Africa Growth Initiative at BROOKINGS

AFRICAN DEVELOPMENT BANK GROUP

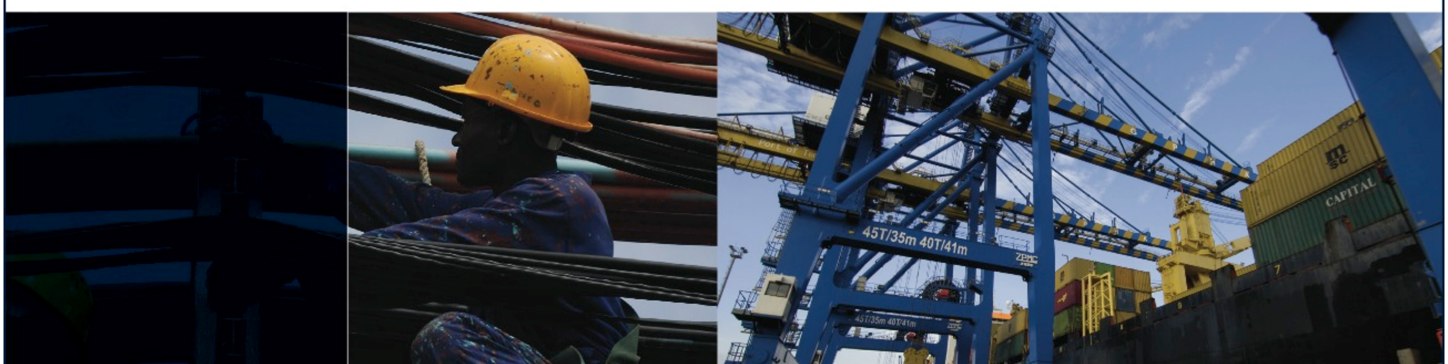
UNITED NATIONS UNIVERSITY

UNU-WIDER

World Institute for Development

LEARNING TO COMPETE

Working Paper No. 14



Productivity-enhancing manufacturing clusters: Evidence from Vietnam

Emma Howard,[1] Carol Newman,[2] John Rand,[3] and Finn Tarp[4]

Abstract

In this paper we explore the extent to which firms experience productivity spillovers from clustering using a rich data source from Vietnam for 2002 to 2007, a period of significant transition. We address issues of simultaneity, self-selection and endogenous location choice of firms in an estimation of firm level productivity. Controlling for competition effects and distinguishing between urbanization and localization economies, we find strong evidence for productivity spillovers from clustering. The effects of these spillovers are found to be particularly large for foreign-owned firms. Our results provide support for spatial clustering policies in developing countries aimed at attracting foreign investment.

Keywords: Clustering, productivity, endogenous location choice, spillovers, Vietnam
JEL classification: D2, L2, L6, O1

[1] Mansfield College, University of Oxford; [2] Trinity College Dublin; [3] University of Copenhagen; [4] UNU-WIDER; corresponding author: cnewman@tcd.ie

The Brookings Institution is a private non-profit organization. Its mission is to conduct high-quality, independent research and, based on that research, to provide innovative, practical recommendations for policymakers and the public. Brookings recognizes that the value it provides is in its absolute commitment to quality, independence and impact. Activities supported by its donors reflect this commitment and the analysis and recommendations are not determined or influenced by any donation.

Learning to Compete (L2C) is a collaborative research program of the Africa Growth Initiative at Brookings (AGI), the African Development Bank, (AfDB), and the United Nations University World Institute for Development Economics Research (UNU-WIDER) on industrial development in Africa. Outputs in this Working Paper Series have been supported by all three institutions.

AGI-Brookings is grateful for the contribution of an anonymous donor for funding its work under the collaborative research program.

1 Introduction

Enterprise concentration and the associated agglomeration economies have long been recognized as an important mechanism for facilitating industrial growth, both in the theoretical literature and by policy makers who have used industrial policy to promote geographic clustering of firms. The theoretical literature on the benefits to firms in clusters is well established. The seminal theoretical work by Marshall (1920) stated that it is of benefit to firms to cluster because it reduces the cost of transporting goods, people, and ideas. By reducing transport costs firms along the supply chain can buy inputs and sell their output more cheaply. A cluster of firms is also likely to attract a pool of suitably skilled labour reducing search costs and facilitate the matching of workers to jobs (Helsley and Strange 1991; Krugman 1991). Moreover, firms located in clusters will be better able to exchange ideas through knowledge sharing or transfers (Krugman 1991; Krugman and Venables 1996; Fujita et al. 1999).[1]

The above factors are agglomerative forces that motivate firms to cluster. They are, however, also the reasons why we expect firms in clusters to be more productive. First, being located close to suppliers or customers reduces transport costs but also increases competition. Under competitive pressures firms are incentivized to reduce slack, cut costs, and organize production more efficiently in order to compete. Second, the clustering of firms is expected to facilitate better matching of workers to jobs as a pool of labour emerges. Workers that are better suited to their jobs will be more productive (Böckerman and Ilmakunnas 2010; Overman and Puga 2010). Third, firms that are located in clusters are more likely to experience technology or knowledge spillovers which directly impact on firm productivity. In a developing country context where firms are operating far away from the 'best practice' frontier, spillovers are likely to have a large effect on firm performance as there is considerable scope for improvements in technology and practices (Siba et al. 2012). Through all three of these mechanisms we expect firms to want to locate in clusters but also that firms in clusters will be more productive. This makes the analysis of causal mechanisms challenging.

