

Agglomeration, Productivity and Technology Spillover: Evidence from Vietnam

Bin Ni¹, Ha Thi Thanh Doan²

Abstract

Agglomeration brings costs (e.g., intensified local competition) as well as benefits (e.g., knowledge spillover). It is important to examine the net impact of agglomeration to understand the geographic distribution of economic activities. In this study, we use firm performance to capture the net impact of agglomeration. Using data from Vietnamese firm data in the 2000-2015 period, we first calculate the agglomeration index using various methods, and then use Vietnam's entry into WTO as a quasi-experiment to identify the impact of agglomeration on firms' total factor productivity (TFP) as well as other performance. Our difference-in-differences (DID) estimation shows that agglomeration has a negative impact on firms' overall performance, suggesting that the devastating competition effect dominates the beneficial spillover effect in Vietnam. Moreover, we find that the impact of agglomeration on firms' productivity varies across different industries and types of firms.

JEL classification: D22, F21, F64, Q56

Keywords: agglomeration, productivity, technology spillover, Vietnam

1. Corresponding author. Faculty of Economics, Hosei University, 4342 Aiharamachi, Machida, Tokyo 194-0298 Japan.

E-mail: jjadaniel@hotmail.com

2. Economic Research Institute for ASEAN and East Asia (ERIA). E-mail: hato.doan@gmail.com

1. Introduction

From the geographical point view, firms in the manufacturing industries tend to be highly concentrated. The clustering¹ phenomenon has long been investigated in the existing studies. The theoretical verification can be traced back to Marshall (1920) which postulated three reasons why firms locate in close proximity to each other. First, the transportation cost can be reduced. Second, when firms cluster, a pool of workers will be available which makes it easier to hire new workers, and also facilitates better matching of workers to jobs. Third, knowledge spillovers, in particular informal exchanges of ideas, are more likely to occur when firms are in close geographic proximity (Krugman, 1991; Fujita et al., 1999).

In reality, the analysis of agglomeration economies also has potentially important policy implications, especially for developing countries. Since the 1980s, agglomeration economies have been used to justify cluster policies by national and local governments in various countries, but the study has been mainly limited to advanced economies. With the fast-growing economic situation in the developing nations, such as Vietnam, the welfare analysis of agglomeration is also drawing attention. Among many that focus on the gains from agglomeration, we are particularly interested in two questions: First, given the positive correlation between the increase in the productivity of a firm and the level of agglomeration in the same industry (Figure 1), what's the causality behind it? Second, to what extent can such kind of productivity gain be explained by technology spillover? The answers to these questions should help understand how much economic gains can be expected from clusters and through what kind of channels.

To answer the questions above, we use Vietnamese firm-level data from GSO in the 2000-2015 period to conduct an empirical analysis. First, we apply the method as in Ellison and Glaeser (1997), and Howard et al. (2016) alternatively, to calculate the index of agglomeration by industry in each city. We then use the relaxation of FDI regulations on different industries upon Vietnam's entry into WTO in 2006 as a quasi-experiment to identify the impact of agglomeration on firms' performance, such as productivity, markup, sales, and employment etc. In practice, we apply a difference-in-differences (DID) estimation strategy, where the treatment group is defined as the foreign firms in the industries that were not entitled to special investment preferences before year 2006, whereas the control group refers to the firms in the industries that were given numerous kinds of investment preferences even before the WTO entry. Thus the relaxation of the regulations on foreign direct investment will give equal opportunity to foreign firms in

¹ In this paper, we will use clustering and agglomeration interchangeably.

these industries compared to those in the FDI-promoted industries. After controlling for the other characteristics between the treatment and control groups, we show that causal influence of industry agglomeration on firms' productivity is negative, suggesting that the competition effect dominates the beneficial spillover effect in the case of Vietnam.

Furthermore, after we obtain the above results, the next question that we can't help asking is: through which kind of channel does agglomeration affect the productivity of firms in proximity? In this study, we focus on the channel of technology spillover, which has been verified by the existing literature as one of the most efficient agglomerative forces. Despite the difficulty of accurately measuring technology spillover, we made an attempt to use different proxies to capture such channel. We calculate the horizontal spillover and vertical spillover index respectively, and interact them with the EG index. The same set of control variables are included and the results show that, vertical technology spillover is a more prevailing and efficient way for agglomeration to promote the productivity of firms located in that particular region. This is consistent with the findings in the previous studies.

