Macalester College
DigitalCommons@Macalester College

Economics Honors Projects Economics Department

Spring 4-25-2018

Does Agglomeration Account for Process Innovation in Vietnamese Small and Medium Enterprises?

Van Anh Le
Macalester College Economics Department, vananhle.157@gmail.com

Follow this and additional works at: http://digitalcommons.macalester.edu/economics_honors_projects

Part of the Regional Economics Commons

Recommended Citation
Le, Van Anh, "Does Agglomeration Account for Process Innovation in Vietnamese Small and Medium Enterprises?" (2018). *Economics Honors Projects*. 83.
http://digitalcommons.macalester.edu/economics_honors_projects/83

This Honors Project - Open Access is brought to you for free and open access by the Economics Department at DigitalCommons@Macalester College. It has been accepted for inclusion in Economics Honors Projects by an authorized administrator of DigitalCommons@Macalester College. For more information, please contact scholarpub@macalester.edu.

Does Agglomeration Account for Process Innovation in Vietnamese Small and Medium Enterprises?

Van Anh Le

Macalester College '18

Economics Honors Thesis

April 25, 2018

Advisor: Professor Sarah West
Readers: Professor Gary Krueger, Professor Samantha Çakır

Abstract: Although small and medium enterprises (SMEs) play a crucial role in the Vietnamese economy, this sector's growth is hindered by low level of technology and innovation. This paper uses firm-level panel data to examine whether process innovation activities in SMEs are influenced by their industrial environments. It measures the effects that agglomeration, the geographic concentration of firms within the same locality, has on firms' total outputs and their propensity to introduce new technology. Using a logistic model with firm fixed-effects, I find that agglomeration decreases outputs of informal firms and the likelihood of new technology introduction in all firms. However, there are evidence of positive lagged effects of agglomeration on innovation and heterogeneous effects across industries.

1. Introduction

In 2016, small and medium enterprises (SMEs), defined as enterprises with fewer than 300 employees, accounted for 93% of total registered enterprises in Vietnam. This sector is the main source of employment, accounting for 49% of all jobs and 99% of private-sector jobs. Despite their role in the economy, SMEs' growth is hindered by low levels of technology. According to the Ministry of Planning and Investment, Vietnamese SMEs rely largely on old machines, outdated technologies, and manual production (Le, 2016). Therefore, introducing new production technology, or process innovation, is a priority for SMEs' development. Several challenges to innovation for SMEs in developing countries similar to Vietnam have been identified. Ayyagari et al. (2011) find that lack of access to external finance and poor corporate governance negatively affect innovation in SMEs across 47 emerging economies. Owners' ability and personality traits are found to affect innovative activity in Bangladeshi micro-enterprises (De Mel et al, 2009).

On the other hand, there is an abundance of anecdotal and academic evidence on the relationship between agglomeration, the geographic concentration of firms in a specific locality, and innovative activity. This has led to the popularity of cluster-based industrial policies around the world, with governments trying to create industrial clusters. The theoretical logic behind this phenomenon is simple. Physical proximity increases the frequency of face-to-face interactions and facilitates the spread of knowledge between firms (Glaeser, 1992). Newly-acquired information allows firms to introduce new innovations to their production. Despite being widely studied in advanced economies, evidence of this phenomenon in developing countries is still lacking (Carlino and Kerr, 2014, Audretsch and Feldman, 2004, Overman and

Venables, 2005). Unlike in developed countries, innovative activity in developing countries often involves adapting existing technologies rather than inventing new processes. Furthermore, the presence of informal enterprises operating alongside formal ones in developing cities, and whether they enjoy the same benefits from agglomeration, have not been properly addressed in previous literature (Moreno-Monroy, 2012). Considering these differences, it is uncertain that cluster-based policy would be effective in spurring innovation in small firms in countries such as Vietnam. Thus, more empirical evidence on the relationship between innovation and agglomeration specific to this context is needed.

This paper asks whether the agglomeration of formal enterprises within the same industry accounts for process innovation in Vietnamese SMEs.[1] By measuring the impact of firm-clustering on process innovation decisions and total firm output, it aims to identify the presence of local intra-industry knowledge spillovers. It adds to the literature on SME innovation in developing countries. More importantly, it contributes additional evidence on the relationship between agglomeration and innovation in this context. The panel nature of my data allows me to address the reverse causality problem inherent to firm location choice. The results from fixed effect models with a panel of formal and informal SMEs suggest that the agglomeration of formal enterprises in the current year has negative effects on total outputs of informal firms. This implies that informal firms are disadvantaged by the clustering of formal firms. Agglomeration in the current year also decreases the likelihood of new technology introduction in both informal and formal enterprises. However, process innovation in formal firms increases

[1] Another form of agglomeration is urbanization, where firms from a diverse set of industries concentrated in a locality. Due to limited access to data, this form of agglomeration is not studied in this paper. Additionally, due to limited access to data, I am not able the measure the agglomeration of the informal sector.

with agglomeration from the previous year, suggesting a dynamic aspect of spillovers. Inspections of two subsamples of food products and fabricated metals manufacturers demonstrate how agglomeration forces operate differently across industries.

The next section reviews existing literature on agglomeration and innovation. Section 3 provides a theoretical framework. The empirical strategy is presented in Section 4. Section 5 describes the data and variables of interest. The results and caveats associated are discussed in Section 6 and Section 7 concludes the paper.

2. Review of the Literature on Agglomeration and Innovation

2.1 Innovation as a Mechanism of Agglomeration Economies

Agglomeration economies refer to productivity gains due to the geographical concentration firms in a specific locality. According to Marshall (1890), the concentration of many businesses of a particular industry creates large local markets for inputs, allowing firms to employ more specialized and productive inputs (Carlino and Kerr, 2014). Marshall calls this type of agglomeration industrial localization and productivity gains from it "localization economies".[2] Since Marshall, a large literature has developed around the mechanisms of agglomeration economies. Duranton and Puga (2004) classify these mechanisms under three broad classes: matching, sharing, and learning. Large local markets allow for the efficient sharing of local facilities, input suppliers, and skilled workers. They also lead to better matching between firms, workers, and suppliers. Lastly, proximity to other firms with knowledge on technologies, production methods or business practices facilitates learning through

[2] Jacobs (1960) later argues that it is the concentration of a wide variety of economic activities, rather than localization of one industry, that allow business to thrive through the cross-fertilization of ideas and knowledge. This type of agglomeration is often referred to urbanization economies and is beyond the scope of this paper.

face-to-face interactions (Carlino and Kerr, 2014). In this framework, process innovation is mostly associated with learning rather than matching and sharing.