The empirical literature linking clustering to improved firm performance is not well-developed. Notable exceptions include Ciccone and Hall (1996) and Henderson (1986), who, using US and Brazilian data respectively, find that productivity increases in areas with a concentration of similar activities. Siba et al. (2012) find that agglomerating firms in Ethiopia have higher productivity, but only if they produce similar products to other firms in the cluster. In this paper we establish the extent to which clustering impacts on firm productivity in Vietnam. Specifically, we use panel data to determine what effect locating in a cluster has on the productivity of manufacturing firms and attempt to uncover the mechanisms through which the productivity of firms is impacted.

Isolating productivity spillovers to firms as a result of locating in a cluster is difficult for three reasons. First, we encounter the usual identification and simultaneity problems in finding a suitable estimate of productivity using firm level data. Second, the identification of the impact of clustering on productivity is confounded by the possibility of self-selection and the 'reflection problem'. In other words, it may be that the most productive firms choose to locate in productive clusters rather than clustering impacting on firm level productivity. Third, it is

[1] For empirical evidence see Ellison et al. (2010) who find evidence for all three Marshallian theories of agglomeration in the USA with input-output links (goods) between firms found to be the most important agglomerative force. In a developing country context, Howard et al. (2012) find that technology transfers (ideas) are an agglomerative force in Vietnam, but that only high-tech firms pool labour (people).

important to distinguish localization economies, i.e. whether firms experience spillover effects from firms located in close proximity, from urbanization economies, which are benefits of locating in an area with more economic activity generally that are not related to spillovers. Identifying the former is of interest in this context and so care is required in ensuring that these factors are separately identified.

To address the first two challenges we use a modified version of the Olley and Pakes (1996) (OP) approach to take account of the endogenous location choice of firms while at the same time controlling for self-selection of surviving firms and simultaneity in the estimation of productivity. Given recent critique of the structure imposed by the OP approach on the underlying behaviour of firms, and the expected sensitivity of results to deviations in these assumptions (Bond and Söderbom 2005), we also employ an index number approach to estimate total factor productivity (TFP). Using a non-parametric approach has the advantage of imposing no structure on the underlying technology and it does not require the estimation of a production function. We relate both measures of productivity to the characteristics of the clusters in which the firm is located controlling for firm, sector, and time-fixed effects. We separate out the impact of localization economies by controlling separately for the impact of both the density of the cluster (capturing urbanization economies) and the proportion of firms in the cluster that are in the same sector (capturing competition effects). Controlling for these factors we can detect productivity spillovers through a variable which captures the average productivity of firms in the same (4-digit) sector in the cluster. We also consider interactive effects and the extent to which other firm characteristics such as investment, technological characteristics and firm ownership impacts on the productivity relationship.

The results suggest that there are significant productivity spillovers associated with the clustering of firms in Vietnam even when urbanization economies and competition effects are controlled for. Of particular note is that foreign-owned firms benefit the most from spillovers providing support for industrial policy aimed at attracting foreign investment through the creation of clusters.

The reminder of the paper is organized as follows. Section 2 presents our empirical approach. Section 3 describes our data and presents summary statistics. Results are presented in Section 4 while Section 5 concludes.

2 Empirical approach

As highlighted in the introduction, isolating productivity spillovers to firms as a result of locating in a cluster is difficult due to (i) challenges associated with estimating productivity and (ii) the possibility of self-selection of productive firms into productive clusters. We use two different approaches to estimating productivity: a modification of the Olley and Pakes (1996) semi-parametric estimator, and a standard non-parametric measure of TFP. Both approaches have their respective advantages and disadvantages. The former approach allows us to estimate the productivity of firms controlling for simultaneity in the choice of inputs, selection bias due to firm survival and, with our modification, selection bias in the location choice of firms. This approach, however, has been criticized as imposing too much structure on the underlying behaviour of firms. In particular, Bond and Söderbom (2005) argue that where inputs are costly to adjust the OP estimator and other control function approaches do not work well. For this reason we also use a non-parametric measure of TFP that imposes no structure on the underlying technology and allows the data to speak for itself. In what follows we outline each approach and then turn to discussing our identification strategy.

2.1 Approach 1: Semi-parametric estimation of productivity

The estimation of firm productivity within a production function framework suffers from two sources of bias that are well documented in the literature: simultaneity bias and selection bias related to the survival of the most profitable firms. Simultaneity bias occurs in the estimation of the production function as the firm knows their own productivity when they choose the levels of capital and labour inputs. This productivity is, however, unobserved by the econometrician. A positive correlation between input choices and productivity will result in an upward bias on the ordinary least squares (OLS) estimate of the coefficient on capital in the production function. Selection bias occurs due to the relationship between productivity and the probability of exit from the market. Firms with different levels of capital stock are likely to react differently to productivity shocks. If we assume that more capital-intensive firms are more profitable, then a more capital-intensive firm is less likely to leave the market due to a productivity shock. This negative correlation will result in a downward bias on the coefficient of capital in the estimation of the production function. On net, it is uncertain as to which direction the OLS estimates will be biased.

The OP approach has become standard in the literature for addressing these biases and in this paper we adopt the idea of this approach. In our case, however, we have the additional problem that more productive firms are more likely to choose to locate in better performing clusters and so within productive clusters firms are a self-selected group. This will also lead to biased estimates of productivity at the firm level if location choice is not controlled for. Specifically, the location choice of the firm will affect its decisions regarding input choices which will bias estimates of the coefficients of a production function. We therefore include a modification to the OP approach that also allows us to take account of possible location choice self-selection in the estimation of firm level productivity.[2] This extension implies that the firm decisions to exit the market and to invest depend on which cluster they are in or their geographic location.