Based on what has been achieved so far, this research will supplement the existing literature in at least three ways: **1.** Despite numerous studies that have studied this topic, very few have explored the cases of developing countries. Thus, this will be one of the pioneering studies that provide empirical evidence on the impact of agglomeration in Vietnam. **2.** To our best knowledge, few studies have investigated the channels through which agglomeration can promote firm-level productivity. Howard et al. (2016) is the closest to this study in terms of Vietnam, however, they focus on the determinants of industry coagglomeration. In this research, we attempt to fill in the blanks. **3.** We also contribute to the literature by carefully dealing with the endogeneity issues.

The remainder of the article is organized as follows. In Section 2, we introduce the current situation of industry agglomeration in Vietnam, which is in accordance with its high economic growth. Section 3 reviews relevant literature concerning this topic. In Sections 4 and 5, we present our data and methodology, and provide evidence on impact of the extent of industrial agglomeration in Vietnam. Section 6 discusses the further issues and robustness checks, whereas Section 7 concludes the paper.

2. Agglomeration of manufacturing industries in Vietnam

Vietnam is composed of sixty-three provinces and five centrally-governed cities, which stand on the same administrative level as provinces (namely Hanoi, Ho Chi Minh City, Can Tho, Da Nang and Hai Phong) in eight regions based on geographical and socio-

economic conditions. The eight regions are Red River Delta, Northeast, Northwest, North Central Coast, South Central Coast, Central Highlands, Southeast, and Mekong River Delta. The Red River Delta, the Southeast, and the Mekong River Delta have much smaller areas compared with the others, but they are the most densely populated areas, accounting for 59.4% of the country's population in 2011. By contrast, the Northwest and the Central Highlands are the least populated regions with less than 10.8% of the country's population in 2011.

The Red River Delta including Hanoi, the capital and the Southeast including Ho Chi Minh City, the largest city of Vietnam are also the two main economic hubs in Vietnam. These regions are the major industrial centers of the country, producing 24.0% and 50.1% respectively of the country's industrial output in 2010. The Northwest and the Central Highlands, on the other hand, are the least industrialized regions with industrial output less than 1.5% of the nation's total in 2010 (The Statistical Yearbook of Vietnam in 2011). Regarding agricultural production, the Mekong River Delta and the Red River Delta are the two major rice-producing areas in Vietnam, accounting for 50.7% of the country's agricultural output in 2010. The Southeast, the Mekong River Delta, and the Red River Delta are also the most important centers for services in Vietnam, and they have the three largest cities of Ho Chi Minh City, Can Tho, and Hanoi, respectively. Those regions accounted for 75.6% of the country's total service output in 2010 (see Table 1). As a result of being the biggest centers in agriculture, industry, and services, the living standards of people in the South East, the Red River Delta, and the Mekong River Delta are the highest in Vietnam.

3. Literature review

Many of the recent studies have recognized the importance of geography in spillover research. Apart from Marshall (1920), Krugman (1979, 1980), and Dixit & Stiglitz (1977) also lay the theoretical foundation on the pattern of agglomeration. When it comes to empirical studies, there is considerable amount of evidence that has shown the positive relationship between productivity and agglomeration phenomena, such as Ciccone and Hall (1996), Andersson and Loof (2011), etc. (see also the survey of Rosenthal and Strange, 2004).

As for the measurement of technology spillover, Audretsch and Feldman (1996) use industry R&D, university R&D and skilled labour as measures of knowledge spillovers. Greenstone et al. (2010) quantify agglomeration spillovers as the change in total factor productivity (TFP) experienced by incumbent manufacturing firms when a large

manufacturing firm locates in the same area. Ellison et al. (2010) use two measures. First, a technology matrix similar to the input–output matrix for the USA which captures how R&D activity in one industry flows out to benefit another industry. Second, they use a patent database to construct measures of patents in and out of pairs of industries. These measures, however, only capture official exchanges of technology. It is likely that many technology exchanges are more informal or accidental, particularly in developing country contexts.

In spite of the large presence of the Vietnamese economy in transition economies, studies on the effect of agglomeration in Vietnam are still rare. As an exception, Ercole (2013) investigates the relationship between industry agglomeration and economic growth in Vietnam. They find that Ho Chi Minh City, among the other few regions, has the highest extent of economic concentration and experiences the country’s most rapid economic growth. Furthermore, low - tech industries are more clustered than the mid - high - and high - tech industries are. On the other hand, Howard, Newman, and Tarp (2012) identify the determinants of agglomeration in Vietnam by applying the approach of Ellison, Glaeser, and Kerr (2010). The result is that they cannot find robust determinants of agglomeration, with the identified determinants varying with the choice of the measure of agglomeration. However, most of the studies on agglomeration in Vietnam focus only on the location patterns of economic activities with the regionally aggregated data, and there are few papers that examines the effects of agglomeration on firms’ performance. One recent study by Gokan et al. (2017) has applied Vietnamese enterprise data to conduct a firm-level analysis. By using the cluster detection method proposed by Mori and Smith (2013), they explore the causality between the industry agglomeration and firm-level productivity. They find that agglomeration actually improves firm-level productivity in Vietnam. However, such “spillover effect” only exists for foreign-owned firms, but not for state-owned or private firms. Similar to their strategy, in this paper, we will also investigate how agglomeration affects firms’ productivity as well as other performances, while using more decomposed industry categorization. On top of that, we make an attempt to take into account the endogeneity issue of clustering measurement, to better identify the causal effect of industry agglomeration.