2.2 Measuring Innovation

Most studies on innovation and agglomeration conclude that there is a positive correlation between industrial localization and innovation activities (Carlino and Kerr, 2014).[3] Focusing on advanced economies, these studies measure innovation with indicators such as patents and R&D expenditures, which limit the definition of innovation to the invention of new technologies, an activity separated from production (Lall, 1992). On the other hand, firms in developing countries often innovate through learning-by-doing, adapting or modifying existing technologies to fit their productions. For these firms, process innovation is more relevant. This term refers to the implementation of a new or significantly improved production method and encompasses changes to equipment, technology or production techniques (OECD, 2011).

Because process innovation often leads to higher productivity, one way to measure it is to estimate the production functions. If agglomeration increases firms' outputs given the same quantities of inputs, then it is likely that knowledge spillovers are present, enabling firms to engage in process innovation to increase productivity. Empirical studies generally agree that firms in regions with higher levels of agglomeration are more productive (Audretsch and Feldman, 2004). However, all three mechanisms of agglomeration may lead to this result. It is impossible to distinguish knowledge spillovers from the matching and sharing mechanisms (Duranton and Puga, 2004). Alternatively, a more direct measurement of innovation is a binary variable

[3] See Audretsch and Feldman (2004) and Carlino and Kerr (2014) for the most comprehensive summaries of this literature.

indicating whether a firm has introduced new or significantly improved production process or technology. If firms in more concentrated regions are more likely to innovate than firms in less concentrated regions, then this suggests knowledge spillovers from agglomeration. Using this measurement allows me to isolate process innovation from other mechanisms of agglomeration. Although this is preferred over productivity, the Vietnam SME survey only captures the introduction of new technology. Any activities that change the production process without the purchase of new equipment are not reported.[4] Therefore, the level of innovation measured with a binary variable is likely to be underestimated. To address the weaknesses of both measurements, I first measure agglomeration's effects on firms' total outputs. I then use a binary variable indicating whether firms have introduced a new technology. The rest of this section reviews empirical studies on these two outcomes.

2.3 Empirical Literature

The predominant concern in estimating the impact of local environment on individual firm outcomes is biases due to unobserved regional or firm-level characteristics. Certain regional characteristics such as institution and infrastructure quality correlate positively with both agglomeration and firm productivity. Furthermore, there is an endogeneity issue due to unobserved firm-characteristics. While agglomeration may increase productivity and the likelihood to innovate, firms with more innovative or productive owners are also more likely to locate in these locations due to the local advantage. Thus, unobserved regional and firm-level characteristics tend to cause positive bias on the agglomeration coefficient. Ways to correct for these unobserved heterogeneity vary in the literature as they often depend

[4] One example of such activity is changing stages of the production process such as the Japanese Kaizen methods.

on the nature of the data available to the researchers.

One of the first studies to focus on firms' decisions is conducted by Kelley and Helper (1999), who use a logistic model to estimate the likelihood of adopting computer numerically controlled machine technology in a cross-section of machined products manufacturers in the U.S. Localization economies are found to increase the likelihood of adoption. The relationship holds after controlling for confounding firm characteristics such as firm size and owner's experience. While greater availability of firm-level data has allowed economists to estimate the relationship on different types of firms and industries, the issue of endogeneity remained. Smit et al. (2015) use the 2012 European Community Innovation Survey (CIS) to measure the effects of localization on firms' innovation across 8 sectors in the Netherlands with a probit model. After controlling for firm characteristics, the positive relationship between agglomeration and innovation becomes insignificant, suggesting that spatial distribution of firms is not random. The cross-section nature of the data limits the paper's ability to control for other unobserved time-invariant firm characteristics such as owners' characteristics. Several studies on firms' decisions based on CIS data also face this issue.[5]

Henderson (2003) is the first to find evidence of productivity gains from agglomeration using firm-level panel data. Using a panel of U.S. high-tech and machinery manufacturing plants from the Longitudinal Research Database, Henderson estimates the effect on plants' outputs of agglomeration at the county level, controlling for capital, labor and intermediate inputs. Localization, measured as total number of own-industry establishments in the county, has a positive relationship on plant outputs

[5] Examples include De Beule and Van Beuren (2012), Johansson and Hans Loft (2006)

while total own-industry employment does not. A tenfold increase in the number of high-tech plants increases output by 20 percent. Henderson (2003) attempts to address the endogeneity problem using three different methods. First, individual plant fixed effects are included to control for time-invariant local and individual characteristics, leading to a higher coefficient on agglomeration than normal OLS. This contradicts with the expectations of positive bias mentioned above. Henderson attributes this to the fact that clusters of high-tech plants might be "accidents of history". To better control for time-variant unobserved characteristics, Henderson experiments with MSA-time fixed effects for multi-county MSAs, instrumental variables (IVs) for agglomeration variables and generalized method of moments (GMM). IV regressions (local business environment and air quality are used as instruments for agglomeration) suffer from the problem of weak instruments while MSA-time fixed-effects do not change the results significantly. Both MSA-time fixed-effects and GMM reduce the sample size significantly. Henderson concludes that using plant fixed effects are sufficient.

Since Henderson (2003), multiple studies have used firm-level panel data to estimate the productivity gains from agglomeration in different contexts. These studies employ different ways to estimate productivity and construct the agglomeration variables. Although papers' results vary depending on the country studied, they typically find that productivity is positively correlated with agglomeration and that the relationship is heterogeneous across sectors and types of firms.[6]

Two studies in particular focus on Vietnamese firms. Howard et al. (2014) use a semi-parametric estimation of productivity on a panel of Vietnamese large

[6] Lee et al. (2010) studies agglomeration's impact on output per workers in South Korea. Lall et al. (2004) uses an input-sharing model to estimate productivity in Indian manufacturing firms. Martin et al. (2011) explored non-linearity in agglomeration economies using a panel of French manufacturers.

manufacturing firms from 2002 to 2007. The paper finds that productivity increases with total number of manufacturing firms, the share of own-industry firms, and average productivity of other firms in the same clusters. Focusing particularly on the informal sector, Tran and La (2017) measure the effects of formal and informal agglomeration on a cross-section of informal enterprises. While informal localization generally has a negative impact on productivity, the effects of formal localization are heterogeneous across sectors and urban/rural settings. Agglomeration of formal enterprises at the district level decreases output per labor for wearing apparel firms but increases that of food products manufacturers. This is the only study so far to examine whether the informal sector benefits from a concentration of formal enterprises. Most notably, Tran and La (2017) find evidence in support of agglomeration diseconomies in certain industries. Another study focusing on the informal sector by Chhair and Newman (2014) finds a negative correlation between firms' outputs and the share of own-industry firms within the area. They suggest that agglomeration increases competitive pressures, erodes mark-ups and total revenues. Results from the few existing studies on informal firms in developing countries suggest that agglomeration diseconomies could be present. However, both Tran and La (2017) and Chhair and Newman (2014) use cross-sectional data, which limit their ability to address firms' location selection bias.