Provided that investment is monotonically increasing in productivity, OP argue that we can proxy productivity by a function in investment and capital. When the firm makes investment or exit decisions they face different market conditions and different potentials for spillovers depending on their geographical location. Therefore the productivity of the cluster in which the firm is located is a state variable which should also be included in the investment function. Extending OP, our investment equilibrium relation can be represented as

$$i_{it} = i_{it}(\omega_{it}, k_{it}, p_{mt}) \Leftrightarrow \omega_{it} = h_t(i_{it}, k_{it}, p_{mt}) \qquad (1)$$

where i_{it} represents the log of investment of firm i at time t, ω_{it} represents the firm's productivity, k_{it} the log of capital stock and p_{mt} the log of the productivity of cluster m at time t.

Each period the firm also decides whether or not to exit the market. They will exit the market if they experience a productivity shock greater than some threshold amount, given their current level of capital stock. Their decision will also depend on geographic location; if the firm is in a high productivity cluster their potential to experience productivity spillovers is higher and so their shock threshold will be higher.

Specifically, a firm will decide to exit the market (χ_{it}=0) if its productivity is less than some threshold amount, given the current level of capital stock and the productivity of the cluster, otherwise it will remain in the market (χ_{it}=1). Formally we write the decision rule as

[2] Our approach follows De Loecker (2013) who extends the OP framework by controlling for the export status of the firm in the estimation of the firm's production function.

$$\chi_{it} = \begin{cases} 1 \ if \ \ \omega_{it} \geq \underline{\omega}_{it}(k_{it}, p_{mt}) \\ 0 \ otherwise \end{cases} \quad (2)$$

Assuming a Cobb-Douglas production function, the first stage of the estimation procedure is

$$y_{it} = \beta_0 + \beta_l l_{it} + \Phi_t(i_{it}, k_{it}, p_{mt}) \quad (3)$$

where y_{it} is the log output of the firm, l_{it} is the log of the labour input, and

$$\Phi_t(i_{it}, k_{it}, p_{mt}) = \beta_k k_{it} + h_t(i_{it}, k_{it}, p_{mt}) \quad (4)$$

where *h*(.) is a polynomial in three variables; the productivity of the cluster enters into the estimation as a state variable. Equation (3) can be estimated by OLS and the coefficient for labour, a variable input, will be consistently estimated as *h*(.) controls for the unobserved productivity of the firm and the productivity of the cluster. We also obtain an estimate of $\Phi_t(.)$, which we term $\hat{\Phi}_t(.)$ from this stage of the analysis.

The second step of the estimation procedure calculates the probability of survival in period *t* by fitting a probit model of χ_{it} on i_{it-1}, k_{it-1} and p_{mt-1}, their squares and cross products. This gives us predicted probabilities of survival which we term $\hat{\chi}_{it}$.

The third and final step of the estimation is to fit Equation (5) by non-linear least squares, using the estimates of the coefficient of labour and $\Phi_t(.)$ obtained in the first stage ($\hat{\beta}_l$ and $\hat{\Phi}_t(.)$, respectively), and the predicted probabilities of survival ($\hat{\chi}_{it}$) estimated in the second stage.

$$y_{it} - \beta_l l_{it} = \beta_0 + \beta_k k_{it} + g\left(\hat{\Phi}_{t-1} - \beta_k k_{it-1}, \hat{\chi}_{it}\right) + v_{it} \quad (5)$$

The unknown function *g*(.) is approximated by a second order polynomial in $\left(\hat{\Phi}_{t-1} - \beta_k k_{it-1}\right)$ and $\hat{\chi}_{it}$.

The OP method relies on an exogenous process for productivity determination and assumes that productivity follows a first order Markov process (Olley and Pakes 1996). Productivity in period *t+1* is simply given by expected productivity, conditional on the firm's current productivity, and a shock or news term in the Markov process.

$$\omega_{it+1} = E(\omega_{it+1} \mid \omega_{it}) + \xi_{t+1} \quad (6)$$

The error term v_{it} in Equation (5) includes an *i.i.d.* shock and the news term in the Markov process. Note that here the productivity shock is based on results from the first step and thus also controls for the productivity of the cluster in which the firm is located.

The estimation of Equation (5) obtains a consistent estimate for the coefficient of capital. As we also have a consistent estimate of the coefficient of labour from the first step (the estimation of Equation (3) we can now recover consistent productivity estimates by calculating

$$\hat{\omega}_{it} = y_{it} - \hat{\beta}_l l_{it} - \hat{\beta}_k k_{it} \quad (7)$$

