment. The positive intermediate inputs are reported for over 80% of plant per year.

$$\omega_{it} = f^{*-1}(k_{it}, m_{it}) \tag{3.14}$$

The equation above implies that ω_{it} could be "measured" through the variables observed by the econometricians. Although the functional form is unknown, f^{*-1} could be approximated by the polynomial in k_{it} and m_{it}. Substituting Equation (3.14) in Equation (3.11), one could obtain the estimator of β_l by regressing y_{it} on l_{it} and the high order polynomial in (k_{it}, m_{it}).[33] However, ACF criticise this implement by questioning the independence of l_{it} and the polynomial in k_{it} and m_{it}. In other words, if l_{it} is functionally dependent on k_{it} and m_{it}, β_l cannot be distinctly identified from the polynomial function of (k_{it}, m_{it}). The authors propose an alternative procedure which relaxes the assumption of a plant's labour choice in OP/LP methodology to account for the functional dependence problem and hence corrects the intermediate input demand function. Accordingly, l_{it} might have dynamic implications and is set at period t, $t-1$ or at some point $t-b$ with $0 < b < 1$. This implies that l_{it} belongs to the state space of the plant's dynamic problem. Since m_{it} is chosen at t, plant's intermediate input demand at t will depend on the l_{it} decided in advance. Therefore, the plant's intermediate input demand function is given by:

$$m_{it} = m_t(k_{it}, l_{it}, \omega_{it}) \tag{3.15}$$

The monotonicity assumption allows to inverted Equation (3.15) to obtain ω_{it}:

$$\omega_{it} = m_t^{-1}(k_{it}, l_{it}, m_{it}) \tag{3.16}$$

Substituting Equation (3.16) in Equation (3.11) does not provide directly any β because of the functional dependence of l_{it} on m_{it}. All coefficients will be estimated together in the second step.

The ACF procedure is performed in two stages. Substituting Equation (3.16) in Equation (3.11), the production function could be rewritten as:

$$y_{it} = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + m_t^{-1}(k_{it}, l_{it}, m_{it}) + \epsilon_{it} \tag{3.17}$$

Since it is not possible to differentiate the β from the coefficients of k_{it} and l_{it} in the function m_t^{-1}, the first stage is to estimate the model:

$$y_{it} = \phi_t(k_{it}, l_{it}, m_{it}) + \epsilon_{it} \tag{3.18}$$

where

$$\begin{aligned} \phi_t(k_{it}, l_{it}, m_{it}) &= \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \omega_{it} \\ &= \beta_0 + \beta_l l_{it} + \beta_k k_{it} + m_t^{-1}(k_{it}, l_{it}, m_{it}) \end{aligned} \tag{3.19}$$

We approximate ϕ_t by the third order of polynomial of k_{it}, l_{it} and m_{it}. From this stage, we obtain the estimates of expected output, $\hat{\phi}_{it}$.

[33] This step is called the first step. β_k will be estimated in the second step.

The second stage is related to estimating the consistent β from Equation (3.19). Recall that this approach supposes that ω_{it} follows the first order of Markov process, i.e., $\omega_{it} = E[\omega_{it}|I_{it-1}] + \xi_{it}$ where $E[\xi_{it}|I_{it-1}] = 0$. So, Equation (3.19) could be written as:

$$\hat{\phi}_t = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + E[\omega_{it}|I_{it-1}] + \xi_{it} \tag{3.20}$$

This approach relies on the timing of the plant's input choice to identify different moment conditions and then estimates β. Indeed, as mentioned above, since k_{it} is actually decided at $t-1$, $k_{it} \in I_{it-1}$, this implies that $E[\xi_{it}|k_{it}] = 0$. The uncorrelated relation between ξ_{it} and k_{it} comes from the mean independence, meaning that:

$$E[\xi_{it} k_{it}] = 0 \tag{3.21}$$

In the ACF procedure, labour is allowed to have dynamic implications, that means, it could be chosen at $t-1, t$ or somewhere in-between, $t-b$. If labour is chosen after $t-1$, for example at a point $t-b$, l_{it-1} was chosen at $t-b-1$. l_{it-1} belongs to I_{it-1}, *i.e.,* [34]

$$E[\xi_{it} l_{it-1}] = 0 \tag{3.22}$$

For some industries, there is a possibility that workers need a significant amount of training time before entering into the production. In that case, labour could be actually chosen at $t-1$, *i.e.,* $l_{it} \in I_{it-1}$. So, we have:

$$E[\xi_{it} l_{it}] = 0 \tag{3.23}$$

We use the three over-identifying conditions above to have to the moment conditions, which help to estimate β, as follows:

$$E\left(\xi_{it}(\beta)\begin{pmatrix}1\\ l_{it-1}\\ l_{it}\\ k_{it}\end{pmatrix}\right) = 0 \tag{3.24}$$

In practice, the second stage is executed as follows. We choose the candidate values of β, denoted as $(\beta_0^*, \beta_k^*, \beta_l^*)$ or β^*. Recalling that from the first stage, we obtain the estimate of ϕ_{it}, $\hat{\phi}_{it}$. From Equation (3.19), we have the estimate of productivity given β^*:

$$\omega_{it}^*(\beta^*) = \hat{\phi}_{it} - \beta_0^* + \beta_l^* l_{it} + \beta_k^* k_{it} \tag{3.25}$$

Since ω follows the first order Markov process, it implies that:

$$\omega_{it} = g(\omega_{it-1}) + \xi_{it} \tag{3.26}$$

Regressing $\omega_{it}^*(\beta^*)$ on its lags, ω_{it-1}^*, we obtain $\xi_{it}(\beta^*)$.[35] Applying the moment conditions

[34] LP also use this moment as an over-identifying restriction in the second stage.

[35] We regress $\omega_{it}^*(\beta^*)$ on the 4th-order in $\omega_{it-1}^*(\beta^*)$. We also include the probability of survivors, as in OP,

in Equation (3.24) and using the standard GMM techniques and the block bootstrapping for the standard errors, we obtain the estimated parameters of the production function, β. The output elasticity of labour, θ_{it}^{L} in Equation (3.10), corresponds to the estimates of β_l in the production function.[36]

Markup correction

Since the output elasticity of inputs are estimated based on the value added specification of production function, it is necessary to make some correction in calculating markups in Equation (3.10).[37]

Firstly, we use the restricted profit value added specification to estimate the production function. In this setup, intermediate input is ruled out in both sides of Equation (3.11). So, the estimate β_l or θ_{it}^{L} is the elasticity of value added with respect to labour. Bruno (1978) suggests using the plant-level ratio of value added to gross output to approximate locally the value added elasticity of labour to the gross output counterpart:

$$\tilde{\hat{\theta}}_{it}^{L} = \hat{\theta}_{it}^{L} \times \frac{VA_{it}}{GO_{it}} \tag{3.27}$$

where $\tilde{\hat{\theta}}_{it}^{L}$ is the elasticity of gross output with respect to labour, $\hat{\theta}_{it}^{L}$ is the estimate of the elasticity of value added with respect to labour from equation Equation (3.11), VA_{it}, GO_{it} are value added and gross output, respectively, of plant i at time t.[38]

The second element that we need to calculate markups is the expenditure share of labour, α_{it}^{L}, which can be directly calculated from available data. According to De Loecker and Warzynski (2012), since variation in output which is not correlated with $\phi_t(l_{it}, k_{it}, m_{it}, z_{it})$ can create variation in expenditure share, this variable of expenditure share of labour should be corrected as follows:

$$\hat{\alpha}_{it}^{L} = \frac{w_{it} L_{it}}{P_{it} \frac{\tilde{Q}_{it}}{exp(\hat{\epsilon}_{it})}} \tag{3.28}$$

where $\hat{\epsilon}_{it}$ is obtained from the first stage. Then, markups of plant i at time t can be simply estimated as:

$$\mu_{it} = \tilde{\hat{\theta}}_{it}^{L} \left(\hat{\alpha}_{it}^{L}\right)^{-1} \tag{3.29}$$

to control for the selection bias. Although we use an unbalanced panel data which, according to LP, accounts for the plant dynamic, introducing the probability of survivors could make our estimates well identified. We use the probit model to regress the probability of survivor at $t+1$ on plant's characteristics at t, such as labour, capital, intermediate inputs. We also introduce plant's ownership status and region dummies.

[36] Since we use the Cobb-Douglas production function, the output elasticity of labour is the same within an industry. The more flexible specification of production function (the trans-log model) is used for the robustness check. In that case, the output elasticity of labour is different across plants.

[37] See Scott et al. (2017) for computing markups under various assumptions on i) technology, ii) which type of production inputs do not depend on adjustment cost and can be set at will by the plant's manager and iii) how productivity progresses.

[38] In particular, gross output is measured as deflated sales.

Markup dispersion

After estimating the plant's markups, the calculation of markup dispersion of an industry is straightforward. To measure industry's markup dispersion, we use the Theil index in the main specification:

$$\text{Theil}_{it} = \frac{1}{N_{it}} \sum_{p=1}^{N_{it}} \frac{\mu_{pit}}{\bar{\mu}_{it}} \log\left(\frac{\mu_{pit}}{\bar{\mu}_{it}}\right) \tag{3.30}$$

where μ_{pit} is the markups of plant p in industry i at time t, $\bar{\mu}_{it}$ is the average markup of industry i at time t, N_{it} is the number of plants in industry i at time t. Besides, we also use other measures of dispersion for robustness check such as Gini index, the relative mean deviation $\left(\text{RMD}_{it} = \frac{1}{N_{it}} \sum_{p=1}^{N_{it}} \left|\frac{\mu_{pit}}{\bar{\mu}_{it}} - 1\right|\right)$, the mean log deviation $\left(\text{MLD}_{it} = \frac{1}{N_{it}} \sum_{p=1}^{N_{it}} \log\left(\frac{\mu_{pit}}{\bar{\mu}_{it}}\right)\right)$.

3.3 Results

3.3.1 Markup investigations

Markup evolution

To figure out the markup evolution in the Vietnamese manufacturing sector for the period 2000-2013, we derive aggregate markup, which is weighted average of plant-level markups:

$$\mu_t = \sum_i w_{it} \mu_{it} \tag{3.31}$$

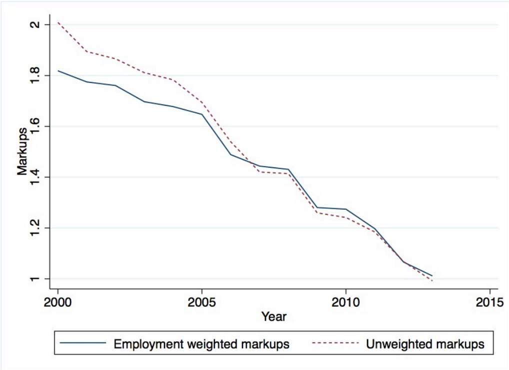
where $w_{it} = \frac{X_{it}}{\sum_i X_{it}}$ $(\sum_i w_t = 1)$ is the market share of plant i in the sample at time t, with X_{it} represent variables chosen as weights, such as number of employees, value added, etc.

It should be noted that the weighting choice will affect how average markups are computed. For example, from the model of imperfect competition, one can choose sales as weight, aggregate markup is then sales-weighted harmonic mean of plant markups (Edmond et al., 2015). That is also equivalent to the cost-weighted arithmetic mean of plant markups which is, as pointed out by Edmond, Midrigan, and Xu (2018): the "wedges" in the aggregate employment and investment maximising conditions, which represent the distortion to employment and investment decisions, are proportional to the cost-weighted average markups and not the sales-weighted average. However, output weighted average markups are not without meaning. Indeed, using output weighted measure and comparing it to input counterpart may give us more information about how plants generate output using their resources.[39] Therefore, we investigate average markup with different input weights (employment, harmonic mean of sales that is equivalent to cost weighted measure) and output weight (value added, since the technology we consider is restricted profit value added).

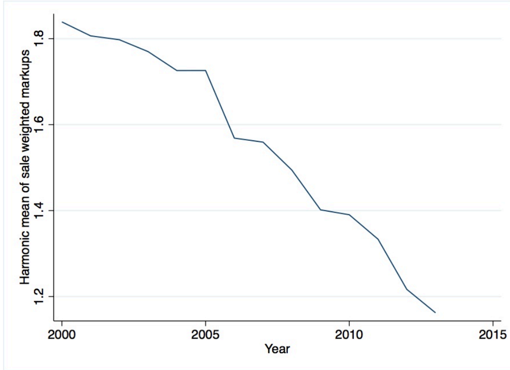
Figure 3.1 presents the evolution of unweighted (dashed line) and weighted (solid line) average markups. In Figure 3.1a and Figure 3.1b, we use two input weights, employment and cost of goods sold (COGS).[40] We also use value added weighted (output weighted) markup in

[39]see De Loecker and Eeckhout (2018b) for more discussion about weights for average markups.