The first study using panel data to examine the effect of agglomeration on firms' innovation decisions is Zhang (2015), who estimates the effects of localization on firms' decisions to introduce a new product (product innovation). In addition to firm fixed-effects, Zhang also includes other time-varying characteristics such as skills-level of the labor force and government subsidies in the linear probability model. Results from this model show that industrial diversity has significant and positive impacts on the probability of innovation while localization does not. Unlike Henderson (2003), Zhang considers only localization in terms of total own-industry employment. It is unclear whether localization in terms of number of establishments would have any effects.

Overall, data availability is a key constraint in studying agglomeration. Compared to productivity studies, the number of studies focusing on firms' decisions is much smaller with few studies using panel data. This is because firm-level panel datasets (such as those used in Henderson (2003), Howard et. al (2016), Zhang (2015)) often come from government-led surveys aimed to measure productivity. These surveys provide fewer details on qualitative information such as whether firms have engaged in process innovation. The Vietnamese SME survey allows me to examine at both of these measures. Even with greater availability of firm-level datasets, good panel data are still hard to find. The ideal dataset would track a large number of firms over a sufficiently long period of time. This is because regional industrial characteristics tend to vary little year-by-year and using firm fixed-effects would require sufficient time variation in the agglomeration variables (Henderson, 2003). On the other hand, studies that use shorter panels than Henderson's such as Martin et al. (2011), Lee et al (2010) and Zhang (2015) typically have much larger sample sizes. Even with a long panel and large sample size, these studies generally find that GMM reduces the sample size significantly. Because the method requires first-differencing all variables and instrumenting first-differenced variables with their levels in previous periods, an observation is included only if for the same firm, two previous consecutive observations are available. Moreover, results from GMM are often not significantly different from using firm fixed-effects. Given these findings, I use fixed effects to alleviate the endogeneity issue. My main dataset covers firms within a short period of 8 years. Due to the survey's smaller scope and the high entry and exit rate for SMEs, my sample size is limited to 3718 distinct firms.[7]

Apart from differences in data and availability and methodologies, choices of agglomeration variables also vary. The first difference is whether agglomeration is measured by the total number of own-industry establishments or employment. Henderson (2003) uses both approaches and finds that total number of own-industry establishments in the county has a positive effect of firms' outputs while total own-industry employment does not. This is also the case for the two studies conducted on Vietnamese firms by Tran and La (2017) and Howard et al. (2014). The second difference lies in whether the effects come from absolute size of local industry, relative size compared to the locality's overall economy, or density. In section 6 and 7, I define these variables and experiment with all of these approaches.

Finally, the general consensus that agglomeration increases innovation might only be applicable to firms in developed countries and the formal sector in developing countries. Although agglomeration diseconomies are less discussed in this literature, evidence from cross-sectional studies on informal firms suggests that diseconomies could be present. Using a panel dataset that includes both formal and informal SMEs, this paper's findings confirm this hypothesis and contribute to the literature on agglomeration for informal enterprises in developing countries.

[7] Henderson (2003) uses a panel that spans 25 years at 5-year intervals. Zhang (2015) uses a much shorter panel but have 78000 firms in the sample.

3. Theoretical Framework

3.1 Firm Decision

Kelley and Helper (1999) provides a simple model on firm's innovation decisions. Suppose a firm is considering in period 1 whether to engage in process innovation in the next period, it will forecast costs and revenues in period 2's based on the environment it faces in period 1. Let the subscript *i* denote the scenario where the firm innovates and *n* denotes the scenario where the firm does not. A profit-maximizing firm will innovate if the innovation increases its total expected profit over both periods. That is, the profit from innovating, V^i, exceeds the profit from not innovating, V^n, as shown in the equation below.

$$\Pr(innovate = 1) = \Pr\left[\left(V^i - V^n\right) > 0\right]$$

Let Q be the quantity that the firm produces in each period. The firm's total variable cost is given by $C(Q)$. Assuming that the product is sold on a competitive market and that price is given as P, the difference in expected profits is given by:

$$V^i - V^n = \left[PQ_2 - C^i(Q_2, E_1) - I^i(E_1)\right] - [PQ_2 - C^n(Q_2)]$$

If the firm decides to innovate, its expected profit (V^i) is the difference between the revenue earned in period 2 (PQ_2) and total cost, which includes variable cost $C^i(Q_2, E_1)$ and a fixed cost of implementing the new technology $I^i(E_1)$. Since the firm relies on external information to predict cost of production under the new technology, C^i and I^i depend on the level of knowledge spillovers and agglomeration E_1. If the firm decides not to innovate, its expected profit (V^n) is simply the difference between revenue and the existing variable cost. Rearranging the above equation, the condition for process innovation depends on the expected savings in variable cost and fixed cost of implementation:

$$C^n(Q_2) - C^i(Q_2, E_1) - I^i(E_1) > 0$$

Therefore, the higher the expected cost savings from innovation ($C^n(Q_2) - C^i(Q_2, E_1)$) and the lower the fixed cost of investment ($I^i(E_1)$), the more likely to firm is to innovate.

3.2 Spatial concentration and knowledge spillovers

The central logic of knowledge spillovers is that proximity increases the probability of face-to-face interaction and the diffusion of information about new technology and production methods. Consider a city *j* where each firm occupies a unique location. The probability that an agent located at location l interacts with another at location *r* decreases with distance between the two points. Comin et al. (2012) express this probability as $e^{-\delta|l-r|}$, where the parameter δ dictates the ease of transportation. In well-connected cities, frequency of interaction decreases less drastically with distances. On the other hand, let G_l be a binary variable indicating whether the agent at l has knowledge relevant to an innovation. Assume that all interactions with others lead to some kind of learning, the amount of knowledge spillovers an individual firm enjoys can be expressed as the total number of agents with the relevant knowledge inversely weighted by their distances from the firm:

$$Knowledge\ Spillovers_{\,r} = \sum_{l \neq r} G_l\, e^{-\delta|l-r|}$$

However, due to constraints in data accessibility, calculating continuous distance $|l - r|$ between any two firms is not always feasible. Thus, it is common in the literature to take a regional approach, which is to assume that an agent is more likely to interact with someone within the same region than others outside. With this assumption, every firm in region j has the same amount of knowledge spillovers available to them, measured by the total number of agents in the city with the relevant knowledge.[8]

$$Knowledge\ Spillovers_{\,j} = \sum_{l \neq r} G_l$$

3.3 Hypotheses

From the theoretical analysis, an individual firm is more likely to engage in process innovation if the expected cost savings is sufficient to offset the expected fixed cost of innovation. Since expected costs are dependent on the information received, if interactions with other firms leads to knowledge spillovers, then we would expect agglomeration to affect the likelihood of innovating. However, according to Moreno-Monroy (2012), interactions between formal and informal entities do not always lead to successful transfer. If the interaction between the two sectors are exploitative, or if there are incentives to restrict information flow, then we would not expect to see any effects of formal agglomeration on the innovation activity of informal firms.