4. Data and empirical methodology

Data

This study uses a panel dataset constructed from the Vietnam Enterprise Survey at firm level. The data was collected annually by the General Statistics Office (GSO) of

Vietnam covering 10 years from 2002 to 2015. The data was collected annually by the General Statistics Office (GSO) of Vietnam for all industrial sectors as at March 1 of each year. The general objectives of this survey are: (i) to collect the business information needed to compile national accounts; (ii) to gather up-to-date information on business registration; and (iii) to develop the statistical database of enterprises. This panel dataset covers 10 years from 2002 to 2011, during which Vietnam experienced two major economic changes, namely, WTO accession and the impact of global economic crisis. The majority of the firms in the dataset are found in the list of Vietnam Standard Industrial Classification (VSIC) codes², including all 22 manufacturing sectors out of 42 sectors in total. Profiles of firms with regard to ownership, labor, capital stock, turnover, assets, FDI, wage, materials inputs, and other information are provided. In the estimation, we measure capital and labor by fixed assets and total labor at the end of year t . Output and capital are deflated using annual GDP deflator³. The statistical summaries of all production inputs are shown in Table 2.

[Table 2 inserted here]

In addition, each firm is given a unique “enterprise code,” which is used together with a province code to identify firms and construct the panel dataset. Incomplete information about exports and imports, missing data for materials, and inconformity of units among different years lead to a reduction in observations that can be used in the analysis. We eliminate the missing observations in calculating firms’ productivity and we delete outliers⁴. Apart from that, we correct irregular jumps by dropping observations with enormous growth in some years; for missing location, we use information of the other years to impute. Moreover, the GSO surveyed all firms with foreign capital. An advantage of this dataset is that it reports the country of origin for the foreign capital⁵. The numbers of observations for each year are presented in Table 3.

[Table 3 inserted here]

² We use the first two digits indicated in VSICcode2007 and VSICcode1993 to identify industries. For simplicity, we aggregate some sectors.

³ The producer price index at the sectoral level is a preferred deflator but such data are not available for Vietnam.

⁴ Firms in the top and bottom 1 percentile of all firm-specific output and input variables (in the means of annual growth) were deleted from the sample. In addition, the top and bottom 1% of output/capital and output/labor are excluded.

⁵ If there are multiple foreign investors, we consider only the one with the largest investment share. If, for a foreign owned firm, Japan’s share of investment is the largest, we consider the firm to be a Japanese-invested firm. In our dataset, more than 90% of foreign owned companies have just one foreign investor. However, as firms do not report the location of their headquarters, we can only judge their country of origin by the ownership share information.

Empirical strategy

We first recover the total factor productivity for each firm using standard production function method, such as the one used in Wooldridge (2009). As alternatives, we apply the method as in Olley and Pakes (1996), and stochastic frontier as in Kumbhakar and Lovell (2000). The statistics are in Figure 2.

[Figure 2 inserted here]

Second, as for industry agglomeration index, we will follow Ellison and Glaeser (1997) and construct EG index. The EG index, later adapted by Ellison et al. (2010), has become the standard way to measure the extent of the clustering. In the model, N firms choose among M locations. A firm must choose whether to follow the prior firm's decision or to choose a location randomly by pointing out a dot at a map. The original EG index for a specific industry l is thus given by:

$$\gamma_l^{EG} = \frac{\sum_{i=1}^M (s_i - x_i)^2 - (1 - \sum_{i=1}^M x_i^2) \sum_{j=1}^N z_j^2}{(1 - \sum_{i=1}^M x_i^2)(1 - \sum_{j=1}^N z_j^2)}$$

where s_i is the share of industry l 's employment in area i . In our dataset, i refers to each city. x_i is the share of total employment in city i , and z_j is the size of the plants j of industry l . The Gini index G_l is defined as the first term in the numerator, whereas the Herfindahl index is defined as $H_l = \sum_{j=1}^N z_j^2$. We can obtain the more commonly-used expression:

$$\gamma_l^{EG} = \frac{G_l - (1 - \sum_{i=1}^M x_i^2)H_l}{(1 - \sum_{i=1}^M x_i^2)(1 - H_l)} = \frac{\frac{G_l}{1 - \sum_{i=1}^M x_i^2} - H_l}{1 - H_l}$$