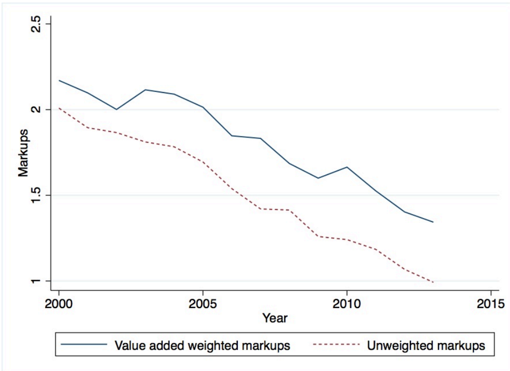
[40]Since we do not have data on COGS to compute directly cost-weighted average markup, we use har-



(a) Employment



(b) Harmonic mean of sale (or cost of goods sold)



(c) Value added

Figure 3.1: Markup evolution

Figure 3.1c. The comparison between weighted and unweighted average markups provides us information on the relationship between markups and plant size. For all weights, despite differing magnitude in markup evolution according to weight choice, we obtain the same pattern: markups decrease overtime. In addition, the level of output weighted markups is higher than input weighted markups. That means, when we use input weights, we have lower weight on high markup plants than when using output weights. The demand curve faced by high markup plants shifts up since they charge higher prices, these plants will use fewer inputs but still generate more sales. This is evident for the reallocation of sales from low to high markup plants.

Distribution of markups

Figure 3.2 show the distribution of plant markups in 2000 and 2013. We use the Kolmogorov-Smirnov test for equality of distribution functions. As shown in Table 3.1, there is evidence that markups values in 2013 are smaller than in 2000, so there is a decrease in markups over this period. Moreover, a variance ratio test shows the evidence that markup dispersion in 2000 is higher than that in 2013.

The contour plot of the kernel density of the joint distribution of plant's markup and three variables used for weights (employment, sales, value added) in Figure 3.3 details the distribution of plant-level markups. Firstly, since the joint distributions are moved to the left, we obtain the same results as above that markups decrease over time. Secondly, due to the closer contour lines

monic sale weighted markup. The latter is proved equivalent to the former, see Edmond et al. (2018).

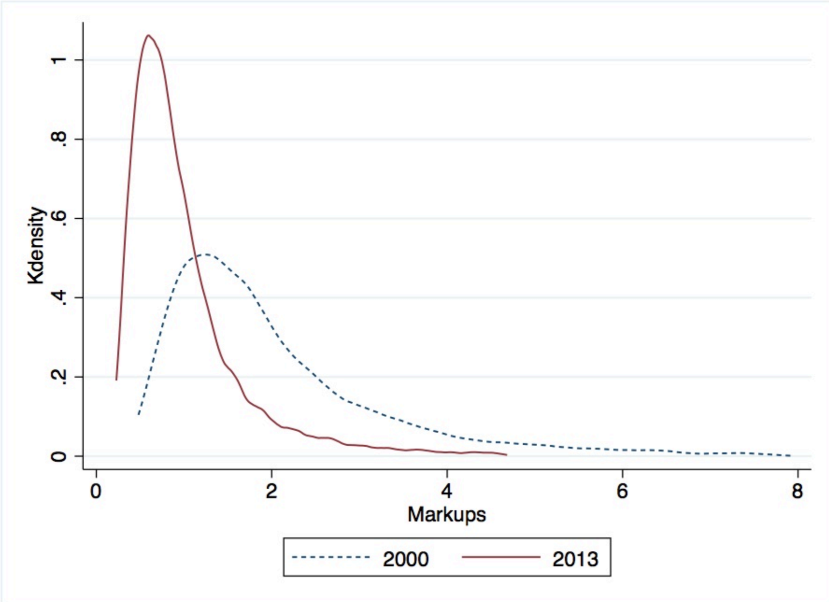


Figure 3.2: Empirical distribution of markups, 2000-2013

Table 3.1: Distribution test

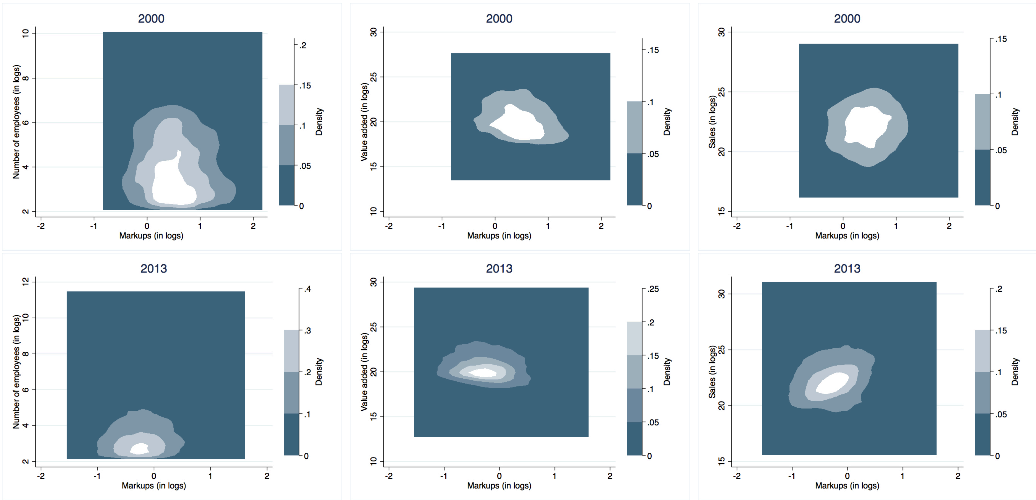
Kolmogorov-Kmirnov test		
Smaller group	D	p-value
2000	0.000	1.000
2013	−0.496	0.000
Combined K-S	0.496	0.000
	Variance ratio test	
Markups in	Mean	Std. Dev
2000	2.009	1.234
2013	0.992	0.661
Combined	1.175	0.886

In the two-sample Kolmogorov-Smirnov test, the first line tests the null hypothesis that markups in 2000 is smaller than in 2013, the second line tests the null hypothesis that markups in 2013 is smaller than in 2000; D indicates the difference between the distribution functions. In the variance ratio test, $ratio = sd(\mu_{2000}/\mu_{2013}), f = 3.491, df = 4749, 216; H_0 : ratio = 1, H_1 : ratio < 1, Pr(F < f) = 1.000, H_1 : ratio \neq 1, Pr(F < f) = 0.000, H_1 ratio > 1, Pr(F < f) = 0.000$.

in 2013, the joint distribution between markups and weights variables, such as labour and value added, becomes less dispersed.

Markup and plant size

Markup and plant size - The main variable in this research is the plant markup, which is obtained from equation Equation (3.29). To understand how markups could be determined across heterogeneous plants, one could be interested in how markups relate to plant characteristics, such



(a) Labour (b) Value added (c) Sales

Figure 3.3: Contour plot

as plant size. The relationship between markups and plant size is found in some recent literature, in both theoretical and empirical studies. Theoretically, in the model of imperfect competition proposed by Edmond et al. (2015), since the elasticity of substitution across sectors is lower than that across goods within a sector, plant markups are positively correlated with plant's share of sectoral revenue. In other words, larger producers tend to charge higher markups. Empirically, in the Taiwanese manufacturing data, the authors observe that larger producers set higher markups, measured by producer's inverse labour shares. By defining markups as the level of profit in the sense that markup plus cost is equivalent to price, Atkin et al. (2015) also find markups are positively correlated with employment (plant size) in the case of Pakistan. In addition, Kugler and Verhoogen (2011) show that large plants impose higher output prices in Colombian manufacturing industries. With a narrow definition of industry, De Loecker et al. (2020) find a positive correlation between markup and different types of plant size measurements (sales, employment, variable input - cost of goods sold) within sectors.

In Figure 3.3, we notice that the contours' ridge of the joint distribution of plant's markups and sales is upward slopping (in 2013), indicating a positive correlation between markups and sales. Unfortunately, the relation between markups and plant size, measured as either employment or value added, are less clear. However, as mentioned above, the comparison between unweighted and weighted average markup, represented graphically in Figure 3.1, could provide information on the relationship between markups and plant size. We also observe that for three weights considered, weighted average markup seems to be higher than unweighted. This means larger plants have a tendency to set higher markups. This point consolidates the prediction of the model of imperfect competition by Edmond et al. (2015).

Markup across industries

In the main part of this study, since we use the Cobb-Douglas specification to estimate the production function for each industry, output elasticities of labour are the same across plants.[41] As

[41]Note that since output elasticities of labour are time-invariant and invariant across plants within an industry, the higher (corrected) expenditure share of labour induces lower plant's markup. We also perform

Table 3.2: Output elasticities of inputs

VSIC	Industry	Cobb-Douglas L	K	Return to scale
15	Food products and beverage	0.882	0.268	1.150
16	Tobacco products	0.948	0.149	1.097
17	Textiles	0.701	0.218	0.919
18	Wearing products	0.872	0.0545	0.927
19	Leather products	0.885	0.0468	0.932
20	Wood and cork manufacturing	0.834	0.214	1.048
21	Paper products	0.803	0.132	0.935
22	Publishing, priting, recording media	1.010	0.183	1.193
23	Manufacture of coke, refined petroleum products and nuclear fuel	0.913	0.327	1.240
24	Chemical manufacturing	0.835	0.162	0.998
25	Rubber and plastics products	0.847	0.190	1.037
26	Other non-metallic mineral products	0.780	0.181	0.961
27	Basic metals manufacturing	0.965	0.286	1.251
28	Fabricated metal products, except machinery and equipment	0.889	0.172	1.061
29	Machinery and equipment	0.929	0.133	1.062
30	Office, accounting and computing machinery	0.766	0.269	1.035
31	Electrical machinery and apparatus	0.877	0.190	1.068
32	Radio, television and communication equipment	0.874	0.218	1.092
33	Medical, precision and optical instruments, watches and clocks	0.871	0.146	1.017
34	Motor vehicles, trailers and semi-trailers	0.924	0.290	1.213
35	Other transport equipment	0.942	0.123	1.065
36	Furniture	0.792	0.109	0.901

shown in Table 3.2, for all industries, output elasticities of labour are higher than those of capital. Although the methodology of De Loecker and Warzynski (2012) in estimating markups does not impose any prior assumptions on the constant return to scale, looking at the returns to scale would provide more details on plant characteristics. Moreover, it helps to verify whether our procedure in estimating production function is accurate. For example, as shown in De Loecker et al. (2016), their initial procedure that uses total wage as labour input (instead of the number of employees as in our research) gives nonsensical estimates of the production function with negative, very low or very high returns to scale. After correcting for input price variation due to the use to total wage as labour input, the authors obtain a rational return to scale picture (positive and not very high or very low compared to 1). Since we use the number of employees as labour input, to check whether we have a good procedure for estimation, we also report the output elasticities and the return to scale. We get a qualitatively similar result to that of De Loecker et al. (2016), meaning that our procedure is acceptable in this sense.

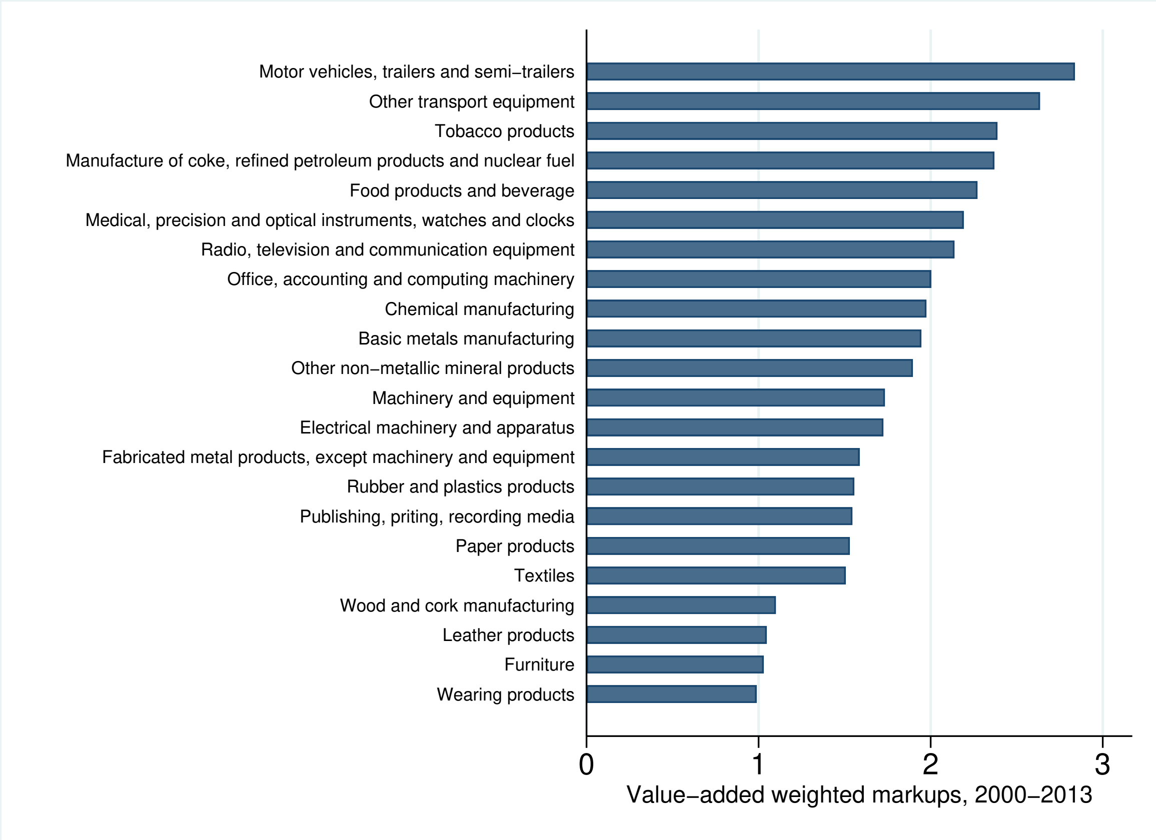
To compute the industry markups, we follow the insight of Olley and Pakes (1996). The industry markups are then weighted averages of plant-level markups:

$$\mu_{st} = \sum_{i \in s} s_{it} \mu_{it} \tag{3.32}$$

where μ_{st} is markup of industry s at time t, s_{it} is plant i's share of value added at time t, μ_{it} is markup of plant i at time t. Then, we compute the mean of μ_{it} over 2000-2013 to compare markups across industries. As shown in Figure 3.4, during the period 2000-2013, all industries record (weighted) markups higher than 1. We observe that highly labour intensive industries (such

the translog specification. In this case, we obtain output elasticities of input varying across plants, see Table 3.A.1 in Appendix 3.A.