In addition, the decision also depends on the true profitability of the innovation

[8] The flaws of this approach are elaborated in Section 6.

in consideration. According to the theoretical literature on social learning and new technology adoption timing, the more information the firm receives, the closer its prediction of profitability is to the true value (Hoppe, 2002). Therefore, knowledge spillovers would decrease the likelihood of innovation if the information the firm receives lead it to believe that the innovation is not profitable. If the information received does not make the firm more certain about the expected payoff, it will choose to delay the innovation decision until a later period where more information is available.

4. Empirical Strategy

To determine whether agglomeration accounts for process innovation in Vietnamese SMEs, I first estimate production functions at the firm level, looking for effects on total output of agglomeration. Because productivity is not a direct measurement of innovation, I then estimate the effects of agglomeration of the probability of firm introducing a new technology. The estimation equations, following Henderson (2003) and Zhang (2015) respectively, are presented below.

Equation 1:

$$\ln Y_{it} = \alpha \ln X_{it} + \gamma Z_{it} + \beta_1 \ln E_{jkt} + \delta_t + f_i + \varepsilon_{it}$$

Equation 2:

$$Y_{it} = \begin{cases} 1\ if\ y_{it}^* > 0 \\ 0\ if\ y_{it}^* < 0 \end{cases}$$

$$\text{where:} \quad y_{it}^* = \alpha \ln X_{it} + \gamma Z_{it} + \beta_1 \ln E_{jkt} + \delta_t + f_i + \varepsilon_{it}$$

In (1), output of firm i at time t, Y_{it}, depends on the vector of firm inputs X_{it} (labor, capital and materials), a set of firm characteristics Z_{it} and agglomeration variables of province j, industry k, E_{jkt}. In equation (2), Y_{it} is an indicator variable on whether a firm has introduced a new technology. Both equations contain the same set of fixed-effects. A time fixed-effect δ_t control for nation-wide shocks in productivity. Most importantly f_i is the firm fixed-effect. As the sample is restricted to firms that did not change provinces over time, f_i controls for both unobserved firm and province characteristics that are correlated with agglomeration and productivity. To account for nation-wide productivity shocks, I include δ_t as the time fixed-effects. Finally, ε_{it} is the error term.

The set of firm characteristics Z_{it} includes whether the firm has received a bank loan, whether it has received government assistance, the share of workers in the firm who do not received a salary, and the percentage of skilled labor. According to Ayyagari (2011), government assistance and access to formal external credit are important determinants of innovation activities within emerging markets SMEs. In the context of Vietnam, where most household enterprises use unpaid family labor, enterprises that can afford to hire salaried workers are more productive (McCaig and Pavcnik, 2014). Thus, the higher the share of waged workers, the more productive the firm is.

To determine if household and formal enterprises are affected differently by agglomeration, I estimate equation 1 and 2 separately for two groups of firms. This is because the firm fixed-effects have already controlled for the legal status of each firm and the share of firms changing their status is small (less than 4%). Since previous studies have suggested heterogeneous agglomeration effects across industries, I also focus on two particular industries with the highest number of observations in the data, food and fabricated metal products manufacturing.

5. Data and Summary Statistics

The sample of firms in this paper comes from the Vietnam SME Survey, collected biennially since 2005 by several Vietnamese government agencies and foreign NGOs[9]. In Vietnam, private domestic firms can operate either as a formally registered private enterprise or as household business. Formal enterprises are required to follow formal accounting standards, make social insurance contributions on behalf of employees, and are subjected to corporate income tax. Firms with less than 10 employees have the option of operating as household firms and are not subjected to the above standards. The SME survey covers both types of enterprises. The sample of firms surveyed is taken from the General Statistics Office's Annual Enterprise Survey (for formal enterprises) and the Establishment Census (for household firms). It covers 9 out of 63 provinces in Vietnam, including the two largest urban centres (Ha Noi and Ho Chi Minh) and less densely populated provinces. Firms are surveyed on a range of topics including production, investment, and labour force. It also provides data on locations of firms at the province and district levels, the two largest administrative units in Vietnam. Firms' industries are reported at the 2-digit level following the Vietnam Standard Industrial Classification system (VSIC).

From this survey, I removed firms that changed their provinces to limit location selection bias. I also removed firms that changed their industries over time. This is because no information is available for the reason why firms would change industries. SMEs might decide to switch to another industry due to low profitability in the current sector. Alternatively, they might also switch to pursue opportunities for innovation in

[9] The agencies involved include: Central Institute for Economic Management, the Institute for Science and Social Affairs, the Development Economics Research Group at the University of Copenhagen, and the United Nations University.

other industries. As the relationship between changes in industries with the dependent variables cannot be accounted for properly, it is best to leave out these observations.[10] Observations that do not meet the Vietnamese government's definition of SMEs are excluded from the sample.[11] My final sample consists of 9682 observations, representing 3718 firms across 12 manufacturing industries from five waves of the survey (from 2007 to 2015 at 2-year intervals).[12] The resulting panel is unbalanced with only 748 firms appearing in all five waves. The largest portion of observations in the sample are manufacturers of food and fabricated metals. With regards to geographical distribution, the majority of firms are located in either Ha Noi or Ho Chi Minh city.[13] The Vietnam SME survey provides firm-level data used in the analysis. Table 1 lists firm-level variables, their definitions and summary statistics.

Table 2 provides average values of key variables by innovation decisions. Out of 9682 observations, firms indicated that had introduced new technologies in 1003 observations. In the majority of the observations (8679), firms had not introduced any new technology. Innovating firms have higher level of output and inputs than non-innovating firms. They also have a higher percentage of professional-degree holders in their workforce. A larger proportion of them received bank loan from a formal source and was assisted by the government within the last two years. The majority of non-innovating firms (70.35%) are also household enterprises.