2.2 Approach 2: Non-parametric measurement of productivity

The second approach we use to measure productivity is an index number approach. This approach is commonly used and has recently been applied in the Vietnamese context (Newman et al. 2013). This is a relative measure of productivity that imposes no structure on the underlying technology or behaviour of firms. We measure the productivity of a firm, relative to the mean level of productivity of firms in the same 4-digit sector in each year. The measure captures the change in productivity over time by linking the sectoral differential to changes in the reference level of productivity from year to year and is given by

$$\begin{aligned}\omega_{ijt} = & \left(lnY_{ijt} - \overline{lnY_{jt}} \right) + \sum_{\tau=2}^{t} \left(\overline{lnY_{jt}} - \overline{lnY_{jt-1}} \right) \\ & - \sum_{m=1}^{k} \frac{1}{2} \left(s_{mijt} + \overline{s_{mjt}} \right) \left(ln\, X_{mijt} - \overline{ln\, X_{mjt}} \right) \\ & + \sum_{\tau=2}^{t} \sum_{m=1}^{k} \frac{1}{2} \left(\overline{s_{mjt}} + \overline{s_{mjt-1}} \right) \left(\overline{ln\, X_{mjt}} - \overline{ln\, X_{mjt-1}} \right)\end{aligned} \qquad (8)$$

where Y_{ijt} measures output of firm *i* in sector *j* in year *t*; X_{mjit} the amount of input *m* used by the firm; and s_{mjit} the expenditure of the firm on input *m* as a share of the total expenditure.

2.3 Identification of impact of clustering on productivity

In the second stage we relate the productivity of firms to the characteristics of the cluster they are in to see whether there are benefits to firms from agglomeration. Our main specification is given in the following model

$$\omega_{ijmt} = \mathbf{X}_{mt}\boldsymbol{\beta}_1 + \mathbf{X}_{ijmt}\boldsymbol{\beta}_2 + \mathbf{X}_{jt}\boldsymbol{\beta}_3 + \alpha_i + \varphi_j + \tau_t + \varepsilon_{ijmt} \qquad (9)$$

where: ω_{ijmt} is the productivity of firm *i* in sector *j* in cluster *m* at time *t*; $\mathbf{X}_{mt}$ is a vector of time-varying cluster-specific variables that capture the features of the cluster that firm *i* is located in, some of which are sector-specific; $\mathbf{X}_{ijmt}$ is a vector of time-varying firm-specific control variables for firm *i* at time *t*; and $\mathbf{X}_{jt}$ is a vector of sector-specific variables. Key to our identification strategy is the inclusion of a range of fixed effects that control for time invariant characteristics of firms, α_i, and sectors φ_j; time dummies that control for the general trend in productivity, τ_t are also included. Using this specification, the impact of clustering on firm productivity is captured by the within-firm variation in productivity over time. Controlling for all time invariant characteristics of the firms also controls for any initial firm characteristics that determine selection into a particular location. Moreover, the time invariant features of sectors (for example, the tendency for firms in certain sectors to naturally locate near each other) are also controlled for through the inclusion of sector fixed effects. Time-varying sector characteristics are included to pick up general trends in particular sectors such as an increase in foreign investment, competition, or a decline in state ownership in a sector.

The impact of the time-varying cluster-specific variables included in $\mathbf{X}_{mt}$ on productivity is the question of interest. First, we define the size of the cluster as the number of manufacturing firms located in the cluster. Following Henderson (2003) we count the number of firms in each cluster rather than focusing on employment. He argues that the sources of agglomeration externalities are individual firms rather than individual employees. Fujita and Ogawa (1982) also consider that the count of firms is related to externalities finding that the count of employees is unimportant. This variable will capture the natural tendency of firms to locate close to where there is more economic activity (i.e. urbanization economies).

Second, we compute the fraction of firms in the cluster that are in the same sector (measured at the 4-digit level) as the firm. The higher the proportion of firms in the cluster in the same sector the greater the competition, consequently firms need to be more productive in order to survive. This variable will therefore isolate the impact of clustering on firm productivity through the competition channel.

Third, we compute the average productivity of the cluster. We calculate this as a cluster and firm-specific variable in the following way; for firm *i* in area *m* we calculate the average productivity of all other firms in area *m*, excluding firm *i*. If the average productivity of other firms in the same cluster has an impact on firm productivity we interpret this as evidence of productivity spillovers. Knowledge and technology spillovers are more likely to occur when more productive firms are in close proximity. If productivity spillovers occur within clusters then we would expect firms to have higher productivity when they are in clusters with higher average productivity of firms in the same sector. In other words, this variable will capture localization economies associated with agglomeration.

It is also possible that the larger the cluster, the greater the competition effects and the greater the possibility of spillovers. To examine whether this is the case we also include interaction terms between the density of the cluster (number of firms) and: (i) the proportion of firms in the same sector in the cluster (isolating interactive effects between urbanization and competition effects) and (ii) the productivity of firms in the same sector in the cluster (isolating inter-active effects between urbanization and productivity spillovers).

In the final stage of our analysis we focus on the ownership structure of firms. The extent to which foreign-owned firms experience productivity spillovers from clustering is particularly important from a policy perspective given that industrial policy often attempts to encourage foreign-owned firms to locate in a particular area such as industrial parks or export-processing zones. If clustering is found to yield productivity spillovers then this will provide a sound basis for promoting such policies in the future. Moreover, if other firms, such as private domestic firms and state-owned firms are also found to benefit it suggests that such policies could also be extended to include other ownership types.

3 Data and summary statistics

The data are taken from the Vietnamese Enterprise Survey for 2002-2007 inclusive, provided by the General Statistics Office of Vietnam. The dataset includes all registered manufacturing enterprises at the end of each year with more than 30 employees, plus a random sample of 15 per cent of small registered enterprises with less than 30 employees. For the purpose of our analysis we drop firms with fewer than 30 employees and so we have the population of large manufacturing firms in Vietnam. Along with the standard financial information the data also include the name of the commune in which each firm is located. There are three levels of administrative areas in Vietnam: communes, districts, and provinces. In 2007 there were 4,325 communes, 631 districts, and 64 provinces. We restrict our sample to firms that do not change location over the sample period as an additional control for endogenous location choice.