Figure 3.4: Estimated mean markups for 2-digit VSIC 1993 manufacturing industries



as wearing, leather goods, etc.) have lower markups than less labour intensive industries.[42]

3.3.2 Markup decomposition

With the objective of studying how change in aggregate markup over time is contributed by different composition changes, we decompose aggregate markups using various approaches.

Firstly, we follow De Loecker and Eeckhout (2018a) by decomposing aggregate markups change over time (from $t = 1$ to $t = 2$), $\Delta\mu = \mu_2 - \mu_1$, into change of markup at the industry level ($\Delta within$), change in the composition between industries ($\Delta between$) and the joint change in markups and the industries composition ($\Delta crossterm$):

$$\Delta\mu = \sum_j s_{j1}(\mu_{j2} - \mu_{j1}) + \sum_j (s_{j2} - s_{j1})\mu_{j2} + \sum_j (s_{j2} - s_{j1})(\mu_{j2} - \mu_{j1}) \tag{3.33}$$

where μ_{jt} is markup of industry j at time t ($t = 1$ or 2); s_{jt} is value added share of industry j at time t, $s_{jt} = \frac{VA_{jt}}{\sum_j VA_{jt}}$ where VA_{jt} is value added of industry j at t. The first term ($\Delta within$) represents the change in markup at industry level, the second term ($\Delta between$) involves the change in the composition of plants, and the last term ($\Delta reallocation$) is the joint change in markup and the plant composition.

[42]This finding still holds when we use employment and cost (or harmonic mean of sales) weights to compute industries markups, see Figure 3.B.1 in Appendix 3.B.

Then, we use the insight of productivity decomposition proposed by Melitz and Polanec (2015) to decompose aggregate markup change at plant-level. This method is derived from Olley and Pakes (1996), but accounts for the dynamic of plants, i.e. the contribution of continuing, exit and entry plants. In this approach, we compare markups of each group of plants with (weighted) average markups of continuers. In the first step, we compute the value-added share and (weighted) average markup of each group of continuers, exiters and entrants. We denote $s_{gt} = \sum_{i \in g} s_{it}$ the value added share of a group g, $\mu_{gt} = \sum_{i \in g} \frac{s_{it}}{s_{gt}} \mu_{it}$ the weighted average markup of each group. Aggregate markup for each period is then a weighted average markup of three groups of plants:

$$\mu_1 = s_{C1}\mu_{C1} + s_{X1}\mu_{X1} = \mu_{C1} + s_{X1}(\mu_{X1} - \mu_{C1})$$
$$\mu_2 = s_{C2}\mu_{C2} + s_{E1}\mu_{E2} = \mu_{C2} + s_{E2}(\mu_{E2} - \mu_{S2})$$

The change of aggregate markup $\Delta\mu$ in terms of those components is then:

$$\Delta\mu = (\mu_{C2} - \mu_{C1}) + s_{E2}(\mu_{E2} - \mu_{S2}) - s_{X1}(\mu_{X1} - \mu_{C1}) \tag{3.34}$$

Considering the continuing plants only, we can decompose markup of the continuing group as a function of (unweighted) mean markup of continuers and a covariance term:

$$\begin{aligned} \mu_{Ct} &= \bar{\mu}_{Ct} + \sum_{i \in C} (s_{it} - \bar{s}_{Ct})(\mu_{it} - \bar{\mu}_{Ct}) \\ &= \bar{\mu}_{Ct} + cov_C(s_{it}, \mu_{it}) \end{aligned} \tag{3.35}$$

where $\bar{\mu}_{Ct} = \frac{1}{N_C} \sum_{i \in C} \mu_{it}$ is the unweighted average markup of continuing plants, $\bar{s}_{Ct} = \frac{1}{N_C}$ (N_C is the number of continuing plants, which is the same at $t = 1$ and 2). The change of aggregate markup is then:

$$\Delta\mu = \Delta\bar{\mu}_C + \Delta cov_C + s_{E2}(\mu_{E2} - \mu_{S2}) - s_{X1}(\mu_{X1} - \mu_{C1}) \tag{3.36}$$

The first two terms indicate the markup change for the continuing group: the first component represents the change in unweighted mean markups, the second represents market share reallocation (the covariance change between markups and market share for continuing plants). The interesting point of this decomposition is that the second term helps us to know whether the allocation across plants of value added share has become more or less markup-boosting over time. For example, a positive Δcov_C implies that the reallocation of value added is markup boosting. The third and fourth terms represent the contribution of entrants and exiters, respectively. Since we use the markups of continuing plants as the benchmark, the contribution of entrants (exiters) can be interpreted as the change in aggregate markups $\Delta\mu$ created by adding/removing the group of entrants (exiters).

Table 3.3: Decomposition aggregate markup

(a) Decomposition within and between industries

Year	Markup	Δ Markup	Δ Within	Δ Between	Δ Cross term
2000	2.170				
2001	2.097	−0.073	−0.076	0.011	−0.007
2002	2.001	−0.097	−0.083	−0.022	0.008
2003	2.115	0.115	0.099	−0.005	0.020
2004	2.090	−0.025	−0.004	−0.032	0.010
2005	2.014	−0.076	−0.090	0.007	0.007
2006	1.847	−0.167	−0.140	−0.052	0.025
2007	1.832	−0.015	−0.021	0.016	−0.010
2008	1.686	−0.146	−0.132	−0.011	−0.004
2009	1.600	−0.086	−0.102	0.011	0.005
2010	1.664	0.064	0.055	0.003	0.007
2011	1.526	−0.139	−0.092	−0.048	0.001
2012	1.403	−0.123	−0.138	−0.011	0.026
2013	1.343	−0.060	−0.061	−0.006	0.007

(b) Decomposition at group-level markups

Year	Markup	Δ Markup	Δ Within	Δ Reallocation	Δ Net entry
2000	2.170				
2001	2.097	−0.073	−0.163	0.120	−0.030
2002	2.001	−0.097	−0.099	0.026	−0.023
2003	2.115	0.115	−0.135	0.030	0.220
2004	2.090	−0.025	−0.065	0.143	−0.103
2005	2.014	−0.076	−0.129	0.056	−0.003
2006	1.847	−0.167	−0.175	0.026	−0.018
2007	1.832	−0.015	−0.133	0.092	0.026
2008	1.686	−0.146	−0.032	−0.114	−0.000
2009	1.600	−0.086	−0.189	0.159	−0.056
2010	1.664	0.064	−0.026	0.017	0.073
2011	1.526	−0.139	−0.060	0.027	−0.105
2012	1.403	−0.123	−0.088	−0.052	0.017
2013	1.343	−0.060	−0.064	0.003	0.001

Table 3.3 shows the markup decomposition results using these approaches.[43] The analysis at the industry-level (Table 3.3a) shows two main findings. First, the change of markups within industries drives that of aggregate markups. Second, the joint change in industry markups and its shares is positive, implying a positive correlation between industries' markups and their value added shares.

Then, we decompose aggregate markup while accounting for the dynamic of plants, *i.e.,* exit and entry plants based on the insight of Melitz and Polanec (2015). The results are reported in Table 3.3b. To have a briefer way of showing the aggregate markup change, as in De Loecker et al. (2020) we try to perform three counterfactual experiments based on the decomposition results in Table 3.3b. We set the original level to 2000 and then cumulatively add the changes of each of the component terms .

[43] We use the year-by-year decomposition for all approaches.

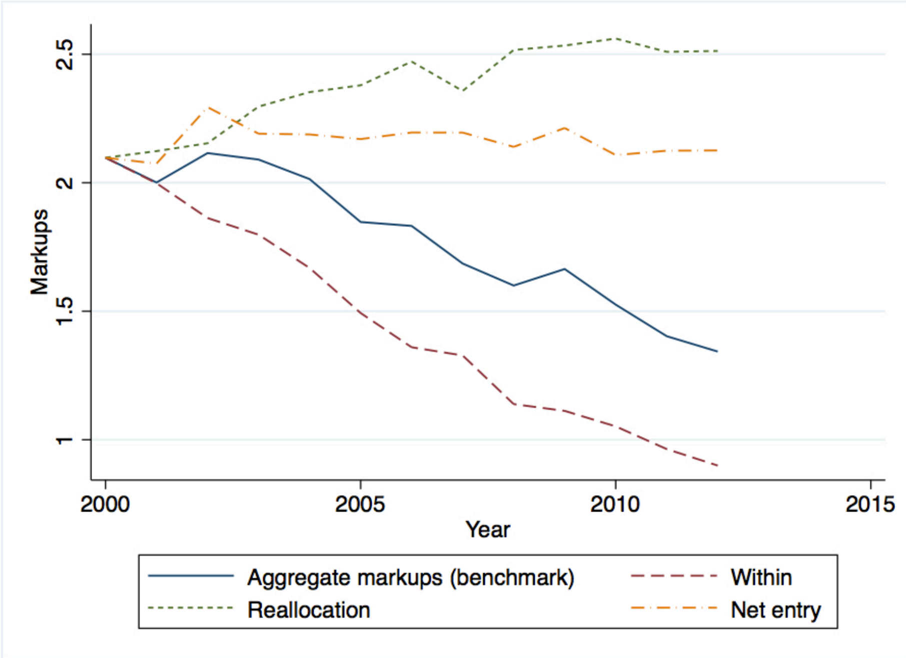


Figure 3.5: Decomposition of markup growth (aggregate markup is weighted by value added share)

Figure 3.5 shows the decomposition of aggregate markup at plant-level. The blue solid line represents the actual aggregate markups (our benchmark), the red dashed line represents the first experiment when aggregate markup change is due only to the change within industries ($\Delta within$), holding other components zero, the green short-dashed line shows the aggregate markups when it is only due to $\Delta reallocation$, the yellow short dashed-dotted line represents the aggregate markups which are only due to the net entry component.

According to this figure, the hypothetical markup evolution that is due to $\Delta within$ component has a similar motion as the benchmark, while that due to either $\Delta reallocation$ or $\Delta netentry$ is less pronounced. Interestingly, the experimental aggregate markup, which is due only to the $\Delta reallocation$ term (the green short dashed line), increases gradually over time from 2.17 to 2.5. We can recall that $\Delta reallocation$ is calculated as the change in covariance between markups and share of value added of continuing plants. The larger this covariance, the higher the share of value added that goes to plants with higher markup. In addition, the evolution of aggregate markups in this experiment indicates an increase in this covariance. That means, there is a reallocation of value added activity away from low markup plants towards high markup plants. The experimental markup evolution due to $\Delta netentry$ seems to be more or less constant overtime, suggesting that the rise in markup is not solely driven by the plant dynamics. Indeed, the incidence that the panel of plants is unbalanced with more entering than exiting could be the driving force for the net entry component.

In sum, decomposition aggregate markups either at industry-level or at plant-level clarify the source of decreasing aggregate markup over time. First, the diminishing aggregate markup is driven by the change of markups within plant or within industry. The fact that plants or industries reduce markups over time implies that the market could be more competitive. Second, the positive $\Delta reallocation$ term at plant-level that a reallocation of activity (value added) enhances higher markups.

3.3.3 Markup and export status

Recently, some studies in international trade explore the relationship between plants' markups and their export status. Theoretically, Bernard et al. (2003) show that plants that are more efficient tend to export and are more likely to have higher markup. This implies a positive correlation between plants' markup and their export status. Exporting plants can also charge higher markups because they succeed in increasing their market share in their home market by lowering their prices (Atkeson and Burstein, 2008). Higher markups for exporters could be the results of the learning-by-exporting process, that means, while exporting, producers could be more productive, lower their prices and then raise their markups. Empirically, De Loecker and Warzynski (2012) compare markups between exporters and non-exporters in Slovenian manufacturing. They find that on average, exporters have higher markup than domestic plants. However, their results are cross-sectional and do not provide implication on the causal relationship between markups and export status. It could be the case that if a plant has lower costs, then it can have higher markups, and is also more likely to export. Therfore, in this section, before examining the relationship between markups and export status, we need to clarify whether plants with higher markups in previous years tend to export. If so, this endogeneity problem must be accounted for. Then, following De Loecker and Warzynski (2012), we investigate plant markups before and after their entering (or exiting) the export market.