[10] There are 728 out of 4444 firms that switch industries during the period from 2007 to 2015.
[11] To be classified as SMEs, firms must have revenue below 300 billion VND (about USD 13.3 million using 2015 exchange rate) and total workforce below 300 people.
[12] The Vietnam SME survey also includes a small percentage of enterprises in the agriculture and service sector. However, these are reported at the 1-digit industry level and cannot be used for this analysis.
[13] See Appendix for distribution of firms across industries and provinces

Table 1: Definition of Firm Variables

Variable	*Definition*	*Mean*	*St. Dev.*
Output	End-of-year revenue and value of finished output inventories (in millions VND)[14]	4162.8	14783.5
Labour	End-of-year number of workers the firm (includes part-time, full-time, casual and contracted workers)	14.48	27.63
Capital	End-of-year value of land, building, production and transportation machinery (in millions VND)	3035.16	9343.04
Intermediate Inputs	End-of-year cost of raw materials and other indirect cost (in millions VND)	2959.9	11743.63
Waged Workers	Percent of workers with a salary	59.6	38.7
Skills Level	Percent of workers with professional degrees	0.80	2.50
Variable	*Definition*	*% Yes*	*% No*
Process Innovation	Whether the firm has introduced a new production technology in the last 2 years	10.36	89.64
Household Enterprises	Enterprises that are not registered under the Enterprise Law, do not have a tax code or business registration codes	67.43	32.57
Formal credit	Whether the firm has received for a formal loan in the past 2 years	28.92	71.08
Government assistance	Whether the firm has received any government assistance in the past 2 years.	17.54	82.46
Observations:	9682 Number of firms: 3718		

[14] The exchange rate in 2015 is 1 million VND = 44.35 USD

Table 2: Firm Variable by Innovating Decisions

	Innovating		*Non-Innovating*	
Continuous Variable	*Obs.*	*Mean*	*Obs.*	*Mean*
Output (millions VND)	1003	9245.36	8679	3575.4
Labor (workers)	1003	29.97	8679	12.69
Capital (millions VND)	1003	5265.38	8679	2735.82
Intermediate inputs (millions VND)	1003	6454.47	8679	2556.04
Percent of waged workers	1003	81.01	8679	57.15
Percent of workers with professional degrees	1003	2.09	8679	0.66
	Innovating		*Non-Innovating*	
Binary Variables	*Obs.*	*Percent*	*Obs.*	*Percent*
Number of Household Enterprises	1003	42.17	8679	70.35
Access to Formal Credit	1003	48.45	8679	26.66
Government Assistance	1003	27.42	8679	16.40

To calculate the agglomeration variables, data on the number of enterprises and employment at the industry-province level is required. To do this, I used data from the Annual Enterprise Survey collected by the Vietnam General Statistics Office (GSO) from 2005 to 2010. This survey provides employment of enterprises, their locations and industries. It covers the population of all formally registered enterprises with 10 or more employees and a sample of smaller firms across 63 provinces. In addition, using published provincial statistical yearbooks from 2011 to 2016, I obtain statistics on employment in formal enterprises and number of enterprises for each industry-province found in the SME sample. I combined data from these two sources to obtain a panel of number of formal enterprises and formal employment for each industry-province from 2005 to 2015. From this panel, I experiment with three different approaches to measure industrial localization at the industry-province level: total number of own-industry establishments, own-industry establishment density, and share of own-industry establishment out of total number of establishments. For each of these approaches, I also tried replacing the number of own-industry establishments with total employment. The list of industry-province localization variables, their definitions and summary statistics are presented in Table 3. All variables exhibit strong variability. In particular, distributions of these variables are highly right-skewed. Thus, I follow the literature and log these variables.

Table 3: Definition of Agglomeration Variables

Variable	*Definition*	*Mean*	*Std. Dev.*
Size of local own-industry	• No. of own-industry formal enterprises	474.39	618.37
	• Employment in own-industry formal enterprises	26677.7	47651.89
Density of local own- industry	• No. of own-industry formal enterprises per km	0.191	0.281
	• Employment in own-industry formal enterprises per km	11.19	22.9
Share of local own-industry	• Ratio of own-industry formal enterprises over number of establishments (formal and household)	0.0015	0.0014
	• Ratio of own-industry formal employment over employment in all enterprises (formal and household)	0.02	0.02

I conduct a series of t-tests to determine the difference in industrial localization between innovating and non-innovating firms. There is no significant difference in the share of local own-industry between the two groups of firms. The difference is negative for all other variables, suggesting that on average, innovating firms are located in provinces with more formal enterprises and formal employment in their own industries.

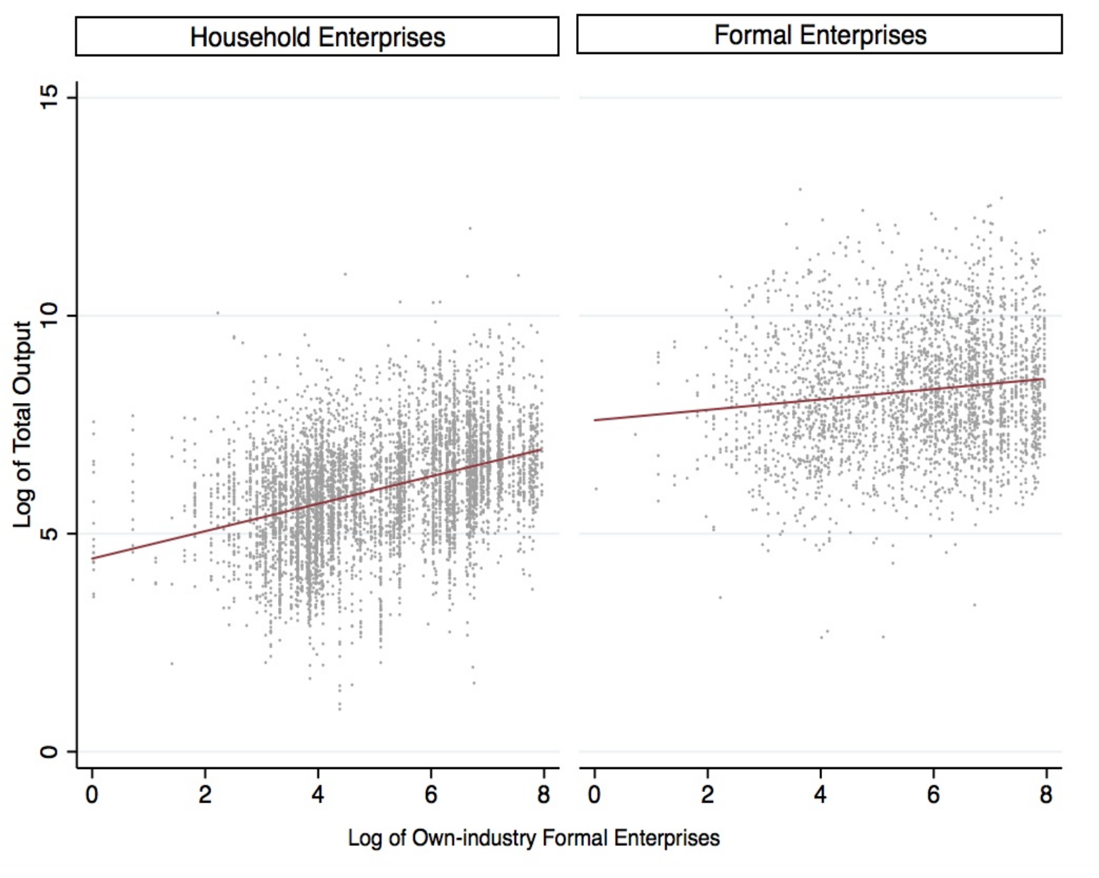
Table 4: T-test for Differences in Agglomeration between Innovating and Non-innovating Firms