To estimate our modified OP production function we measure output using value added. Capital is measured as the total assets of the firm and labour is the total number of employees. As noted in Section 2 we assume a Cobb-Douglas production function and take logs of each of these variables. For each firm we calculate investment as the change in total assets/capital from one period to the next assuming a depreciation rate of 2 per cent. Table 1 shows the summary statistics for the logs of output, capital, labour, and investment.

Table 1: Descriptive statistics

	Mean	Std. Dev.
Production function		
Log output	9.96	2.37
Log capital	7.23	1.90
Log labour	3.51	1.48
Log investment	5.51	2.42
Firm-specific controls		
Number of employees	145	627
Capital-labour ratio	3.82	14.38
Foreign-owned	0.110	0.313
State-owned	0.078	0.270
Sector-specific controls		
HHI	0.066	0.093
Weighted TFP	0.707	0.394
Foreign concentration	0.112	0.097
State concentration	0.066	0.086
Cluster variables		
Size of cluster (commune)	108	148
Fraction same sector (commune)	0.242	0.290
Average productivity (commune)	0.736	8.370
Fraction foreign (commune)		
Fraction high-tech (commune)	0.405	0.254
Size of cluster (district)	888	905
Fraction same sector (district)	0.132	0.206
Average productivity (district)	0.042	0.077

Source: authors' calculations using the Vietnamese Enterprise Survey 2002-07.

In order to isolate the benefits to firms of locating in a particular cluster in terms of productivity spillovers we need to control for the specific cluster the firm is in when estimating firm productivity. The most important characteristic of the cluster in terms of impacting on firm productivity is how productive the other firms in the cluster are. We compute the average productivity of the cluster (in logs) using the index number approach described above. For the index number approach output is revenue sales and inputs include capital, labour, and other intermediate costs of production. We use the TFP index to compute an estimate for p_{mt} for inclusion in the estimation of our modified OP model.[3]

In defining the cluster we consider two different administrative areas: the commune and the district. We construct cluster-specific variables at each of these administrative levels and consider them separately in the estimation of Equation (9). Summary statistics for all cluster-specific measures and other control variables are presented in Table 1.

4 Results

Production function estimates are presented in Table 2 when estimated using OLS, the standard OP approach and the modified OP approach with clustering defined at commune and district level. It is clear from the table that OLS estimates of capital coefficients are upward-biased relative to the standard OP and the modified OP approach. Also of interest is the fact that when the correction is made for the location choice of firms we find that both OLS and the standard OP approach appear to overstate the coefficients on the labour input and also on the capital input, although to a lesser extent. This suggests that there is a positive correlation between the productivity of firms and their decision to hire workers and invest in capital that is related to the productivity of the cluster where they locate.

Table 2: Production function estimates

	OLS	OP	Modified OP Commune	Modified OP District
Capital	0.430***	0.267***	0.247***	0.253***
	(0.001)	(0.005)	(0.008)	(0.010)
Labour	0.763***	0.795***	0.746***	0.745***
	(0.002)	(0.004)	(0.004)	(0.005)
Av. cluster productivity			0.010***	0.010***
			(0.003)	(0.003)
Number of firms		46,001	46,001	46,001
Number of obs.	128,419	136,001	136,001	136,001

Note. Bootstrapped standard errors are presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Source: authors' calculations using the Vietnamese Enterprise Survey 2002-07.

We use the parameters of the production function to back out a firm-specific productivity measure as described in Equation (7). As a robustness check we estimate productivity using an index number approach as described in Equation (8). Summary statistics for each measure are provided in Table 3. The correlation between the OP measures when clustering is defined at commune and district level is high. There is also a positive and statistically significant correlation between the indexed TFP measure and the OP measures.

[3] As we need the average productivity of the cluster for the OP estimation procedure we cannot calculate p_{mt} using the OP approach and so use the non-parametric TFP estimate instead. We also perform robustness checks using labour productivity of the cluster and find similar results.

Table 3: Productivity estimates

	Mean	Std. Dev.	Correlation Index TFP	Correlation OP Commune	Correlation OP District
Indexed TFP	0.553	0.416	1.000		
OP Prod (Cluster=Commune)	1.527	0.735	0.357	1.000	
OP Prod (Cluster=District)	1.487	0.733	0.356	0.999	1.000

Source: authors' calculations using the Vietnamese Enterprise Survey 2002-07.

To reiterate, the aim of our paper is to investigate the extent to which agglomeration and, in particular, the characteristics of clusters, have an impact on firm level productivity. Moreover, using the modified OP approach firm level productivity is estimated controlling for the self-selection of productive firms into productive clusters. We can therefore use this as the dependent variable in our analysis of clustering on productivity without being concerned about bias due to location selection. We focus on clustering at the commune level with the results for clustering defined at the more aggregate district level presented in the Appendix. As a further robustness check we resent the results using the TFP index in the Appendix.

Our baseline specification is presented in Table 4.[4] Each model includes firm, 4-digit industry and time fixed effects. In addition, we control for observed time-varying firm- and sector-specific effects as described in Sections 2 and 3. As such, the identification of the impact of agglomeration on firm level productivity comes from the within-firm variation in productivity over time that can be explained by changes in the characteristics of the clusters they are located in. Also key to our identification strategy is that we control for endogenous location choice in the estimation of firm level productivity and so this confounding factor is therefore no longer a concern.