The EG-index has a number of advantages over the other measurements, which are discussed in detail in Ellison and Glaeser (1997). However, as pointed out by previous literature (e.g. Howard et al. 2016), by using employment shares in the construction of the agglomeration index, the EG index implicitly focuses on individual employees as the source of agglomeration economies. In a developing country context, rather than employees, firms owners are more likely to play the role in the decision-making of agglomeration, particularly in relation to technology transfers, given that the majority of employment is unskilled. Moreover, a potential caveat to use the EG index is that it compares each city industry employment share to the mean industry employment share in that city. In doing so, it controls for population density in that city, rather than controlling for the overall employment distribution. This can lead to the EG index over-

weighting clusters in low population areas⁶. This is likely to be a problem particularly in Vietnam where the government actively encourages labour intensive firms to locate in rural areas where there are smaller populations. Also there are a large number of firms that have less than 10 employees in our dataset. Thus a possibility is that there might be many firms that cluster in one city, but the EG index is too small to capture the extent of agglomeration. Alternatively, we choose the new method as developed by Howard et al. (2016) to confirm. Instead of using firm employment, this new method applies the number of firms located in one area to construct the agglomeration index.

The baseline specification will be as follows:

$$\ln Y_{ikt}^j = \beta_0 + \beta_1 \cdot EG_{kt}^j + \mathbf{X}_{ikt} + \gamma_i + \gamma_j + \gamma_k + g_t + \varepsilon_{it}^j \quad (1)$$

The dependent variable is the logarithm of performance index for firm i in industry j , city k at time t . EG_{kt}^j in equation (1) is the EG index, calculated by industry-city pair. β_1 thus becomes our key interest, representing the causal effect of agglomeration on firm productivity. We also include control variables \mathbf{X}_{ikt} that might affect firm productivity, time, city, industry and firm fixed effects.

To further identify the causal effect of agglomeration on firm productivity, we follow Lu et al. (2017) and conduct a quasi-experiment analysis by applying difference-in-difference (DID) estimation.

The economic shock we choose is Vietnam's entry into WTO in 2006. As can be seen in Table 1, for example, there was an increasing clustering trend in the 6 major manufacturing regions in Vietnam after 2006. A part of this phenomenon can be explained by relaxing the limitation on foreign investors' "agglomeration" in certain regions (specifically, the provinces that have more industrial parks to attract foreign investment, such Dong Nai-28 parks, Binh Duong-27 parks, HCMC and Hanoi-16 parks each city, etc.). Thus the identification strategy relies on the comparison of the productivity of domestic firms that are located in cluster areas (the treatment group) with the productivity of firms that are located outside the clusters (the control group) before and after WTO entry⁷:

$$\ln Y_{it}^j = \beta_0 + \beta \cdot Treatment_i \times Post_{it} + \gamma \cdot EG_t^j \times Post_{it} + \mathbf{X}_{it} + \gamma_i + g_t + \varepsilon_{it}^j \quad (2)$$

⁶ Indeed, Billings and Johnson (2015) use generated data and find that the EG index may generate statistically significant localisation for an industry with high spatial variance due to high industry concentration in low population density (rural) areas.

⁷ An alternative way to do this is to compare productivity of domestic firms that changed their provinces (the treatment group) with the productivity of firms that stayed in the same areas (the control group) before and after WTO entry.

$Post_{it}$ in equation (2) is a dummy to show the time of Vietnam's WTO entry. We also include control variables X_{it} that might affect firm productivity, time and firm fixed effects. γ thus becomes our key interest, representing the causal effect of agglomeration on firm productivity. Note that the inclusion of $Treatment_i \times Post_{it}$ controls for the identification assumption, that is, any effects due to the change in the reaction to the change of relaxation of regulations on FDI, beyond the change in the degree of agglomeration. In other words, whether the estimated coefficient γ from equation (2) captures the true effect only hinges upon the satisfaction of the identification assumption. To deal with the potential heteroskedasticity, we cluster the standard errors at the industry level.