Variable	*Non-innovating Mean*	*Innovating Mean*	*Difference*
No. of own-industry formal enterprises (in log)	5.18	5.25	-0.078*
Own-industry formal employment (in log)	9.15	9.28	-0.13***
Density of own-industry formal enterprises (in log)	-3.16	-2.95	-0.208***
Density of own-industry employment (in log)	0.81	1.07	-0.258***
Share of own-industry formal enterprises (in log)	-6.91	-6.93	0.014
Share of own-industry employment (in log)	-4.43	-4.45	0.001

Note: * p-value < 0.10 ** p-value < 0.05 *** p-value < 0.001

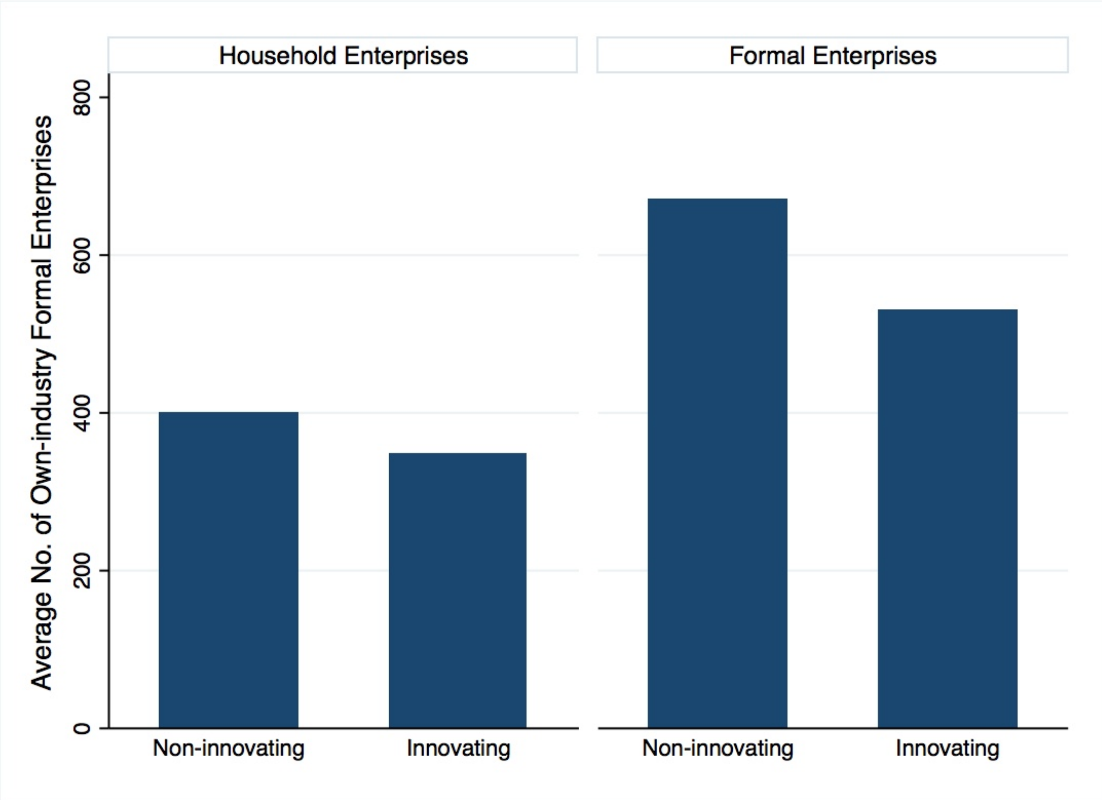
Figure 1 shows scatter plots of total firm output against number of own-industry formal enterprises within the province for household and formal enterprises separately. For both groups, there seems to be a positive relationship between output and agglomeration, as shown by the slopes of the fitted lines. However, this correlation does not take into account quantities of inputs and other firm-characteristics.

Figure 1: Scatter Plot of Total Output Against Agglomeration



Next, I divide formal and household enterprises into firms that have introduced a new technology (innovating firms) and firms that have not (non-innovating firms). The bar graphs in figure 2 shows that for both household and formal enterprises, firms that innovate are faced with a lower level of agglomeration than firms that do not. This could suggest a negative relationship between innovation and agglomeration. However, results from the student t-tests on the subsamples show that the differences are not statistically significant.

Figure 2: Average Agglomeration by Innovation Decisions



6. Results

6.1 Main Results

a) Output

Table 5 presents the basic results using OLS and fixed-effects models with different agglomeration variables on two groups of firms: household and formal enterprises. Columns 1 and 3 show results from OLS specifications with industry, year, and province fixed-effects. Columns 2 and 4 present the results from the preferred specification with firm fixed-effects. For both models, only the results using number of own-industry formal enterprises are presented. No significant effects are found for own-industry formal employment, density or share of own-industry employment across all models. The results for density and share of own-industry enterprises are not presented either as they resemble the results in Table 5 in magnitude, signs and significance.