Do plants with higher markups become exporters?

To investigate the causality that producers with higher markups will be exporters, we follow the insight of Bernard and Jensen (1999) to investigate the decision to export by producers. Accordingly, a production unit decides to export at t, $Exp_{it} = 1$, when its current and expected revenues, $\hat{R}_{it}$, could cover the variable costs, c_{it} and any (sunk) costs of entry, N, if no export in the last period, $Exp_{it-1} = 0$:

$$Exp_{it} = \begin{cases} 1 & \text{if } \hat{R}_{it} > c_{it} + N \cdot (1 - Exp_{it-1}) \\ 0 & \text{otherwise} \end{cases}$$

This model is estimated by a binary choice non-structural approach of the form

$$Exp_{it} = \begin{cases} 1 & \text{if } \zeta X_{it} - N \cdot (1 - Exp_{it-1}) + \varsigma_{it} > 0 \\ 0 & \text{otherwise} \end{cases}$$

where X_{it} includes the exogenous variables that might affect the probability of exporting, such as plant size, productivity, markups and plant's average wage. The model could be estimated by a linear probability specification with fixed effects to circumvent the issues of unobserved plant heterogeneity. Because the direction of the causality is uncertain, one year lagged regressors are used to lessen the possible simultaneity problem (Bernard and Jensen, 2004). That means,

$$Exp_{it} = \zeta_0 + \zeta_1 X_{it-1} + NExp_{it-1} + \kappa_i + \varsigma_{it} \tag{3.37}$$

However, because the fixed effect specification with a lagged endogenous variable could

Table 3.4: Decision to export

VARIABLES	First difference (GMM)
Exported last year	0.219***
	(0.007)
Last exported two years ago	0.068***
	(0.005)
Number of employees (last year)	0.081***
	(0.002)
Average wage (last year)	0.004***
	(0.000)
Productivity (last year)	0.009***
	(0.002)
Markups (last year)	0.037***
	(0.004)
Number of observations	70,457

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

induce biased and inconsistent estimates, the authors suggest estimating Equation (3.37) in first differences using the Arellano and Bond (1991)'s GMM methodology. We use a lag of levels of all regressors as instruments:

$$\Delta Exp_{it} = \zeta_1 \Delta X_{it-1} + N \Delta Exp_{it-1} + \Delta \varsigma_{it} \tag{3.38}$$

As shown in Table 3.4, the coefficients of lagged export status are positive and statistically significant. This suggests that having exported last year increases the probability of exporting today by 22%, having exported two years ago increases the probability by only 7%. This finding also suggests that there are sunk costs for plants entering export markets. The positive sign and statistical significance of all remaining coefficients verify the hypothesis that previous success, as measured by number of employees, plant productivity, average wages and markups, increases a plant's probability of exporting. Before entering the export markets, the potential exporters are large, more productive and charge higher markups. This finding provides evidence that success leads to exporting for Vietnamese manufacturing.

Markups and export status

To examine the markup difference between exporters and domestic plants, as in De Loecker and Warzynski (2012)'s research, we use the following model:[44]

$$y_{it} = \gamma_0 + \gamma_1 e_{it} + \mathbf{b}'_{it} \varphi + u_{it} \tag{3.39}$$

[44]Since data about export status are not well identified in every year, we use alternative variables, such as export tax, to determine whether plants export or not. When investigating markup and export status, we include export dummy in the first stage of the production function estimation, *i.e.,* Equation (3.18), and in the first order Markov process of ω, *i.e.,* Equation (3.26). This specification takes into account different market structures for exporting plants (De Loecker, 2007; De Loecker and Warzynski, 2012; Van Biesebroeck, 2005).

Table 3.5: Markups, exporters and non-exporters

VARIABLES	
Export status	0.554**
	(0.249)
Capital-labour ratio	0.000**
	(0.009)
Number of employees (in logs)	0.042***
	(0.014)
Constant	−0.390***
	(0.058)
Observations	80,078
Number of plants	34,474
Plant fixed effect	Yes
Year dummies	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

where y_{it} is plant's markup in natural logarithm, e_{it} is export status dummy which equals 1 for exporters, 0 otherwise. We use plant size and capital-labour ratio as control variables. These variables are collected in vector $\mathbf{b}_{it}$. The coefficient of interest is γ_1 which denotes the markup difference in percentage between exporters and non-exporters. The positive value of this coefficient indicates that exporters experience higher markups than non-exporters. Since plants with higher markups tend to become exporters, we use some instruments for export status, that are correlated with export status, but do not directly affect markups. We choose two dummy variables, one that indicates whether a plant is state-owned and the other for foreign-ownership because foreign-owned plants and some state-owned plants have relatively more advantages in resources than other plants, so they are more likely to export. We also include plants' number of computers as having many computers could facilitate connections to foreign trading partners.[45]

The results are reported in Table 3.5.[46] The estimate of interest, $\hat{\gamma}_1$, is statistically significant and positive. That means, the markups for exporters is slightly higher than that for non-exporters 0.55% (or 0.38 in markup premium).[47] These results are consistent with the findings of De Loecker and Warzynski (2012) that exporters charge higher markup than non-exporters.

Markup before and after export market entry/exit

In this section, we investigate markups and exporters in a greater detail. In the previous subsection, we find out that markups of exporters are higher than non-exporters because they have had success before exporting. Our results are consistent with recent models of international trade. For example, Bernard et al. (2003) find that exporters charge higher markups on average than the

[45] Besides, the correlation between these instruments and export dummy is statistically significant at 1% (the correlation with export status is 0.06 for state-owned dummy, 0.61 for foreign-owned dummy, and 0.35 for the number of PC).

[46] We have information on number of computers in some years: from 2001 to 2005, 2008, 2010 and 2011. Therefore, we restrict this IV analysis to those years.

[47] The instrument set passes the Sargan test for overidentifying restrictions with Sargan-Hanses statistic of 3.952, p-value is 0.139.

others since they are more productive and hence can challenge their competitors in the market. Besides, one could have a question about what happens to plants that enter and exit the export market. To take a close look into the possible benefits from exporting and the changes that happen when plants enter and exit export activities, we follow the insight of Bernard and Jensen (1999) and De Loecker and Warzynski (2012), in a cross-section analysis while considering markups before and after plants enter or exit an export market in a cross-section analysis. We define *Entrant* = 1 for plants entering the export market in the middle of the period, surviving until 2013, and not switching again, 0 otherwise; *Exiter* = 1 for plants exporting from the beginning of the period but ceasing exporting during the period (and not re-entering) and 0 otherwise, *Always*=1 for plants exporting during all years and 0 otherwise. The base group is thus *Never* as plants that do not export in every year.[48] We construct the model as follows:

$$\ln\mu_{it} = \lambda_0 + \lambda_1 Entrant_i + \lambda_2 Exiter_i + \lambda_3 Always_i + \mathbf{b}'_{it}\varphi + \vartheta_{it} \tag{3.40}$$

The constant term λ_0 represents average log markups of the base group, i.e. plants that have no export activities in all years. λ_1 represents the entry effect, that captures the markup percentage difference for starters, between the post–export and pre–export entry periods. In the same way, λ_2 measures the markup percentage difference for exiters before and after exiting the export market, i.e., plants before exiting have λ_2 higher in markup percentage than after exiting export market. λ_3 represents the markup percentage difference between $Always$ and the base group $Never$, the expected sign for this coefficient is positive. We expect a positive value of λ_1, that is, plants' markup is higher when they enter an export market. Plants decide to entering an export market when their productivity is higher enough. When they are more productive, they can challenge their competitors and then experience higher markups. Empirically, De Loecker and Warzynski (2012) find out a substantially higher markups for export entry in the case of Slovenian manufacturers. Due to the limitations in reporting export status in the data, we consider various periods to regress function Equation (3.40): the whole period of 14 years (from 2000 to 2013) and two alternatives sub-periods where plants' export status is identified directly from the data (from 2002 to 2004 and from 2010).[49] We also include year dummies and industries dummies. The results are reported in Table 3.6. We find that export entrants have considerably higher markups. The estimates imply that export entry is associated with a considerable increase in markups of around 16 to 17 percent (or 0.18 to 0.3 in level) according to the period considered. Moreover, we also observe that markups of plants before exiting is greater than after exiting an export market. In other word, it indicates the reduction of markups when plants stop exporting. Markups of $Always$ group is higher than that of never-export plants, especially when considering only the period 2010-2013.

To sum up, in this section, our main finding is that exporters have higher markup than non-exporting producers, supporting the results of De Loecker and Warzynski (2012). It is because on the one hand, plants have had success in their activities, so they have charged higher markups before exporting. On other hand, plants experience higher markups when entering the export market.

[48] We exclude plants switching export status more than once.

[49] For the whole sample, unfortunately, we observe no plants exporting in all 14 years. For each sub-period, we re-estimate markups and then regress Equation (3.40).

Table 3.6: Markups before and after export market entry/ exit

VARIABLES	Whole period (1)	2002-2004 (2)	2010-2013 (3)
Entry effect	0.172***	0.156***	0.171***
	(0.007)	(0.019)	(0.008)
Exit effect	0.049	0.036*	0.090***
	(0.031)	(0.022)	(0.018)
Always		0.026**	0.110***
		(0.012)	(0.007)
Constant	0.557***	0.145***	0.320***
	(0.012)	(0.016)	(0.010)
Observations	135,738	20,853	73,108
Control variables	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

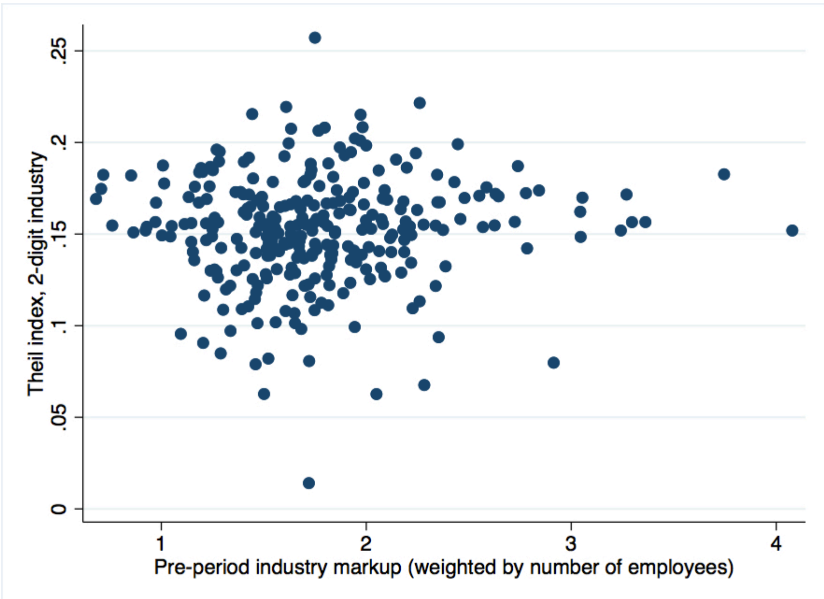


Figure 3.6: Correlation between markup dispersion and pre-period markup (2-digit industry), 2000-2013

3.3.4 Does trade liberalisation reduce markup dispersion?

Beside markup levels as such, the second variable of interest in this research is markup dispersion. Figure 3.6 represents the correlation between the markup dispersion, the Theil index for 2-digit industries and their average markup in the previous year. Overall, it seems that there is no correlation between these two variables, which means that they have different underlying factors which are likely to be uncorrelated.

We use the average MFN ad-valorem tariffs from the WITS database. Figure 3.7 represents the correlation between tariffs changes during 2006-2013 and tariffs in 2006. Except for the case of tobacco products, there is a negative correlation between tariffs in 2006 and their changes in

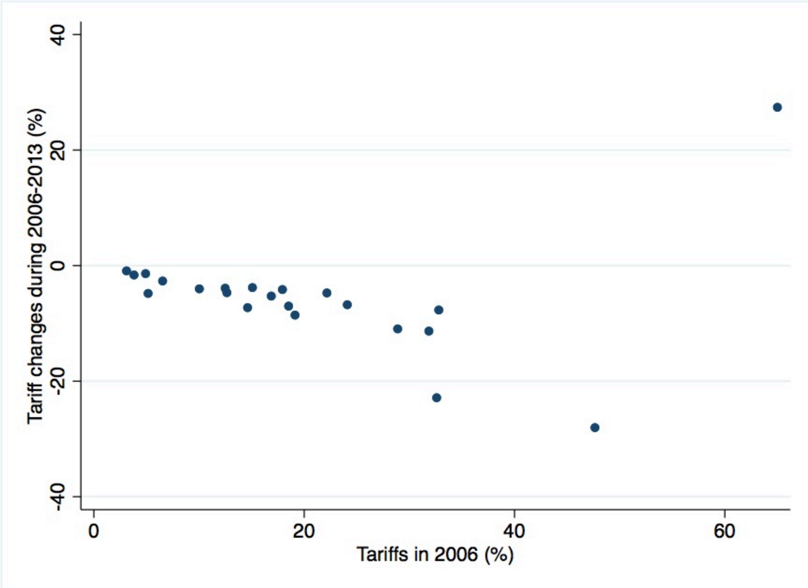


Figure 3.7: Correlation between tariff changes during 2006-2013 and tariffs in 2006

2013 relative to 2006. In other words, industries with higher tariffs in 2006 experience greater reduction of tariffs by 2013.