5. Estimation results

[Table 5 inserted here]

The results of baseline estimation are shown in Table 5, first we use the full sample to conduct the analysis. In accordance with Howard et al. (2016), we add the average firm size measured by value added for industry-city pair. The average TFP in industry j , city k is also included for further verification. Apart from that, we control for capital-labor ratio, and SOE/FDI dummies which represent the ownership of the firm. The sign of EG index is always negative, and significant when we include the full control variables. This indicates that firms regardless of their ownership, are negatively affected by the presence of a cluster of "competitors". In other words, the entry of new firms in the same industry can pose a threat to the existing firms that are already located in that particular area. This finding is in sharp contrast with the other literature, which basically verifies the positive impact of industrial agglomeration. We can attribute some phenomenon to the fact that most firms in Vietnam are small and medium ones. And unlike FDI firms, they do not have the resources and capability to compete with newcomers who have more advance technology. As a result, their productivity will be suppressed and even be crowded out the market.

[Table 6 inserted here]

Consistent with the results in the baseline estimation, the results using DID method present similar predictions. No matter whether we use the full sample or domestic firms only, EG index is always negative and significant. Meanwhile the interaction term

between the treatment (whether the firm belongs to the industry that experiences relaxation of regulations on foreign investment) and the post2007 dummy is not significant. This verifies the fact the competition effect outweighs promotion effect in the case of Vietnam, and provides the evidence that agglomeration does not necessarily lead to the prosperity of local firms, as suggested by previous literature.

6. Further issues and robustness checks

Further issues

It might be argued that firms with different ownership might react differently to agglomeration. To account for such potential heterogeneity, we divide firms according to their reported ownership, namely State-owned firms, private firms and FDI firms. We then repeat the practice as in equation (1).

[Table 7 inserted here]

As can be seen from Table 7, when we focus on the samples of FDI firms, EG index turns out to be negative and insignificant. In contrast, when the samples of private firms are applied, EG is negatively significant, indicating that firms with private ownership are relatively easier to be affected by the presence of a cluster of “competitors”. This can be plausible because state-owned firms and FDI firms are assumed to have more resources, and less vulnerable to external shocks. In comparison, private firms are not as productive as the other two types of firms, and can be suppressed when they are faced with fierce competition from their counterparts.

Robustness checks

In this section, we conduct several sensitive checks to verify the robustness of our findings. First, to improve the comparability between the treatment and control groups, we construct a matched control group, that is, unaffected firms (i.e., those without changes concerning their status towards investment restrictions) in the same prior treatment industry and with similar firm characteristics. In practice, we first estimate the probability of belonging to an FDI-restricted industry before WTO entry, based on firm markup, size, age, productivity, ownership structure, and industry and year dummies. The matched control firm is the firm with the closest predicted probability as that of the focus treatment firm.

Second, the change in industrial categorization is defined at the three-digit industry

level. To relieve the concern of misreporting industrial affiliation, we restrict the selection to permanent changers, that is, we exclude those firms that changed industrial affiliation many times in the sample period. As a further check, we repeat the analysis for changes defined at the two-digit industry level.

Furthermore, firms in the treatment and control groups may follow different time trends, which may compound our DID estimator. To address this concern, we allow firm-specific time trends in our DID estimation. The new regression specification becomes:

$$\ln Y_{it}^j = \beta \cdot Treatment_i \times Post_{it} + \gamma \cdot EG_t^j \times Post_{it} + X_{it} + \gamma_i + g_t + \gamma_i \times t + \varepsilon_{it}^j \quad (3)$$

in this way, it relaxed our original assumption that requires the treatment group to have followed the trend of the control group in productivity changes (or other performance), if they had not changed industrial affiliation⁸. The above practices do not change our qualitative predictions.

7. Conclusions

Using data from Vietnamese firm data in the 2000-2015 period, in this study we make an attempt to verify how industrial-geographic agglomeration affects firms' productivity as well as other performance, while taking into account the endogeneity issue that might occur due to firms' self-selective decision-making of agglomeration.

In practice, we apply a difference-in-differences (DID) estimation strategy, where the treatment group is defined as the foreign firms in the industries that were not entitled to special investment preferences before year 2006, whereas the control group refers to the firms in the industries that were given numerous kinds of investment preferences even before the WTO entry. After controlling for the other characteristics between the treatment and control groups, we show that causal influence of industry agglomeration on firms' productivity is negative, suggesting that the competition effect dominates the beneficial spillover effect in the case of Vietnam. Also when we focus on the samples of private firms, EG index is always negative and significant, indicating that firms with private ownership are relatively easier to be affected by the presence of a cluster of "competitors".

In further studies, we will make an effort to narrow down the channels of how agglomeration can affect firms' performance. More rigorous robustness checks will be conducted to confirm our predictive findings.

⁸ The results are not included in the manuscript but are available upon request.