For both models, coefficients on labor, capital and intermediate inputs are of the expected signs and highly significant. Compared to the OLS model, the coefficient on intermediate inputs is lower for the fixed-effects model while coefficients on capital and labor are higher. This suggests that values of intermediate inputs might be positively correlated with some time-invariant firm characteristics. For both models, the coefficients on inputs sum up to something close to one. Consistent with McCaig et al. (2014), the percent of waged workers is positive and significant. Receiving government assistance does not have any impact on total outputs.

Focusing on the industrial localization variable, one can see that OLS inflates the coefficient on number of own-industry firms for household firms (column 3, 4). This is consistent with the hypothesis that the distribution of firms is not random and more productive firms may self-select into areas with higher number of own-industry formal firms. While formal enterprises are not affect by agglomeration, results from the fixed-effects models suggests the presence of agglomeration diseconomies for household firms. The coefficients on number of own-industry firms for this sample are about -0.06, meaning that a doubling in the number of own-industry formal enterprises decrease outputs by 6% for the same set of inputs. These results suggest that agglomeration decreases the productivity of household enterprises. If we interpret productivity as an indicator of process innovation, then agglomeration decreases the level of innovation for household enterprises. Competition is another possible explanation as provided by Chhair and Newman (2014). Since this is a separate mechanism from knowledge spillovers, it will be discussed later in the Limitations and Future Directions section.

Table 5: Effect of Localization on Firm Output

	Formal Enterprise		Household Enterprise	
Dependent Variable: Log of Output	OLS	FE	OLS	FE
Log of own-industry formal firms	0.0132	-0.0127	0.00109	-0.0566***
	(0.0105)	(0.0367)	(0.00607)	(0.0177)
Log of total workforce	0.193***	0.186***	0.168***	0.153***
	(0.00811)	(0.0146)	(0.00663)	(0.0106)
Log of capital	0.0137***	0.0201***	0.0191***	0.0131***
	(0.00388)	(0.00645)	(0.00227)	(0.00376)
Log of cost intermediate inputs	0.776***	0.748***	0.793***	0.763***
	(0.00479)	(0.00793)	(0.00304)	(0.00459)
% workforce with professional degree	0.00338**	0.00131	0.0130	0.0109
	(0.00157)	(0.00239)	(0.00934)	(0.0127)
Received loan from bank	0.0174	0.0290*	0.00694	-0.00219
	(0.0111)	(0.0155)	(0.00718)	(0.00986)
% Waged workers	0.00199***	0.000966	0.00246***	0.00188***
	(0.000456)	(0.000772)	(0.000139)	(0.000211)
Received government assistance	0.0183	0.0188	0.00803	0.00974
	(0.0131)	(0.0155)	(0.00807)	(0.00961)
Year FE	YES	YES	YES	YES
Industry FE	YES	NO	YES	NO
Province FE	YES	NO	YES	NO
Province - Year FE	YES	NO	YES	NO
Firm FE	NO	YES	NO	YES
Constant	1.091***	1.543***	1.102***	1.548***
	(0.0779)	(0.205)	(0.0385)	(0.0824)
Observations	2,931	2,931	6,287	6,287
R-squared	0.966	0.898	0.970	0.915
Number of Firms	-	1,385		2,359

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

b) Introduction of New Technology

The impacts of agglomeration on the odds ratio of innovating over not innovating are estimated using logistic models with and without firm fixed-effects. With the fixed-effect logit models, firms whose innovation decisions do not vary over time are dropped and the number of firms in the sample is reduced. Table 6 presents the results for logit and fixed-effects logit model on the formal enterprise and the household enterprises subsamples.

Table 6: Effect of Agglomeration on Odds-Ratio of New Technology Introduction

	Formal Enterprises		Household Enterprises	
Introduction of New Technology	Logit	FE	Logit	FE
Log of own-industry formal firms	0.879	0.387**	0.997	0.438**
	(0.0975)	(0.178)	(0.113)	(0.157)
Log of total workforce	1.254***	1.270	1.235*	0.954
	(0.102)	(0.225)	(0.143)	(0.230)
Log of capital	1.157***	1.115	1.094*	1.122
	(0.0480)	(0.0899)	(0.0535)	(0.0928)
Log of cost intermediate inputs	1.036	1.049	1.202***	1.121
	(0.0514)	(0.104)	(0.0660)	(0.118)
% workforce with professional degree	1.018	1.063**	1.160	2.016***
	(0.0135)	(0.0306)	(0.130)	(0.526)
Received loan from bank	1.450***	1.694***	1.475***	1.705***
	(0.167)	(0.328)	(0.194)	(0.324)
% Waged workers	1.003	1.003	1.010***	1.005
	(0.00553)	(0.0106)	(0.00290)	(0.00489)
Received government assistance	1.427***	1.718***	1.217	1.231
	(0.178)	(0.303)	(0.179)	(0.216)
Industry FE	YES	NO	YES	NO
Year FE	YES	YES	YES	YES
Province FE	YES	NO	YES	NO
Province - Year FE	YES	NO	YES	NO
Firm FE	NO	YES	NO	YES
Constant	0.122**		0.0147***	
	(0.105)		(0.0111)	
Observations	2,838	1,080	6,260	1,191
Wald Chi-squared	312.02		335.13	
Likelihood Ratio Chi-squared		150.39		68.45
Number of Firms		310		316

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

The odds-ratios for total workforce and capital are positive and significant under OLS. Specifically, doubling the total workforce increases the odds of introducing a new technology about 250%. This is consistent with the findings from Kelley and Helper (1999). Doubling the amount of capital increases the odds by 115%. However, these coefficients become 0 under fixed-effects for both formal and household firms, suggesting that these inputs are correlated positive with other characteristics (such as entrepreneurial talent) that make firms more likely to innovate. Across all models, receiving a bank loan significantly increases the odds of new technology introduction. While government assistance is important for formal enterprises, it is not for households. Skills level of the workforce are significant and greater than one across both types of firms.

Similar to previous estimations on outputs, OLS overstates the coefficient on agglomeration. This is consistent with the expectation that firms that are more likely to innovate tend to locate in areas with a higher level of agglomeration. Although formal enterprises are not negatively affected in terms of total outputs, Results from fixed-effects model show that for both household and formal enterprises, clustering of formal enterprises in the same industry decreases the odds of new technology introduction. In particular, a 1% increase in the number of own-industry formal firms decreases the odds of new technology introduction by around 60% for both formal enterprises household enterprises. Based on the theoretical analysis in Section 3, the fact that agglomeration has a significant impact on innovation likelihood supports the presence of knowledge spillovers. However, odds-ratio smaller than one suggest that the information received by firms lead them to believe that the innovation in consideration is not profitable. Alternatively, it might not be informative enough for firms to overcome their uncertainties.