[4] In all specifications where clusters are defined at the commune level standard errors are also clustered at the commune level. Likewise, when clusters are defined at the district level (see Appendix) standard errors are also clustered at the district level.

Table 4: Impact of clustering on firm productivity

Prod measure: OP Clustering: Commune	(1)	(2)	(3)	(4)	(5)
1: Size cluster	0.030***	0.033***	0.035***	0.039***	0.016*
	(0.006)	(0.006)	(0.006)	(0.007)	(0.010)
2: Fraction same sector		0.047**	-0.008	0.038	0.013
		(0.024)	(0.024)	(0.040)	(0.052)
3: Av. prod. same sector			0.351***	0.536***	0.520***
			(0.043)	(0.091)	(0.127)
Interaction 1&2				-0.014	-0.025
				(0.014)	(0.017)
Interaction 1&3				-0.232**	-0.159
				(0.115)	(0.162)
R-squared	0.015	0.014	0.016	0.017	0.059
Firms	32,018	32,018	32,018	32,018	6,177
Obs.	94,865	94,865	94,865	94,865	35,580

Note: Each model includes firm, sector and time fixed effects, and time-varying firm and sector controls. The full set of results for the main models presented in columns (3), (4) and (5) are provided in Table A2 of the Appendix. Robust standard errors clustered at the commune level are presented in parenthesis. *** p<0.01, ** p<0.05 * p <0.1.

Source: authors' calculations using the Vietnamese Enterprise Survey 2002-07.

We first consider the extent to which the size of the cluster, in terms of the number of manufacturing firms located in the commune, impacts on the productivity of the firm (column 1). Including this variable captures urbanization economies or, in other words, the natural tendency of firms to locate close to where there is more economic activity and other natural advantages. As expected, we find a positive and significant impact of the size of the cluster on the productivity of the firm. This result holds across all specifications and suggests that the productivity of firms is higher on average in locations where there are a large number of firms such as in urban centres or strategic regions. This result also holds for the other measures of productivity and when clustering is defined at the district level.

Second, we include the fraction of firms in the cluster that are in the same 4-digit sector (column 2). This variable captures competition between similar firms in the cluster and the extent to which this is productivity-enhancing or is detrimental to firms in terms of measured performance. We find this variable to be positive and statistically significant suggesting that the competition effects associated with locating close to competitors have a positive impact, on average. However, when we extend our model to include our measure of localization economies (column 3) the sign on the variable turns negative and it is no longer well-determined.[5] The average productivity of other similar firms located in the same commune is our core variable of interest. It is computed excluding the productivity of the individual firm in question. As discussed in Section 2 it is included to capture knowledge and technology spillovers, commonly termed localization economies; and given that we control for urbanization economies (through the inclusion of the size of the cluster) and competition effects (through the inclusion of the

[5] When clustering is defined at the district level we find some evidence of negative competition effects once the productivity of other firms in the cluster is controlled for. See Appendix for details.

fraction of firms in the same sector in the cluster) we can interpret it as such. As revealed in column 3 the variable has a positive and significant effect. The higher the productivity of other firms in the same sector in the commune the higher is the firm's productivity level. Given that we control for endogenous location choice in the estimation of productivity we can interpret this as evidence of technology spillovers.[6]

In column 4, we include interaction terms between the size of the cluster and the measures of cluster competition and localization to determine the extent to which there are differential effects in large compared with small clusters. We find that firms located in small clusters experience technology spillovers to a much greater extent than those in larger clusters. This is consistent with a model of technology diffusion whereby knowledge is easier to transmit within a cluster involving a small group of similar firms than a cluster with many firms. As a robustness check we estimate the model for a balanced panel of firms to ensure that our results are not driven by the selection of productive firms into productive clusters and the selection of less productive firms out of productive clusters. We find (in column 5) that our core result holds in the balanced panel.

Finally, we explore the extent to which the ownership structure of the firm matters for technology spillovers. The rationale for this disaggregation is that industrial policy often focuses on geographically concentrating firms, in particular foreign firms, in, for example, industrial parks or export-processing zones. As such the extent to which foreign firms benefit from productivity spillovers will matter for evaluating whether such policies are likely to work. In T able 5 we consider interaction terms between dummy variables capturing the form of ownership of firms and the average productivity of other firms in the same sector in the cluster.

[6] This result also holds when we use the lag of the productivity of other firms in the same sector in the same cluster. Results are available on request.

Table 5: Impact of clustering on firm productivity – role of ownership structure

Prod measure: OP Clustering: Commune	(1) Foreign	(2) Private domestic	(3) State
1: Size cluster	0.035***	0.035***	0.035***
	(0.006)	(0.006)	(0.006)
2: Fraction same sector	-0.008	-0.008	-0.007
	(0.024)	(0.024)	(0.024)
3: Av. prod. same sector	0.342***	0.443***	0.373***
	(0.044)	(0.127)	(0.044)
4. Foreign firm	0.012		
	(0.099)		
5. Private domestic firm		0.085	
		(0.082)	
6. State-owned firm			-0.066***
			(0.018)
Interaction 3&4	0.471**		
	(0.196)		
Interaction 3&5		-0.089	
		(0.131)	
Interaction 3&6			-0.457***
			(0.150)
R-squared	0.023	0.016	0.016
Firms	32,018	32,018	32,018
Obs.	94,865	94,865	94,865

Note: Each model includes firm, sector and time fixed effects, and time-varying firm and sector controls. Results for the control variables are very similar to those presented in Table A2 and are not presented but are available on request. Robust standard errors clustered at the commune level are presented in parenthesis. *** p<0.01, ** p<0.05 * p <0.1.