References

- Andersson, M. and Lööf, H. (2011) "Agglomeration and Productivity: Evidence from Firm-level Data", *Annals of Regional Science* 46: 601–620.
- Audretsch B. D. and M. P. Feldman. (1996) "R&D Spillovers and the Geography of Innovation and Production", *American Economic Review* 86 (3), 630–640.
- Ciccone, A. and Hall, R. (1996) "Productivity and the Density of Economic Activity", *American Economic Review* 86, 54–70.
- Dixit, A. and J. Stiglitz. (1977) "Monopolistic Competition and Optimum Product Diversity", *American Economic Review* 67, 297–308.
- Ellison, G. and E.L. Glaeser. (1997) "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach", *Journal of Political Economy* 105, 889–927.
- Ellison, G., E. Glaeser, and W. Kerr (2010) "What Causes Industry Agglomeration?" *American Economic Review* 100(3), 889–927.
- Ercole, R. (2013) "Spatial Agglomerations of Two - Digit Manufacturing Industries within and across Vietnamese Locations", mimeo.
- Fujita, M., Krugman, P. R., Venables, A. J. (1999) *The Spatial Economy: Cities, Regions and International Trade*. Cambridge, MA: MIT Press.
- Gokan, T., I Kuroiwa, K. Nakajima, and S. Sakata (2017) "Agglomeration Economies in Vietnam : A Firm-level Analysis", IDE Discussion Paper 636.
- Greenstone, M., R. Hornbeck, and E. Moretti (2010) "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings", *Journal of Political Economy* 118 (3), 536–598.
- Howard, E., C. Newman, and F. Tarp (2012) "Measuring Industry Agglomeration and Identifying the Driving Forces", WIDER Working Paper, No. 2012/84.

Howard, E., C. Newman, and F. Tarp (2016) "Measuring Industry Coagglomeration and Identifying the Driving Forces", *Journal of Economic Geography* 16, 1055–1078.

Krugman, P.R. (1980) "Scale Economies, Product Differentiation, and the Pattern of Trade", *American Economic Review* 70, 950–959.

Marshall, A. (1920) *Principles of Economics*, Macmillan, London.

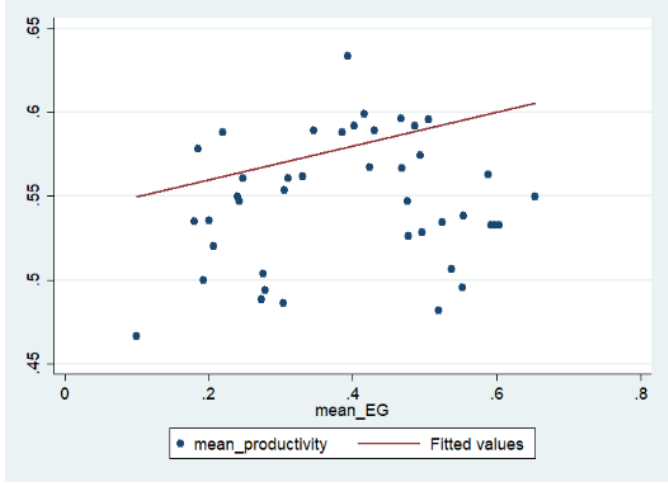
Mori, T. and T. Smith (2013) "A Probabilistic Modeling Approach to the Detection of Industrial Agglomerations", *Journal of Economic Geography* 14, pp.547–588.

Olley, S. and A. Pakes. (1996) "The Dynamics of Productivity in the Telecommunications Equipment Industry", *Econometrica* 64, 1263–1298.

Rosenthal, S. and W. Strange (2001) "The Determinants of Agglomeration", *Journal of Urban Economics* 50(2), 191–229.

Wooldridge, Jeffrey M., (2009) "On Estimating Firm-level Production Functions Using Proxy Variables to Control for Unobservables," *Economics Letters* 104(3), 112–114.

Figure 1: Industry agglomeration and firm-level productivity (EG measured at two-digit industry level)



Source: GSO, Vietnam

Table 1 General indicators of the regions in Vietnam

Region	Population share 2011 (%)	Agricultural share 2010 (%)	Industrial share 2010 (%)	Service share 2010 (%)	Income per capita 2010 (thousand VND)
Red River Delta	22.8	17.4	24.0	24.1	18,960
Northeast	8.1	7.0	5.3	3.0	10,994
Northwest	4.8	3.3	0.6	1.8	8,944
North Central Coast	11.6	7.8	2.2	6.2	11,118
South Central Coast	10.1	6.7	7.0	9.2	13,905
Central Highlands	6.0	14.1	0.8	4.2	13,056
Southeast	16.9	10.4	50.1	33.6	27,648
Mekong River Delta	19.7	33.3	10.0	17.9	14,964

Source: The Statistical Yearbook of Vietnam in 2011.