To investigate the impact of trade liberalisation (represented by the accession of Vietnam to WTO on January 2007) on markup dispersion, we classify industries into two groups: the treatment group which was more protected before WTO accession is subjected to a higher level of tariff reduction; the control group which was more open previously and experienced less tariff reduction after WTO accession. Our objective is to evaluate the impact of trade liberalisation on markup dispersion by comparing the markup dispersion between these groups of industries before and after the intervention (the trade liberalisation in 2007). To do so, we apply the difference-in-differences methodology where the treatment is continuous (Guadalupe and Wulf, 2010). The model specification is then:

$$MD_{it} = \alpha Tariff06_i \times WTO_t + \mathbf{X}'_{it}\delta + \tau_i + \eta_t + \varepsilon_{it} \tag{3.41}$$

where MD_{it} is markup dispersion of industry i at time t, $Tariff06_i$ is most-favoured-nation (MFN) tariff rate of industry i in 2006, WTO_t is a dummy variable which equals 1 from 2007, 0 otherwise; $\mathbf{X}'_{it}$ is a vector of control variables that may affect markup dispersion, τ_i is industry fixed-effect, η_t is year dummies. Our coefficient of interest is α, it reflects the mean change of markup dispersion from before and after between the treatment group and the control group. In other words, it captures the average effect of trade liberalisation on industry markup dispersion. We expect that the trade liberalisation through WTO accession leads to a greater decline in markup dispersion in industries with high tariffs than in low-tariff industries. Therefore, α is expected to be negative. In this specification, we define an industry at the two-digit of the Vietnam Standard Industry Classification.

To gauge the level of exposure of industry to trade liberalisation, we use the tariff rate in 2006 only. However, we also use the average of MFN tariff rate until 2006, $AvT06_i$ for robustness check.

As for control variables, we use time-varying covariates that may affect the industry markup dispersion such as number of plants, average fixed assets and the industrial agglomeration degree. The industrial agglomeration degree is the index of industry geographical concentration developed by Ellison and Glaeser (1997) (hereafter, EG index) and defined as

$$EG_i \equiv \frac{G - \left(1 - \sum_j x_j^2\right) H}{\left(1 - \sum_j x_j^2\right)(1 - H)} = \frac{\sum_{j=1}^{M}(s_j - x_j)^2 - \left(1 - \sum_j^{M} x_j^2\right)^2 \sum_{k=1}^{N} z_k^2}{\left(1 - \sum_j^{M} x_j^2\right)\left(1 - \sum_k^{N} z_k^2\right)} \tag{3.42}$$

where EG_i is the index of geographical concentration of industry i, s_j (with $j = 1, ..., M$), denote shares of industry i's employment in each of M geographic areas, x_j are shares of total employment in each of those areas, so, $G \equiv \sum_{j=1}^{M}(s_j - x_j)^2$ is defined as a measure of "raw" geographic concentration of industry i. z_k ,where $k = 1, ..., N$, is the kth plant's share of the industry whole employment, $H \equiv \sum_k^N z_k^2$ is the Herfindahl index of the industry plant size distribution. As noted in Ellison and Glaeser (1994), EG_i is the probability that a given pair of plants choose their locations jointly or is a measurement of the importance of advantage in location choice. A higher EG_i indicates that industry i is more geographically concentrated.

Some remarks should be noted in the specifications. First, since Vietnam acceded to the WTO in early 2007 after a long negotiation process, one should be concerned that plants could have anticipated this event and changed their behaviour before 2007. To address this concern, we re-perform Equation (3.41) by assuming that the intervention time is one year before WTO accession, in other word, we replace WTO_t by WTO_{t-1} which equals 1 from 2006, 0 otherwise.

Second, using the difference-in-differences method usually could leave the researcher with an essential concern related to the possible correlation between the pre-existing trends and changes in the relevant variable. There is maybe a spurious correlation in the estimates presented in the previous section in cases where the industry-level trends in markups dispersion are correlated with the measures of trade liberalisation (Topalova, 2010). Therefore, we also perform a placebo test by investigating the effect of tariffs on markup dispersion before WTO accession.

Third, the period studied covers not only the time when Vietnam acceded to the WTO but also the 2008 financial crisis. To isolate the potential effect of the crisis on markup dispersion, we use an additional variable in the form of an interaction term between financial crisis and plant's liquidity. The latter is calculated as the industry average of short-term assets over total assets.

Main results

We firstly illustrate in Figure 3.8 the change in markup dispersion relative to 2000 of the control and treatment group to explore their markup dispersion trajectories before and after the intervention (the Vietnam's accession to WTO in 2007). As shown in the figure, the markup dispersion of these two groups seem to have a similar motion before WTO accession. This parallel pre-treatment trend in markup dispersion between treatment and control groups implies the graphical satisfaction of the difference-in-differences identifying assumption and mitigate the issue of ex-ante incomparability between the two groups. It is worth noting that the WTO accession of

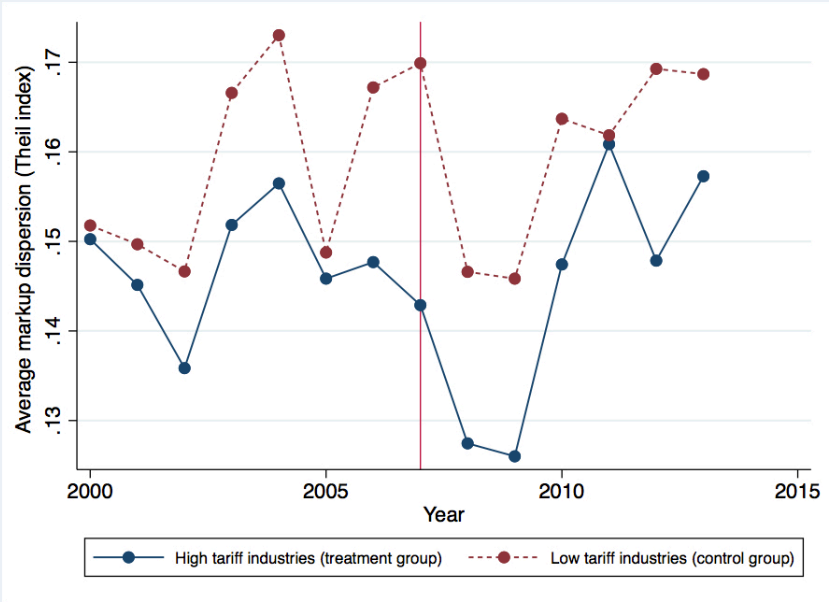


Figure 3.8: Markup dispersion evolution for high and low tariff industries

Vietnam took place in early 2007 while the Census was collected in late 2007. So, a plants' behaviour in 2007 could be influenced (at least partly) by this event. Even though it is not very clear from the figure, the divergence to some extent of the trajectories since 2007 implies that the control group and the treatment group might react differently in terms of markup dispersion following the trade liberalisation. Moreover, since the markup dispersion evolution increases after the financial crisis, it makes the inclusion of the financial crisis variable in our model more reasonable. The figure is merely a visual presentation and superficial view of the question at hand, therefore we rely heavily on other tests to study the direction of markup dispersion.

Table 3.7 reports the main results of Equation (3.41). All regressions include year dummies, the standard errors are clustered at the industry level to deal with the possible issue of heteroskedasticity and serial autocorrelation (Bertrand, Duflo, and Mullainathan, 2004). Since the studied period covers the financial crisis in 2008, one may be concerned that this event could affect markup dispersion. To account for such an impact, we follow Ha et al. (2016) by adding the interaction $Liquid07 \times crisis$, where $Liquid07$ indicates average liquid ratio of an industry in 2007, $crisis$ equals 1 since 2008, 0 otherwise.

The first column shows the difference-in-differences specification without any covariates. The coefficient of interest corresponding to that with the interaction $Tariff06_i \times WTO_t$ is statistically significant and negative. This means that industries with high tariffs before trade liberalisation (the treatment group) have a decreasing markup dispersion as there is a higher level of competitiveness afterwards due to the decline in tariffs in the markets in which the plants operate. For instant, an industry with a higher tariff rate in 2006, for example a manufacturer of wearing products with tariffs of 47.65%, decreased its markup dispersion by 0.28 (0.4765×0.592) following trade liberalisation. In other words, this is evidence that the WTO accession of Vietnam has had a negative impact on industry's markup dispersion.

The results are still consistent when we control for time-varying industry characteristics that may correlate with markup dispersion and trade liberalisation, such as number of plants, average fixed assets and the industrial agglomeration degree (the EG index) (Column 2). Besides,

Table 3.7: Markup dispersion and trade liberalisation

VARIABLES	Dependent variable: Theil-index, in log			
	(1)	(2)	(3)	(4)
$Tariff06_i \times WTO_t$	−0.592**	−0.451**	−0.709***	−0.539***
	(0.247)	(0.211)	(0.193)	(0.173)
Number of plants (in log)		0.074		0.055
		(0.071)		(0.064)
Average industry's assets (in log)		0.000		0.004
		(0.036)		(0.035)
Value added share of SOE		−0.105		−0.049
		(0.146)		(0.130)
The EG index		0.416**		0.403*
		(0.193)		(0.206)
$Liquid07 \times crisis$			−1.201*	−0.618
			(0.642)	(0.522)
Constant	−1.915***	−2.254***	−1.918***	−2.231***
	(0.037)	(0.380)	(0.037)	(0.381)
Observations	303	289	303	289
R-squared	0.153	0.231	0.176	0.240
Number of industries	22	22	22	22
Industry fixed effect	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

there is evidence that an industry's geographic concentration index may affect markup dispersion as its coefficient is statistically significant. The positive value indicates that the more geographically concentrated the industry, the higher the markup dispersion.[50] The entry barriers (number of plants, average fixed assets) and the value added share of SOE have no effect on markup dispersion according to their statistically insignificant coefficients.

We also check the effect of the financial crisis on markup dispersion by adding the interaction between liquidity ratio and crisis dummy (Columns 3 and 4). The negative impact of trade liberalisation on markup dispersion is robust to the influence of the financial crisis. Moreover, the financial crisis' coefficient is statistically significant without controlling for the industries' characteristics. This result implies an impact of the crisis on markup dispersion and could explain a short period of similar trend (2008-2010) of both groups after the crisis, as shown in Figure 3.8.[51]

Diagnostics for parallel trends

The difference-in-differences methodology is performed under the parallel trend assumption without which one cannot obtain credible impact estimates. According to the assumption, if the intervention did not happen, the outcome in the treatment group would have progressed similarly to the outcome in the control group. If the trends were different in the treatment and control

[50] However, when we use the Herfindhal-Hirschman Index, it is not statistically significant.

[51] It should be noted that in Figure 3.8, we compute the unweighted average markup dispersion of industries, that means, we do not account for industries' characteristics.

Table 3.8: Diagnostics for parallel trends

VARIABLES	Theil index (in log) (1)	Theil index before WTO accession (in log) (2)	(3)
$Tariff06_i \times WTO_{t-1}$	−0.335		
	(0.216)		
Tariff from WTO accession		−0.004	
		(0.004)	
Tariff before WTO accession			0.009
			(0.019)
Constant	−2.338***	−1.230	−1.676
	(0.421)	(0.785)	(1.206)
Observations	289	142	122
R-squared	0.215	0.105	0.109
Number of industries	22	21	21
Control variables	Yes	Yes	Yes
Industry fixed effect	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

groups before the intervention, the estimation of treatment effect from difference-in-differences technique would be biased or invalid. Therefore, it is necessary to test the validity of the equal trends assumption.

The first way to test the validity of the difference-in-differences procedure is using a ”placebo” year of intervention. With this setting, we want to make sure that trade liberalisation is unanticipated. Since Vietnam acceded to the WTO in early 2007, it raises the concern that plants might have anticipated the WTO accession and made changes to their behaviour before this event. To address this issue, we choose 2006 as a fake year of intervention. That means, we use the interaction $Tariff06_i \times WTO_{t-1}$ instead of $Tariff06_i \times WTO_t$ and keep the set of control and treatment groups. This new variable would take into account a distinct innovation/shock before it happened if plants anticipated the liberalisation or there was a pre-existing trend. In case of no impact, plants did not anticipate the trade liberalisation and they only started to act from 2007. According to column 1 of Table 3.8, the coefficient is statistically insignificant. This consolidates the fact that trade liberalisation was not anticipated by plants and that the industry markup dispersion only reduced after trade liberalisation in early 2007.

The relationship between the tariff rate and markup dispersion would be biased if the change in tariff is correlated with omitted time-varying factors that affect the industry markup dispersion. To address this issue, we follow Edmonds, Pavcnik, and Topalova (2010); Topalova (2010) by examining whether tariffs changes had similar pre-existing trends in markup dispersion. If there is such a correlation, the difference-in-differences coefficient estimated in the previous section might be spurious. For instance, if an industry, whose markup dispersion has reduced faster before trade liberalisation, experienced a (continuing) faster markup dispersion reduction after trade reform and a significant tariffs cut, there would be an unobservable factor that leads to both tariffs change and markup dispersion reduction. Therefore, we perform a test of whether changes in tariffs from 2007 are correlated with changes in industry's markup dispersion before 2007. If there is no such correlation, the estimates in (Table 3.7) are credible. Column 2 of Table 3.8 shows that the relevent coefficient is statistically insignificant, as expected. This rejects the hypothesis that the trade liberalisation is correlated with the pre-existing trends in markup dispersion.