Source: authors' calculations using the Vietnamese Enterprise Survey 2002-07.

We find strong evidence to suggest that foreign-owned firms benefit most from being located in clusters where firms in the same sector have a higher average productivity. In other words foreign-owned firms experience the greatest productivity spillovers. This provides support for a policy of attempting to attract foreign-owned firms through the creation of clusters with the promise of agglomeration economies. Private-owned firms also experience positive productivity spillovers but not to the same extent suggesting that they too benefit from being located in productive clusters. This suggests that a policy promoting a combination of foreign and private domestic firms may work well.

In contrast, state-owned firms do not benefit from being in clusters with other productive firms suggesting that they are much less likely to experience productivity spillovers. This is consistent with other findings for Vietnam which suggest that state-owned firms have less absorptive capacity when it comes to receiving knowledge and technology transfers associated with export markets (Newman et al. 2014).

5 Conclusion

Empirical evidence on the benefits to firms from clustering is limited, particularly in developing country contexts. We make use of rich panel data on Vietnamese manufacturing firms to empirically analyse the impact of cluster productivity on firm productivity. Analysis of productivity spillovers from spatial clustering is fraught with difficulty, as firms may 'self-select' into productive clusters. Additionally, estimation of firm productivity requires careful consideration as firms select inputs based on their current productivity, which is unobserved to the econometrician. This leads to simultaneity problems in addition to self-selection problems.

In this paper we extend the Olley and Pakes (1996) method of estimating productivity to include average productivity of the cluster as a state variable. Our productivity estimates therefore control for the productivity of the cluster in which the firm is located in the firms' investment and exit decisions. Crucially, we calculate a firm-specific measure of the productivity of the cluster which excludes the firm itself. Our productivity estimates can therefore be used to isolate the impact on the productivity of firms as a result of clustering.

Overall, our results provide strong evidence for the existence of significant agglomeration economies in Vietnam. Unlike other recent work set in developing country contexts, the positive productivity spillovers associated with clustering do not seem to be affected by negative competition effects (see, for example, Chhair and Newman (2014) in the case of Cambodia and Siba et al. (2012) in the case of Ethiopia). Finally, we find evidence that foreign-owned firms benefit the most from agglomeration; private domestic firms also benefit but to lesser of an extent. Taken together this suggests that government policies aimed at attracting foreign investment through the creation of clusters, economic hubs or industrial parks, have worked in the case of an important transition economy, Vietnam, and have also benefited private domestic firms in the process.

Appendix Table A1: Impact of clustering on firm productivity

	(1)	(2)	(3)	(4)	(5)	(6)
Prod measure	OP	OP	Index TFP	Index TFP	Index TFP	Index TFP
Cluster definition	District	District	Commune	Commune	District	District
1: Size cluster	0.042***	0.043***	0.002	0.004*	0.010***	0.010***
	(0.008)	(0.009)	(0.002)	(0.002)	(0.003)	(0.003)
2: Fraction same sector	-0.186***	-0.253**	-0.030***	-0.017	-0.088***	-0.083**
	(0.047)	(0.150)	(0.007)	(0.013)	(0.016)	(0.034)
3: Av. prod. same sector	0.800***	1.098***	0.082***	0.127***	0.147***	0.139***
	(0.111)	(0.152)	(0.013)	(0.029)	(0.025)	(0.049)
Interaction 1&2		0.024		-0.004		-0.001
		(0.022)		(0.004)		(0.007)
Interaction 1&3		-0.438*		-0.060*		0.011
		(0.245)		(0.036)		(0.068)
Firm-specific controls						
Number of employees	-0.00002*	-0.00002*	-0.0003***	-0.00003***	-0.00003***	-0.00003***
	(0.00001)	(0.00001)	(0.00001)	(0.000001)	(0.00000)	(0.000001)
KL ratio	-0.005**	-0.005**	-0.0004**	-0.0004**	-0.0005**	-0.0005**
	(0.002)	(0.002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Foreign-owned	-0.016	-0.015	-0.035	-0.034	-0.034	-0.034
	(0.089)	(0.090)	(0.048)	(0.080)	(0.046)	(0.046)
State-owned	-0.073***	-0.073***	-0.032***	-0.032***	-0.032***	-0.032***
	(0.018)	(0.018)	(0.006)	(0.006)	(0.006)	(0.006)
Sector-specific controls						
HHI	-0.107**	-0.109**	-1.023***	-1.023***	-1.062***	-1.062***
	(0.053)	(0.053)	(0.067)	(0.067)	(0.100)	(0.099)
WTFP	-0.037***	-0.037***	0.889***	0.890***	0.886***	0.886***
	(0.011)	(0.011)	(0.005)	(0.005)	(0.007)	(0.007)
Foreign concentration	0.324***	0.324***	0.141***	0.142***	0.146***	0.146***
	(0.072)	(0.072)	(0.030)	(0.030)	(0.034)	(0.034)
State concentration	-0.060	-0.063	0.010	0.011	0.029	0.029
	(0.071)	(0.071)	(0.023)	(0.023)	(0.026)	(0.026)
R-squared	0.020	0.020	0.685	0.686	0.699	0.699
Firms	36,214	36,214	32,018	32,018	36,214	36,214
Obs.	101,060	101,060	94,865	94,865	101,060	101,060

Note: Each model includes firm, sector and time fixed effects. Robust standard errors clustered at the same level as the definition of the cluster are presented in parenthesis. *** p<0.01, ** p<0.05 * p <0.1.