Note: The agricultural output value is at constant 1994 prices, the other indicators are at current prices.

Table 2 Summary statistics for the variables used for the estimation of production function (unit: million dong)

Variable	Mean	S.D.	N
Material	3.746	2.346	553993
Labor	2.385	1.287	1367707
Output	6.382	2.238	1318029
Capital	5.182	1.936	1197153
Investment	5.223	1.928	472853

Note: All variables are in logarithmic form

Table 3 Number of firms in six major regions

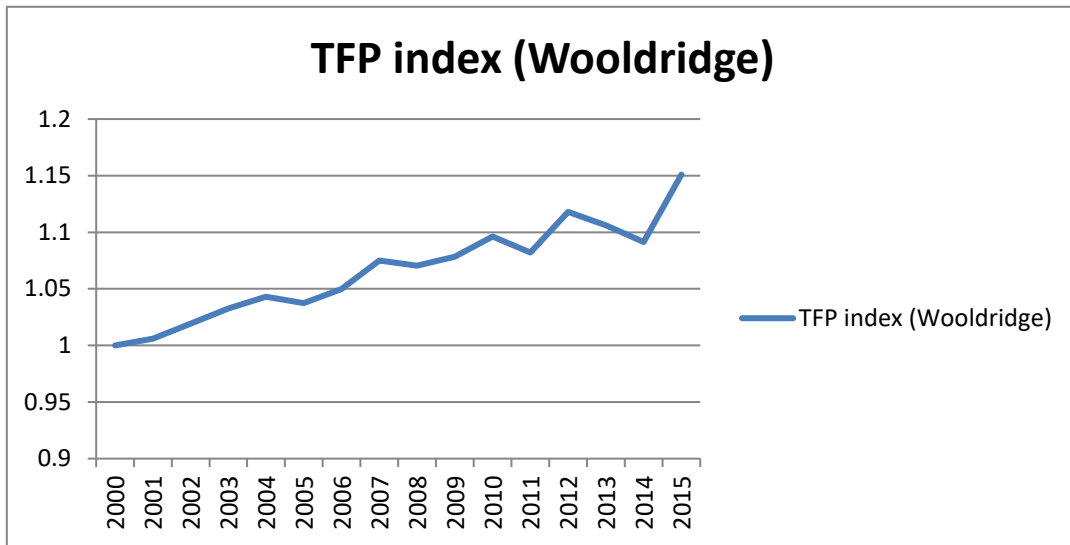
	2000	2007
Red River Delta	9,356	43,707
Northern Highlands	1,988	9,153
Central Coast	6,767	23,476
Central Highlands	1,827	4,597
South East	12,329	57,022
Mekong Delta	9,837	17,652

Source: GSO, Vietnam

Table 4 Summary statistics and variable definition

Variable	Var definition	Obs	Mean	SD	Min	Max
logva	Value-added (real)	489,055	10.06529	1.998344	-12.6262	21.89009
logk	Capital stock (fixed assets) (real)	499,149	10.8851	2.099051	0.633459	21.39645
logl	employment	534,978	2.974531	1.599058	0	11.33525
logtfp	TFP (Wooldridge method)	450,405	6.724117	0.9753179	-0.50665	12.14292
loglp	Labor productivity	479,594	7.062095	0.8682021	0.720575	12.19807
l_GEI	Ellison- glaeser index	534,951	0.0113915	0.0231996	-6.95633	0.914386
aveVA_jk	average firm size in district k, industry j	445,820	9.273007	1.800378	0	18.23991
nfdi_jk	number of FDI firms in dist k, ind j	534,978	2.825939	7.230956	0	66
nprivate_jk	number of private firms in dist k, ind j	534,978	32.56199	52.57785	0	357
n_jk	total number of firms in ind j dist k	534,978	35.70498	55.31652	1	386
aggN_outside	total number of firms in ind j province (excl dist k)	534,978	328.038	526.0435	0	2541
aveVA_outs~e	average firm size, outside dist k, same ind j same province	520,286	9.176865	1.433673	0	20.39221
tfp_jk	average tfp of firms in ind j, dist k (excl firm i)	410,912	5.799613	1.336892	0	12.04389
soe	SOE dummy	534,978	0.0327041	0.1778613	0	1
fdi	FDI dummy	534,978	0.1076343	0.3099183	0	1
private	private dummy	534,978	0.8596596	0.3473401	0	1
logkl	capital intensity	499,149	7.84255	1.437672	-1.56377	15.54081