Another way to deal with the above problem is by testing the correlation between trade liberalisation and markup dispersion before the intervention. Indeed, if there is an effect , this could be evidence of some underlying confounding factors. As shown in column 3 of Table 3.8, tariffs have no impact on markup dispersion before 2007.

Our three falsification tests confirm the validity of the difference-in-differences procedure. This consolidates the effect of trade liberalisation in reducing industry markup dispersion. In the next section, we perform various robustness tests with the alternative measures of markup dispersion and industries' exposure to trade liberalisation.

Robustness checks

First, we perform some robustness tests by using different measuring of markup dispersion and trade liberalisation. The results are represented in Table 3.9.

In columns 1-4, we report the results with four alternative measurements of markup dispersion: Gini index, the relative mean deviation (RMD), the mean log deviation (MLD) and the standard deviation (SD). The coefficients of interest are still statistically significant and negative. Therefore, the effect of trade liberalisation on industries markup dispersion is consistent.

Then, we check the effect by using different tariffs to proxy the industrial exposure to trade liberalisation. Specifically, we use average tariffs during 2001-2006 and the tariff in 2001 instead of the tariff in 2006. As shown in columns 5-8, the estimates are clearly robust.

Second, we use the translog production function as an alternative way to estimate markups and markup dispersion. Different from the Cobb-Douglas production function, the translog model allows plants to have their own output elasticity of labour input. Therefore, the translog model is considered as a flexible specification of production function. Once having obtained the markup dispersion, we re-estimate Equation (3.41) and report the results in Table 3.A.2 in Appendix 3.A. The difference-in-differences estimates are still statistically significant and negative. In other words, there is evidence that trade liberalisation reduces markup dispersion in industries. The financial crisis, which occurred after trade liberalisation, seems to have no effect on markup dispersion .

As in the case of the Cobb-Douglas production function, we examine the validity of the ”equal” trend assumption, that is, the markup dispersion trends in high tariff and low tariff industries groups are parallel. We re-perform three placebo tests to check (i) whether trade liberalisation is unanticipated, (ii) whether tariff rates are correlated with pre-existing trends of markup dispersion, and (iii) whether tariffs rates and markup dispersion are correlated in a prior period. The results are represented in Table 3.A.2b in Appendix 3.A. In most cases, the placebo tests support the assumption that treatment and control groups have the same markup dispersion trends before WTO accession.

While using different measures of markup dispersion and industry's exposure to trade liberalisation for the translog model, we also obtain the robust effect that trade liberalisation reduces markup dispersion of industries (Table 3.A.2c in Appendix 3.A).

Table 3.9: Robustness checks with alternative measuring

VARIABLES	Gini-index	RMD	MLD	SD
	(1)	(2)	(3)	(4)
$Tariff06_i \times WTO_t$	-0.230**	-0.217**	-0.448**	-0.417***
	(0.091)	(0.093)	(0.184)	(0.118)
Constant	-1.389***	-1.036***	-2.286***	-0.288
	(0.194)	(0.214)	(0.374)	(0.335)
Observations	289	289	289	289
R-squared	0.223	0.186	0.228	0.813
Control variables	Yes	Yes	Yes	Yes
Industry fixed effect	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
VARIABLES	Theil index			
	(5)	(6)	(7)	(8)
$AvT07_i \times WTO_t$	-0.591**	-0.451**		
	(0.245)	(0.208)		
$Tariff01_i \times WTO_t$			-0.460*	-0.459**
			(0.235)	(0.201)
Constant	-1.915***	-2.260***	-1.907***	-2.287***
	(0.037)	(0.379)	(0.038)	(0.380)
Observations	303	289	294	286
R-squared	0.153	0.232	0.217	0.241
Control variables	No	Yes	No	Yes
Industry fixed effect	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Heterogeneous effect of trade liberalisation

In this subsection, we aim to investigate the heterogeneous effect of trade liberalisation on markup dispersion between different groups of industries divided by the four alternative criteria: industry labour intensity (as the ratio between total wages and value added), share of number of state-owned producers (SOE) in an industry, share of foreign-owners (FDI) and share of number of inland-plants. These variables are taken for 2006. We use the median to differentiate between industries with higher and lower shares of number of state-owned plants, for instance. More specifically, we define D_i^{l06} as equal to 1 if the industry stays in the top 50% for labour intensity , 0 in the bottom 50% . In the same way, dummy D_i^{soe06} equals 1 for industries with high concentration of state-owned plants, 0 otherwise; $D_i^{fdi06} = 1$ for industries with high numbers of foreign-owned plants, 0 otherwise; and $D_i^{ild06} = 1$ implies an industry with a higher number of in-land plants, 0 otherwise. To depict the heterogeneous treatment effect, we use the triple differences by introducing, for example D_i^{l06}, into Equation (3.41) (dummies remaining can be introduced in the same logic):

Table 3.10: Heterogeneous treatment effect

VARIABLES	Theil index			
	High/low labour intensity (1)	SOE/non-SOE (2)	FDI/non-FDI (3)	Inland (4)
$Tariff06_i \times WTO_t$	−0.606***	−0.299*	−0.640***	−0.508***
	(0.178)	(0.157)	(0.164)	(0.170)
$Tariff06_i \times WTO_t \times D_i^{l06}$	0.120			
	(0.155)			
$Tariff06_i \times WTO_t \times D_i^{soe06}$		−0.439**		
		(0.185)		
$Tariff06_i \times WTO_t \times D_i^{fdi06}$			0.240	
			(0.160)	
$Tariff06_i \times WTO_t \times D_i^{ild06}$				0.303
				(0.195)
Constant	−2.197***	−2.209***	−2.185***	−2.117***
	(0.398)	(0.352)	(0.339)	(0.377)
Observations	275	275	289	289
Number of industries	21	21	22	22
Control variables	Yes	Yes	Yes	Yes
Industry fixed effect	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

$$MD_{it} = \alpha_0 Tariff06_i \times WTO_t + \alpha_1 Tariff06_i \times WTO_t \times D_i^{l06} + \mathbf{X}'_{it}\delta + \tau_i + \eta_t + \varepsilon_{it} \tag{3.43}$$

The coefficient α_1 captures the difference between change in expected outcome (markup dispersion) from before to after trade liberalisation in a high labour intensive industry and that of a low labour intensive industry. Table 3.10 reports the results of the four criteria. Firstly, the difference-in-differences coefficients, α_0, are still negative and statistically significant in all cases, and this confirms the negative effect of trade liberalisation on markup dispersion. Hence, we could compare this effect on different groups of industries. Only when we split industries according to their level of SOEs, is the triple difference coefficient statistically significant, this implies that there is a heterogeneous effect of trade liberalisation between industries with higher numbers of state-owned plants and the counterfactual. The negative sign indicates that following trade liberalisation, markup dispersion reduces more in industries in which there are relatively more state-owned plants. Our result agrees with the findings of Lu and Yu (2015) who show an effect of trade liberalisation on markup dispersion for SOEs and no effect for non-SOEs in the case of China. However, no difference in the impact of trade liberalisation is detected for the three remaining cases.

3.4 Conclusion

Misallocation of inputs is one of the factor that explain the difference in TFP across a country. Reducing trade barriers is proved to mitigate misallocation and then increase aggregate productivity. Instead of examining directly the effect of trade liberalisation on misallocation, we investigate empirically the effect of trade liberalisation on industries' markup dispersion, one of the sources of inefficient resources allocation.

Using rich plant-level data of Vietnamese manufacturing during the period 2000-2013, we apply the methodology of De Loecker and Warzynski (2012) to estimate variable markups of heterogeneous production units. Over 14 years, we observe a downward trend of the aggregate markup and a reduction of markup dispersion. This implies that higher competition leads plants to set prices in favour of the customer, and brings about a better market share reallocation between Vietnamese manufacturers. The decreasing aggregate markups are mostly due to the change in markup level within industries (or within plants). Moreover, the reallocation between plants, even if its contribution is minor, could improve the aggregate markup.

To explore the role of trade on plant pricing, we firstly compare markups according to plant export status. We find a positive correlation between a plant's markup and its export status. In the case of Vietnamese manufacturing, this relationship could be explained by not only the self-selection process of exporters (producers with prior success tend to export) but also by the learning-by-exporting process (producers entering the export market charge higher markups and experience higher markup change). Then, examining the relationship between trade and markup dispersion, we point out that trade liberalisation could reduce two-digit industry markup dispersion. This reduction is more important for industries with a higher proportion of state-owned producers.

Appendix

3.A Tables

Table 3.A.1: Output elasticities of inputs (translog model)

VSIC	Industry	L	K	RS
15	Food products and beverage	0.831	0.268	1.099
16	Tobacco products	0.849	0.207	1.057
17	Textiles	0.507	0.179	0.686
18	Wearing products	0.828	0.083	0.911
19	Leather products	0.808	0.078	0.885
20	Wood and cork manufacturing	0.724	0.199	0.923
21	Paper products	0.613	0.091	0.705
22	Publishing, priting, recording media	0.731	0.148	0.880
23	Manufacture of coke, refined petroleum products and nuclear fuel	0.902	0.335	1.237
24	Chemical manufacturing	0.717	0.149	0.866
25	Rubber and plastics products	0.633	0.159	0.792
26	Other non-metallic mineral products	0.859	0.220	1.078
27	Basic metals manufacturing	0.793	0.296	1.089
28	Fabricated metal products, except machinery and equipment	0.623	0.119	0.742
29	Machinery and equipment	0.747	0.109	0.855
30	Office, accounting and computing machinery	0.569	0.356	0.925
31	Electrical machinery and apparatus	0.818	0.225	1.043
32	Radio, television and communication equipment	0.807	0.267	1.074
33	Medical, precision and optical instruments, watches and clocks	0.843	0.184	1.027
34	Motor vehicles, trailers and semi-trailers	0.820	0.320	1.140
35	Other transport equipment	0.908	0.165	1.073
36	Furniture	0.552	0.069	0.620

Table 3.A.2: Markup dispersion and trade liberalisation using translog production function

(a) Main results

VARIABLES	Dependent variable: Theil-index, in log			
	(1)	(2)	(3)	(4)
$Tariff06_i \times WTO_t$	−0.644***	−0.561***	−0.582***	−0.448***
	(0.148)	(0.186)	(0.103)	(0.144)
Number of firms (in log)		0.003		0.025
		(0.042)		(0.042)
Average industry's assets (in log)		0.067		0.059
		(0.050)		(0.045)
Value added share of SOE		−0.155		−0.212
		(0.155)		(0.163)
The EG index		−0.047		−0.028
		(0.218)		(0.225)
Liquid07 × crisis			0.754	0.848
			(0.545)	(0.585)
Constant	−1.685***	−2.236***	−1.683***	−2.247***
	(0.046)	(0.500)	(0.046)	(0.449)
Observations	302	289	302	289
R-squared	0.184	0.235	0.202	0.255
Number of industries	22	22	22	22
Industry fixed effect	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

(b) Diagnostics for parallel trends

VARIABLES	Theil index (in log)	Theil index before WTO accession (in log)	
	(1)	(2)	(3)
$Tariff06_i \times WTO_{t-1}$	−0.402		
	(0.264)		
Tariff from WTO accession		0.002	
		(0.003)	
Tariff before WTO accession			0.001
			(0.015)
Constant	−2.346***	−1.020**	−1.637***
	(0.506)	(0.395)	(0.525)
Observations	289	142	122
R-squared	0.203	0.148	0.147
Number of industries	22	21	21
Control variables	Yes	Yes	Yes
Industry fixed effect	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes

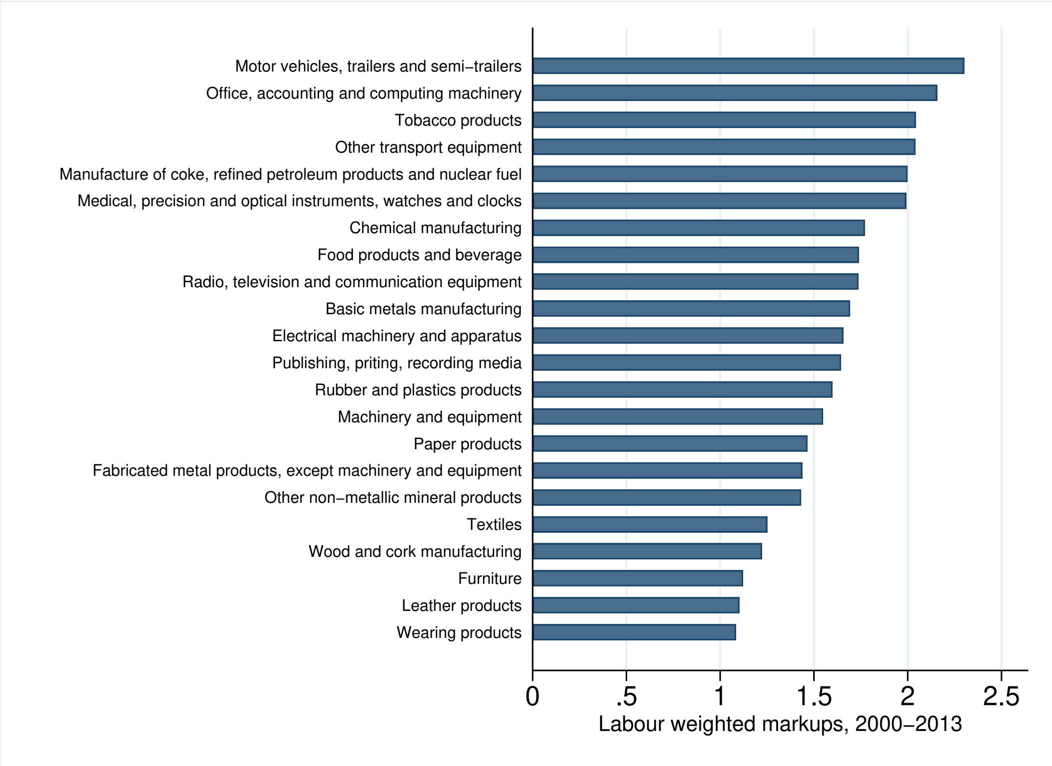
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

(c) Robustness checks with alternative measuring

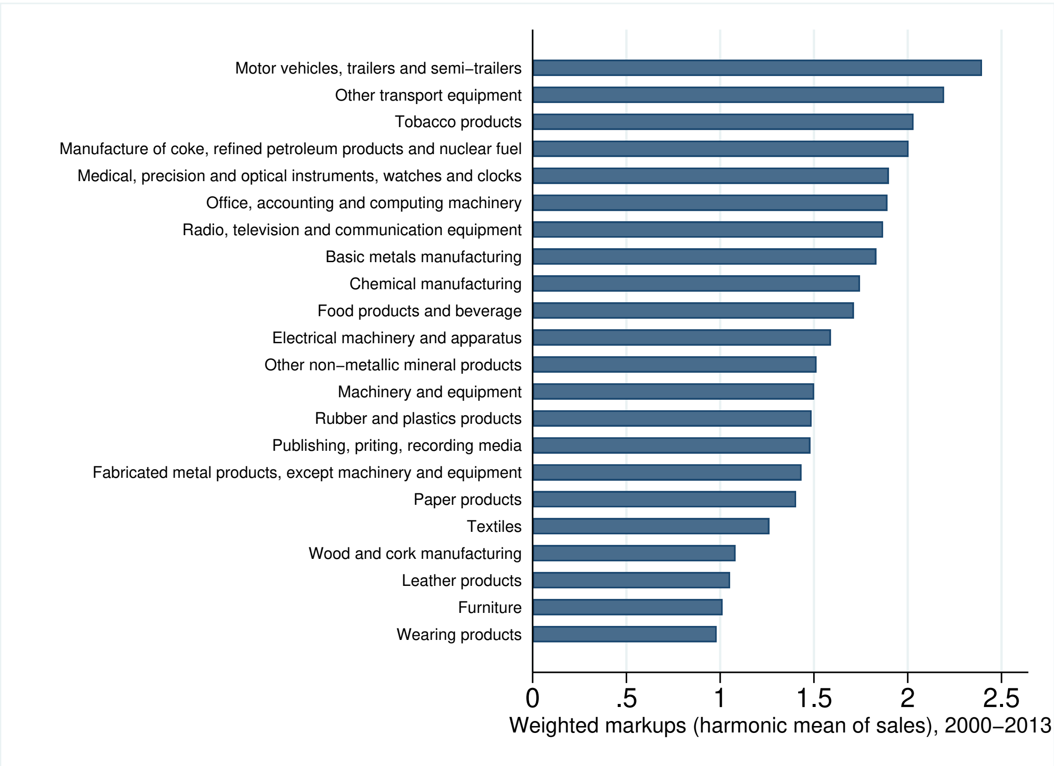
VARIABLES	Gini-index	RMD	MLD	SD
	(1)	(2)	(3)	(4)
$Tariff06_i \times WTO_t$	-0.232**	-0.205	-0.472**	-0.311**
	(0.106)	(0.139)	(0.189)	(0.121)
Constant	-1.325***	-0.975***	-2.065***	-0.087
	(0.235)	(0.256)	(0.447)	(0.221)
Control variables	Yes	Yes	Yes	Yes
Observations	289	289	289	289
R-squared	0.194	0.146	0.189	0.242
Industry fixed effect	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes
VARIABLES		Theil index		
	(5)	(6)	(7)	(8)
$AvT07_i \times WTO_t$	-0.645***	-0.563***		
	(0.149)	(0.185)		
$Tariff01_i \times WTO_t$			-0.636***	-0.568***
			(0.148)	(0.179)
Constant	-1.685***	-2.243***	-1.699***	-2.275***
	(0.046)	(0.497)	(0.046)	(0.492)
Control variables	No	Yes	No	Yes
Observations	303	289	294	286
R-squared	0.185	0.236	0.198	0.244
Industry fixed effect	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

3.B Figures



(a) (Arithmetic) Employment weighted markup



(b) Cost weighted markup (of harmonic mean of sales weighted markup)

Figure 3.B.1: Estimated mean markups for 2-digit VSIC 1993 manufacturing industries

Appendix A

Data

Data description

In this appendix, we describe the plant-level data used in more detail. The data on the manufacturing sector is taken from the Vietnamese Enterprise Survey collected annually during the 14 years from 2000-2013 and provided by the General Statistics Office of Vietnam (VGSO).[52] The unit of observation is a plant. This is the unit we use throughout the thesis. It is an unbalanced panel including every plants in each year. Therefore, we can track the entering and exiting plants. We do not account for plants dropping out of the dataset and showing up again. Investigating the source of aggregate productivity growth (Chapter 2) and aggregate markup change (Chapter 3) between two points in time, noted as $t = 1$ and $t = 2$, we adopt the definition of entering and exiting plants provided by Olley and Pakes (1996) and Melitz and Polanec (2015). Accordingly, exiting plants designate plants that are present at $t = 1$ but not at $t = 2$; entering plants are plants present only at $t = 2$.

In any given year, the dataset covers all registered plants with 10 employees or more and a representative random sample of small registered plants with less than 10 employees. We only account for plants with 10 employees or more and those reporting information on sales, value added, employment, capital, intermediate inputs and wage bill. We end up with an unbalanced panel of 197,361 observations over the sample period and number of plants grew over time (Table A.1).

We deflate all monetary variables by the Producer Price Index in the manufacturing sector provided by the VGSO. The main variables used in the thesis are: value added in dongs (VA), number of employees in a given year (L), total fixed assets in dongs (K), intermediate inputs in dongs (M), total cost of employees in dongs (wage bill). As the dataset does not provide any information on value added, we compute this variable as the sum of the wage bill, depreciation and pre-tax profit. This is a relaxed definition of value added, because we do not have data on interest paid on credit and loans, which is part of the payments to capital (Ha and Kiyota, 2014).

[52]To request access to the census data, researchers should contact the General Statistics Office of Vietnam at the address: 54 Nguyen Chi Thanh Street, Dong Da District, Hanoi, Vietnam; tel: +84(0)2473046666, email: banbientap@gso.gov.vn.

Table A.1: Plants

Year	2000	2001	2002	2003	2004	2005	2006
Number of plants	4,896	6,062	7,223	8,229	9,620	10,801	11,611
Year	2007	2008	2009	2010	2011	2012	2013
Number of plants	13,229	18,212	18,895	19,894	22,991	23,425	22,273

Table A.2: Summary statistics

Variables	Obs. (1)	Mean (2)	Median (3)	Std. Dev. (4)	Min (5)	Max (6)
Value-added (millions of dongs)	197,361	7,394	616.34	88,819	0.177	31,256,728
Number of employees	197,361	180	35	829	10	84,660
Capital (millions of dongs)	197,361	11,023	892.64	94,628	0.0441	18,575,066
Intermediate inputs (millions of dongs)	197,361	37,731	3,058	572,188	0.0456	194,394,064

Note: Authors' calculation from VGSO database.

Appendix B

Production function estimation methods

In this appendix, we discuss further two methods in the estimation of production functions used in the thesis, those of Levinsohn and Petrin (2003) (hereafter, LP) in Chapter 2 and Ackerberg et al. (2015) (hereafter, ACF) in Chapter 3. Since the procedure in estimating production function is detailed in Chapter 2 and 3, in this appendix, we focus on the comparison of the two methods as well as their strengths and weaknesses. Ackerberg et al. (2015) adopt the method of Levinsohn and Petrin (2003) in using the intermediate input as a proxy for unobserved productivity. The main difference, as we will discuss below, is that they relax the assumption of a firm's labour choice.

The two methods follow the most recent approach pioneered by Olley and Pakes (1996) (hereafter, OP) to eliminate the endogeneity issue when estimating production function. Indeed, let us consider the following production function:

$$y_{it} = f(l_{it}, k_{it}; \boldsymbol{\beta}) + \omega_{it} + \epsilon_{it} \tag{B.1}$$

where y_{it}, l_{it} and k_{it} denote (log) output, labour and capital respectively; $\boldsymbol{\beta}$ is the vector of production function coefficients to be estimated; and ϵ_{it} is a pure error term that is unobserved by plants when they decide the input choices, hence uncorrelated with l_{it} and k_{it}. The term ω_{it} denotes productivity shocks, that is unobserved by econometricians but predictable by plants during their decision-making process. This term raises the endogeneity problem, one of the most well-known challenges in estimating the production function: the plant's input choice is correlated with the productivity shock. In that case, the OLS estimates of input coefficients (labour, capital) are biased.

To eliminate the endogeneity issue, the authors propose using a demand function to proxy productivity shocks. As in the original work of Olley and Pakes (1996), the methods of Levinsohn and Petrin (2003) and Ackerberg et al. (2015) make the same assumptions on: (i) the first-order Markov process of the productivity shocks, ω_{it}, which means, $p(\omega_{it+1}|I_{it+1}) = p(\omega_{it+1}|\omega_{it})$; and (ii) the law of motion of capital, $k_{it} = (1-\delta)k_{it-1} + i_{it-1}$ where i_{it-1} is plant i's investment at $t-1$.

The first assumption about the productivity shocks shows that the expected productivity, ω_{it+1}, is based on the current set of information, I_{it}, that consists of current and past productivity and not the future, ω_{it+1}. Therefore, plant's decisions at t are uncorrelated with the unanticipated innovation in ω_{it+1} between t and $t+1$, denoted ξ_{it+1} with $\xi_{it+1} = \omega_{it+1} - E[\omega_{it+1}|I_{it}]$ and $E[\xi_{it+1}|I_{it} = 0]$.

The second assumption about capital choice implies that a plant's capital stock at t is actually decided at $t-1$. It belongs to I_{it-1} and is hence uncorrelated with the unanticipated innovation ξ_{it}. In other words, we have:

$$E[\xi_{it}k_{it}] = 1 \tag{B.2}$$

Then, these approaches identify a control function in which ω_{it} is the only econometrically unobservable term. The function must be strictly monotonic in ω_{it}. Different from the pioneering work of OP, LP propose intermediate input, m_{it}, as a control function for ω_{it}. A plant's intermediate input demand function must be monotonic in ω_{it} for all k_{it} and is given as $m_{it} = f_t(k_{it}, \omega_{it})$.

The main advantage of LP's implementation is that it is suitable for many datasets in developing countries in which the OP proxy (investment demand function) can be zero for many plants in some years and hence unavailable. Indeed, in our data, only 15% of total plants per year have any information about investment. The positive intermediate inputs are reported for over 80% of plants per year.

However, this method has two inconvenient points that are related to the first step of the procedure. Recall that the first stage involves the estimation of the output elasticity of labour and a non-parametrical function, called $\phi_t(k_{it}, m_{it})$. Indeed, by replacing $\omega_{it} = f_t^{-1}(k_{it}, m_{it})$ in the production function, Equation (B.1) becomes:

$$y_{it} = f(l_{it}, k_{it}; \boldsymbol{\beta}) + f_t^{-1}(k_{it}, m_{it}) + \epsilon_{it} \tag{B.3}$$

$$= f(l_{it}; \boldsymbol{\beta}) + \phi_t(k_{it}, m_{it}) + \epsilon_{it} \tag{B.4}$$

where $\phi_t(k_{it}, m_{it})$ is the non-parametrical function and can be approximated by the high-order polynomial of k_{it} and m_{it}. The fact that one can estimate the output elasticity of labour in the first stage comes from the assumption about the intermediate input demand, *i.e.*, $m_{it} = f_t(k_{it}, \omega_{it})$. The right-hand side does not contain l_{it}, meaning that m_{it} does not depend on l_{it}. Moreover, this function implies that m_{it} is chosen after plants have information about ω_{it}. In brief, l_{it} and m_{it} are assumed to be non-dynamic variables and chosen simultaneously at t, after plants have information about ω_{it}. Therefore, l_{it} does not be included in the non-parametrical function and will be estimated in the first step.