Source: authors' calculations using the Vietnamese Enterprise Survey 2002-07.

Appendix Table A2: Impact of clustering on firm productivity – full results

Prod measure: OP Clustering: Commune	(1)	(2)	(3)
1: Size cluster	0.035***	0.039***	0.016*
	(0.006)	(0.007)	(0.010)
2: Fraction same sector	-0.008	0.038	0.013
	(0.024)	(0.040)	(0.052)
3: Av. prod. same sector	0.351***	0.536***	0.520***
	(0.043)	(0.091)	(0.127)
Interaction 1&2		-0.014	-0.025
		(0.014)	(0.017)
Interaction 1&3		-0.232**	-0.159
		(0.115)	(0.162)
Firm-specific controls			
Number of employees	-0.00002*	-0.00002	-0.00003**
	(0.00001)	(0.00001)	(0.00001)
KL ratio	-0.005**	-0.005**	-0.012***
	(0.002)	(0.002)	(0.002)
Foreign-owned	0.002	0.010	-0.098
	(0.099)	(0.099)	(0.102)
State-owned	-0.075***	-0.077***	-0.076***
	(0.018)	(0.035)	(0.020)
Sector-specific controls			
HHI	-0.121**	-0.067**	-0.152***
	(0.049)	(0.035)	(0.054)
WTFP	-0.045***	-0.042***	-0.054***
	(0.011)	(0.010)	(0.013)
Foreign concentration	0.321***	0.104**	0.395***
	(0.076)	(0.054)	(0.093)
State concentration	-0.108	-0.079	-0.152*
	(0.069)	(0.051)	(0.080)
R-squared	0.016	0.017	0.059
Firms	32,018	32,018	6,177
Obs.	94,865	94,865	35,580

Note: Each model includes firm, sector and time fixed effects, and time-varying firm and sector controls. Robust standard errors clustered at the commune level are presented in parenthesis. *** p<0.01, ** p<0.05 * p <0.1.

Source: authors' calculations using the Vietnamese Enterprise Survey 2002-07.

References

Böckerman, P., and Ilmakunnas, P. (2010) 'The Job Satisfaction-productivity Nexus: A Study using Matched Survey and Register Data.' MPRA Paper 23348. Munich: University Library.

Bond, S., and Söderbom, M. (2005). 'Adjustment Costs and the Identification of Cobb Douglas Production Functions.' WP05/04. London: The Institute for Fiscal Studies.

Chhair, S., and Newman, C. (2014). 'Clustering, Competition and Spillover Effects: Evidence from Cambodia.' Working Paper 2014/065. Helsinki: UNU-WIDER.

Ciccone, A., and Hall, R. (1996). 'Productivity and the Density of Economic Activity.' *American Economic Review* 86(1): 54-70.

De Loecker, J. (2013). 'Detecting Learning by Exporting.' *American Economic Journal: Microeconomics* 5(3): 1-21.

Ellison, G., Glaeser, E., and Kerr, W. (2010). 'What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns.' *American Economic Review* 100(3): 1195–1213.

Fujita, M., Krugman, P., and Venables, A. (1999). *The Spatial Economy: Cities, Regions and International Trade*. Cambridge, MA: MIT Press.

Fujita, M., and Ogawa, H. (1982). 'Multiple Equilibria and Structural Transition of Non-monocentric Urban Configurations.' *Regional Science and Urban Economics* 12(2): 161-96.

Helsley, R. and Strange, W. (1991). 'Agglomeration Economies and Urban Capital Markets.' *Journal of Urban Economics* 29(1): 96-112.

Henderson, J. (1986). 'Efficiency of Resource Usage and City Size.' *Journal of Urban Economics* 19(1): 47-70.

Henderson, J. (2003). 'Marshall's Scale Economies.' *Journal of Urban Economics* 53(1): 1-28.

Howard, E., Newman, C., and Tarp, F. (2012). 'Measuring Industry Agglomeration and Identifying the Driving Forces.' Working Paper 2012/84. Helsinki: UNU-WIDER.

Krugman, P. (1991). *Geography and Trade*. Cambridge, MA: MIT Press.

Krugman, P., and Venables, A. (1996). 'Integration, Specialization and Adjustment.' *European Economic Review* 40: 959-68.

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

Newman, C., Rand, J., and Tarp, F. (2013). 'Industry Switching in Developing Countries.' *World Bank Economic Review* 27(2): 357-88.

Newman, C., Rand, J., Tarp, F., and Nguyen Thi Tue Anh (2014). 'Exporting and productivity: the role of ownership and innovation in the case of Vietnam.' Working Paper 2014/070. Helsinki: UNU-WIDER.

Olley, G., and Pakes, A. (1996). 'The Ddynamics of Productivity in the Telecommunications Equipment Industry.' *Econometrica* 64(6): 1263-97.

Overman, H., and Puga, D. (2010). 'Labour Pooling as a Source of Agglomeration: An Empirical Investigation.' In, E.L. Glaeser (ed.), *Agglomeration Economics*. Chicago: University of Chicago Press.

Siba, E., Söderbom, M., Bigsten, A., and Gebreeyesus, M. (2012). 'Enterprise Agglomeration, Output Prices and Physical Productivity: Firm level Evidence from Ethiopia.' Unpublished: University of Gothenburg.