Figure 2 TFP variation



Source: GSO, Vietnam

Note: Note: Estimated TFP at 3-digit industry level is normalized to year 2000

Table 5 Baseline estimation results

VARIABLES	(1) TFP	(2) TFP	(3) TFP
Ellison-Glaeser index	-0.262 (0.193)	-1.594* (0.910)	-1.539* (0.864)
number of firms, ind j dist k			
average firm size by VA, ind j dist k		0.043*** (0.004)	
number of FDI firms, ind j dist k		0.010*** (0.002)	0.010*** (0.002)
number of private firms, ind j dist k		0.001** (0.000)	0.001** (0.000)
average firm size, province		0.089*** (0.011)	0.088*** (0.011)
number of firms in same province, ind j, out of dist k		0.000 (0.000)	-0.000 (0.000)
logkl	-0.056*** (0.005)	-0.053*** (0.005)	-0.054*** (0.005)
soe	-0.029 (0.023)	-0.054** (0.024)	-0.052** (0.024)
fdi	0.159*** (0.044)	0.133*** (0.045)	0.134*** (0.045)
average tfp, ind j dist k, excl firm i			0.060*** (0.005)
Constant	6.439*** (0.067)	5.236*** (0.145)	5.329*** (0.133)
Observations	450,382	404,920	404,920
R-squared	0.118	0.143	0.145
Number of id	102,464	94,416	94,416

Robust standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Year, industry, city and firm fixed effects are included.

Table 6 Estimation results using DID

VARIABLES	Domestic firms		All sample	
	TFP	TFP	TFP	TFP
Ellison-Glaeser index		-2.043*		-1.595*
		(1.143)		(0.908)
number of firms, ind j dist k	-0.003		0.000	
	(0.006)		(0.006)	
average firm size by VA, ind j dist k	0.047***	0.047***	0.043***	0.043***
	(0.004)	(0.004)	(0.004)	(0.004)
number of FDI firms, ind j dist k	0.009	0.006**	0.009	0.010***
	(0.008)	(0.002)	(0.007)	(0.002)
number of private firms, ind j dist k	0.004	0.001**	0.000	0.001**
	(0.006)	(0.000)	(0.005)	(0.000)
average firm size, province	0.102***	0.102***	0.089***	0.089***
	(0.012)	(0.012)	(0.011)	(0.011)
number of firms in same province, ind j, out of dist k	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
logkl	-0.051***	-0.051***	-0.053***	-0.053***
	(0.006)	(0.006)	(0.005)	(0.005)
soe			-0.059***	-0.055**
			(0.022)	(0.024)
fdi			0.132***	0.133***
			(0.045)	(0.045)
post2007 dummy	0.506***	0.524***	0.502***	0.514***
	(0.089)	(0.074)	(0.078)	(0.067)
treatment*post2007 dummy	-0.002	-0.004	-0.002	-0.002
	(0.051)	(0.049)	(0.048)	(0.047)
Constant	4.988***	4.998***	5.227***	5.237***

	(0.166)	(0.162)	(0.148)	(0.146)
Observations	366,265	366,265	404,920	404,920
R-squared	0.139	0.139	0.143	0.143
Number of id	88,031	88,031	94,416	94,416

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Year, industry, city and firm fixed effects are included.

Table 7 Estimation results by firm type

Firm Type	FDI		Private	
	(1)	(2)	(3)	(4)
VARIABLES	TFP	TFP	TFP	TFP
Ellison-Glaeser index	-0.168 (0.542)	-0.174 (0.535)	-2.147* (1.144)	-2.051* (1.078)
average firm size by VA, ind j dist k	0.014*** (0.003)		0.050*** (0.004)	
number of FDI firms, ind j dist k	0.013*** (0.002)	0.013*** (0.002)	0.006** (0.002)	0.006*** (0.002)
number of private firms, ind j dist k	0.000 (0.000)	0.000 (0.000)	0.001** (0.000)	0.001* (0.000)
average firm size, province	0.007 (0.006)	0.008 (0.006)	0.108*** (0.012)	0.107*** (0.012)
number of firms in same province, ind j, out of dist k	-0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
logkl	-0.075*** (0.010)	-0.074*** (0.010)	-0.052*** (0.006)	-0.053*** (0.006)
average tfp, ind j dist k, excl firm i		0.019*** (0.004)		0.067*** (0.006)
Constant	7.313*** (0.130)	7.341*** (0.128)	4.909*** (0.166)	5.022*** (0.150)
Observations	38,655	38,655	355,682	355,682
R-squared	0.201	0.201	0.138	0.140
Number of id	6,854	6,854	86,868	86,868

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Year, industry, city and firm fixed effects are included.