The first inconvenient point of this implimentation is that it is difficult to apply it for a more general specification of the production function, such as the translog model. Indeed, it is unlikely that l_{it} will be estimated separately from the interaction term $l_{it}k_{it}$. Second, ACF point out the identification problems with LP's method that are related to the functional dependence issue in the first step of the procedures. If labour is fully determined by the value of k_{it}, m_{it}, the contribution of labour to output cannot be separately identified from the non-parametric function. To eliminate this problem, they propose relaxing the assumption that l_{it} and m_{it} are chosen simultaneously at t, after plants have information about ω_{it}. They allow l_{it} to be chosen at $t, t-1$ or some point in between. In this case, m_{it}, which is chosen at t, will depend on l_{it} decided in advance. Therefore, the plant's intermediate input demande function becomes:

$$m_{it} = m_t(k_{it}, l_{it}, \omega_{it}) \tag{B.5}$$

The monotonicity assumption allows to invert the above equation to obtain ω_{it}:

$$\omega_{it} = m_t^{-1}(k_{it,it}, m_{it}) \tag{B.6}$$

The production function estimation is given as:

$$y_{it} = f(l_{it}, k_{it}; \boldsymbol{\beta}) + m_t^{-1}(k_{it}, l_{it}, m_{it}) + \epsilon_{it} \tag{B.7}$$

It is not possible to differentiate the β from the coefficients of k_{it} and l_{it} in the function m_t^{-1}. This implies that one cannot obtain the output elasticity of labour in the first stage. In other words, the vector of coefficients $\boldsymbol{\beta}$ will be obtained in the second stage. This is the crucial difference between the methods of LP and ACF. The latter not only provides a dynamic implication about l_{it} but also is suitable for using the translog model specification when all coeffcients are estimated in the second stage.

References

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

Aghion, P., and Howitt, P. (1992). A model of growth through creative destruction. *Econometrica: Journal of the Econometric Society*, 323–351.

Arellano, M., and Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The review of economic studies*, *58*(2), 277–297.

Atkeson, A., and Burstein, A. (2008). Pricing-to-market, trade costs, and international relative prices. *American Economic Review*, *98*(5), 1998–2031.

Atkin, D., Chaudhry, A., Chaudhry, S., Khandelwal, A. K., and Verhoogen, E. (2015). Markup and cost dispersion across firms: Direct evidence from producer surveys in pakistan. *American Economic Review*, *105*(5), 537–44.

Bach, T. N. (2019). State owned enterprises and capital misallocation in vietnam. *Journal of the Asia Pacific Economy*, *24*(3), 430–451.

Baily, M. N., Hulten, C., and Campbell, D. (1992). Productivity dynamics in manufacturing plants. *Brookings papers on economic activity. Microeconomics*, *1992*, 187–267.

Baldwin, J. R., and Gorecki, P. (1998). *The dynamics of industrial competition: A north american perspective*. Cambridge University Press.

Banerjee, A. V., and Duflo, E. (2005). Growth theory through the lens of development economics. *Handbook of economic growth*, *1*, 473–552.

Bartelsman, E. J., and Dhrymes, P. J. (1998). Productivity dynamics: Us manufacturing plants, 1972–1986. *Journal of productivity analysis*, *9*(1), 5–34.

Bartelsman, E. J., and Doms, M. (2000). Understanding productivity: Lessons from longitudinal microdata. *Journal of Economic literature*, *38*(3), 569–594.

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

Bartelsman, E. J., Leeuwen, G. v., and Nieuwenhuijsen, H. (1998). Adoption of advanced manufacturing technology and firm performance in the netherlands. *Economics of innovation and new technology*, *6*(4), 291–312.

Bellone, F., Musso, P., Quéré, M., and Nesta, L. (2006). Productivity and market selection of french manufacturing firms in the nineties. *Revue de l'OFCE*, *97*(5), 319–349.

Bernard, A. B., Eaton, J., Jensen, J. B., and Kortum, S. (2003). Plants and productivity in international trade. *American economic review*, *93*(4), 1268–1290.

Bernard, A. B., and Jensen, J. B. (1999). Exceptional exporter performance: cause, effect, or both? *Journal of international economics*, *47*(1), 1–25.

Bernard, A. B., and Jensen, J. B. (2004). Why some firms export. *Review of economics and Statistics*, *86*(2), 561–569.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, *119*(1), 249–275.

Bruno, M. (1978). Duality, intermediate inputs and value-added. In *Contributions to economic analysis* (Vol. 2, pp. 3–16). Elsevier.

Caballero, R. J., and Hammour, M. L. (1994). The cleansing effect of recessions. *The American Economic Review*, 1350–1368.

Camacho, A., and Conover, E. (2010). Misallocation and productivity in colombia's manufacturing industries.

Chen, K., and Irarrazabal, A. (2015). The role of allocative efficiency in a decade of recovery. *Review of Economic Dynamics*, *18*(3), 523–550.

Collard-Wexler, A., and De Loecker, J. (2015). Reallocation and technology: Evidence from the us steel industry. *American Economic Review*, *105*(1), 131–71.

De Loecker, J. (2007). Do exports generate higher productivity? evidence from slovenia. *Journal of international economics*, *73*(1), 69–98.

De Loecker, J., and Eeckhout, J. (2018a). *Global market power* (Tech. Rep.). National Bureau of Economic Research.

De Loecker, J., and Eeckhout, J. (2018b). Some thoughts on the debate about (aggregate) markup measurement.

De Loecker, J., Eeckhout, J., and Unger, G. (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*, *135*(2), 561–644.

De Loecker, J., Goldberg, P. K., Khandelwal, A. K., and Pavcnik, N. (2016). Prices, markups, and trade reform. *Econometrica*, *84*(2), 445–510.

De Loecker, J., and Warzynski, F. (2012). Markups and firm-level export status. *American economic review*, *102*(6), 2437–71.

Dheera-Aumpon, S. (2014). Misallocation and manufacturing tfp in t hailand. *Asian-Pacific Economic Literature*, *28*(2), 63–76.

Edmond, C., Midrigan, V., and Xu, D. Y. (2015). Competition, markups, and the gains from international trade. *American Economic Review*, *105*(10), 3183–3221.

Edmond, C., Midrigan, V., and Xu, D. Y. (2018). *How costly are markups?* (Tech. Rep.). National Bureau of Economic Research.

Edmonds, E. V., Pavcnik, N., and Topalova, P. (2010). Trade adjustment and human capital investments: Evidence from indian tariff reform. *American Economic Journal: Applied Economics*, *2*(4), 42–75.

Ellison, G., and Glaeser, E. L. (1994). Geographic concentration in us manufacturing industries: a dartboard approach.

Ellison, G., and Glaeser, E. L. (1997). Geographic concentration in us manufacturing industries: a dartboard approach. *Journal of political economy*, *105*(5), 889–927.

Epifani, P., and Gancia, G. (2011). Trade, markup heterogeneity and misallocations. *Journal of International Economics*, *83*(1), 1–13.

Ericson, R., and Pakes, A. (1995). Markov-perfect industry dynamics: A framework for empirical work. *The Review of economic studies*, *62*(1), 53–82.

Foster, L., Haltiwanger, J. C., and Krizan, C. J. (2001). Aggregate productivity growth: Lessons from microeconomic evidence. In *New developments in productivity analysis* (pp. 303–372). University of Chicago Press.

Gandhi, A., Navarro, S., Rivers, D., et al. (2017). How heterogeneous is productivity? a comparison of gross output and value added. *Journal of Political Economy, forthcoming*.

Griliches, Z., and Regev, H. (1995). Firm productivity in israeli industry 1979–1988. *Journal of econometrics*, *65*(1), 175–203.

Guadalupe, M., and Wulf, J. (2010). The flattening firm and product market competition: The effect of trade liberalization on corporate hierarchies. *American Economic Journal: Applied Economics*, *2*(4), 105–27.

Ha, D. T. T., and Kiyota, K. (2014). Firm-level evidence on productivity differentials and turnover in vietnamese manufacturing. *The Japanese Economic Review*, *65*(2), 193–217.

Ha, D. T. T., Kiyota, K., and Yamanouchi, K. (2016). Misallocation and productivity: The case of vietnamese manufacturing. *Asian Development Review*, *33*(2), 94–118.

Hall, R. E. (1988). The relation between price and marginal cost in us industry. *Journal of political Economy*, *96*(5), 921–947.

Haltiwanger, J., Kulick, R., and Syverson, C. (2018). *Misallocation measures: The distortion that ate the residual* (Tech. Rep.). National Bureau of Economic Research.

Harrison, A. E. (1994). Productivity, imperfect competition and trade reform: Theory and evidence. *Journal of international Economics*, *36*(1-2), 53–73.

Harrison, A. E., Martin, L. A., and Nataraj, S. (2013). Learning versus stealing: How important are market-share reallocations to india's productivity growth? *The World Bank Economic Review*, 202–228.

Hashiguchi, Y., et al. (2015). *Allocation efficiency in china: an extension of the dynamic olley-pakes productivity decomposition* (Tech. Rep.). Institute of Developing Economies, Japan External Trade Organization (JETRO).

Hopenhayn, H., and Rogerson, R. (1993). Job turnover and policy evaluation: A general equilibrium analysis. *Journal of political Economy*, *101*(5), 915–938.

Hsieh, C.-T., and Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics*, *124*(4), 1403–1448.

Klenow, P. J., and Rodriguez-Clare, A. (1997). The neoclassical revival in growth economics: Has it gone too far? *NBER macroeconomics annual*, *12*, 73–103.

Kugler, M., and Verhoogen, E. (2011). Prices, plant size, and product quality. *The Review* *of Economic Studies*, *79*(1), 307–339.

Lagos, R. (2006). A model of tfp. *The Review of Economic Studies*, *73*(4), 983–1007.

Levinsohn, J. (1993). Testing the imports-as-market-discipline hypothesis.

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

Lichtenberg, F. R., and Siegel, D. (1990). The effects of leveraged buyouts on productivity and related aspects of firm behavior. *Journal of financial economics*, *27*(1), 165–194.

Lichtenberg, F. R., and Siegel, D. (1991). The impact of r&d investment on productivity–new evidence using linked r&d–lrd data. *Economic inquiry*, *29*(2), 203–229.

Lu, Y., and Yu, L. (2015). Trade liberalization and markup dispersion: evidence from china's wto accession. *American Economic Journal: Applied Economics*, *7*(4), 221–53.

Machicado, C. G., and Birbuet, J. C. (2012). Misallocation and manufacturing tfp in bolivia during the market liberalization period. *The BE Journal of Macroeconomics*, *12*(1).

Melitz, M. J., and Polanec, S. (2015). Dynamic olley-pakes productivity decomposition with entry and exit. *The Rand journal of economics*, *46*(2), 362–375.

Midrigan, V., and Xu, D. Y. (n.d.). Finance and misallocation: Evidence from plant-level data. *American economic review*, *104*(2), 422-458.

Mishra, D. (2011). *Vietnam development report 2012: Market economy for a middle-income vietnam* (Tech. Rep.). The World Bank.

Nishida, M., Petrin, A., Rotemberg, M., and White, T. (2017). Are we undercounting reallocation's contribution to growth? *US Census Bureau Center for Economic Studies Paper No. CES-WP-13-55R*.

Oberfield, E. (2013). Productivity and misallocation during a crisis: Evidence from the chilean crisis of 1982. *Review of Economic Dynamics*, *16*(1), 100–119.

Olley, G. S., and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica: Journal of the Econometric Society*, 1263–1297.

Pagan, A., and Ullah, A. (1999). *Nonparametric econometrics (themes in modern econometrics)*. Cambridge University Press.

Pakes, A., and Ericson, R. (1998). Empirical implications of alternative models of firm dynamics. *Journal of Economic Theory*, *79*(1), 1–45.

Pavcnik, N. (2002). Trade liberalization, exit, and productivity improvements: Evidence from chilean plants. *The Review of Economic Studies*, *69*(1), 245–276.

Peters, M. (2013). Heterogeneous mark-ups, growth and endogenous misallocation.

Petrin, A., and Levinsohn, J. (2012). Measuring aggregate productivity growth using plant-level data. *The RAND Journal of Economics*, *43*(4), 705–725.

Restuccia, D., and Rogerson, R. (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic dynamics*, *11*(4), 707–720.

Restuccia, D., and Rogerson, R. (2013). Misallocation and productivity. *Review of Eco-*

nomic Dynamics, *16*(1), 1–10.

Ryzhenkov, M. (2016). Resource misallocation and manufacturing productivity: The case of ukraine. *Journal of Comparative Economics*, *44*(1), 41–55.

Schumpeter, J. A. (1942). *Socialism, capitalism and democracy*. Harper and Brothers.

Scott, P., et al. (2017). Estimating market power: Evidence from the us brewing industry. In *2017 meeting papers.*

Topalova, P. (2010). Factor immobility and regional impacts of trade liberalization: Evidence on poverty from india. *American Economic Journal: Applied Economics*, *2*(4), 1–41.

Van Biesebroeck, J. (2005). Exporting raises productivity in sub-saharan african manufacturing firms. *Journal of International economics*, *67*(2), 373–391.