6.2 Dynamic Effects

In this section, I consider whether past agglomeration affects current productivity and innovation decisions as past agglomeration can contribute to a stock of knowledge in the region (Henderson, 2003). In particular, I consider the effects of number of own-industry formal enterprises lagged by 1 year and 5 years in separate regressions. As I only have access to the agglomeration data from 2005 to 2015, the sample used to estimate the effect of agglomeration lagged by 5 years only include observations from 2011, 2013 and 2015. Overall, there are no lagged effects on total output found for both the formal and household sub-samples. The results can be found in Table 4 in the Appendix.[15]

Table 7: Dynamic Effects on Odds-Ratio of New Technology Introduction

Dependent variable: Log of Output	Formal Enterprises		Household Enterprises	
Log of own-industry formal firm				
Lagged by 1 years	1.524**		1.094	
	(0.305)		(0.107)	
Lagged by 5 years		1.072		0.482
		(0.934)		(0.335)
Log of total workforce	1.267	0.944	0.990	1.276
	(0.225)	(0.333)	(0.237)	(0.487)
Log of capital	1.096	1.235	1.108	1.134
	(0.0876)	(0.209)	(0.0910)	(0.164)
Log of cost intermediate inputs	1.044	1.237	1.107	1.174
	(0.103)	(0.242)	(0.116)	(0.206)
% workforce with professional degree	1.065**	1.149***	2.011***	3.349**
	(0.0316)	(0.0604)	(0.521)	(1.983)
Received loan from bank	1.704***	2.015**	1.707***	1.739*
	(0.328)	(0.630)	(0.324)	(0.521)
% Waged workers	1.000	0.999	1.005	0.999
	(0.0111)	(0.0142)	(0.00492)	(0.00682)
Received government assistance	1.687***	2.074**	1.223	0.938
	(0.298)	(0.654)	(0.215)	(0.308)
Industry FE	YES	NO	YES	NO
Year FE	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES
Observations	1,080	440	1,191	466
Likelihood Ratio Chi-squared	151.62	83.82	64.02	42.15
Number of firms	310	164	316	169

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

[15] The results are not driven by differences in sample size. Limiting the sample to only observations from 2011 onwards does not alter the results in the previous sections.

Table 7 shows that a 1% increase in the number of own-industry formal enterprises in the previous year increases the odds ratio of new technology introduction for formal enterprises by 50%. The same effect is not observed for household firms, consistent with Moreno-Monroy (2012) arguments that the informal sector might not benefits from interactions with formal firms. The fact that for formal firms, current level of agglomeration negatively impacts new technology introduction yet lagged agglomeration increases the likelihood of innovation can be surprising. This finding can be explained using the theoretical literature on social learning and technology adoption timing. Models within this literature shows how firms accumulate information on the technology in consideration from previous to update their own beliefs of the technology's profitability. Thus, a firm might choose to delay adoption if the information receives signal low profitability or not sufficient to lower their uncertainty. Formal firms in this sample could be receiving knowledge about new technologies in both periods, but information from the current period might not be sufficient to overcome their uncertainties and alter their beliefs of the technology's profitability. Knowledge accumulated over a longer period of time (over a year), however, decreases uncertainties and inform their decisions to innovate.

6.3 Results from Individual Sectors

Lastly, I estimate the relationship between industrial localization and productivity/innovation for both a sub-sample of industries. Since food products (VSIC 10) and fabricated metals products (VSIC 25) have the highest numbers of observations within the data, I focus on these two industries. Similar to previous estimations, for each industry, I divide the sample into formal and household enterprises. Using the fixed effects model with log of own-industry formal establishments as the agglomeration variable, I consider the current and lagged effect of agglomeration on total outputs and innovation decisions. The differences in results from the overall sample support the fact that agglomeration externalities are heterogeneous across sectors.

Table 8 : Agglomeration Economies in Food Manufacturers

Dependent Variable: Log of Output	Formal Enterprises			Household Enterprises		
Log of own-industry formal firm						
No Lag	-0.143			-0.00475		
	(0.119)			(0.0293)		
Lagged by 1 years		0.0107			-0.00913*	
		(0.0241)			(0.00460)	
Lagged by 5 years			0.0359			0.0874***
			(0.120)			(0.0331)
Log of total workforce	0.172***	0.174***	0.172***	0.0856***	0.0862***	0.0717***
	(0.0328)	(0.0329)	(0.0319)	(0.0133)	(0.0133)	(0.0172)
Log of capital	0.0227	0.0215	0.00665	0.0140***	0.0134***	0.0313***
	(0.0153)	(0.0153)	(0.0182)	(0.00415)	(0.00414)	(0.00567)
Log of cost intermediate inputs	0.757***	0.757***	0.774***	0.823***	0.823***	0.827***
	(0.0183)	(0.0184)	(0.0198)	(0.00570)	(0.00569)	(0.00779)
% workforce with professional degree	0.00571	0.00686	0.00404	-0.0250	-0.0256	-0.0275
	(0.00543)	(0.00545)	(0.00513)	(0.0211)	(0.0210)	(0.0234)
Received loan from bank	0.0520	0.0471	-0.0299	-0.00600	-0.00604	-0.00522
	(0.0389)	(0.0389)	(0.0362)	(0.0115)	(0.0115)	(0.0153)
% Waged workers	-0.000316	-0.000435	0.00191	0.00187***	0.00186***	0.00219***
	(0.00150)	(0.00152)	(0.00160)	(0.000231)	(0.000231)	(0.000298)
Received government assistance	0.0252	0.0191	0.0316	0.00117	0.00256	-0.0173
	(0.0402)	(0.0399)	(0.0456)	(0.0104)	(0.0104)	(0.0169)
Year FE	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES
Constant	2.198***	1.401***	1.325*	1.049***	1.068***	0.609***
	(0.650)	(0.221)	(0.687)	(0.134)	(0.0394)	(0.157)
Observations	422	422	286	2,817	2,817	1,808
R-squared	0.929	0.928	0.946	0.945	0.945	0.933
Number of firms	189	189	142	975	975	822

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

For food products manufacturers (Table 7), I find no significant impact on total output from current agglomeration. However, household enterprises benefit from the agglomeration of food manufacturer in the formal sector five years ago. The coefficient on the 5-year lagged agglomeration variable is large and significant at the 5%. A doubling in number of formal food manufacturers in the province five years ago would allow household food manufacturers to increase their output by 9%. The same analysis is conducted for new technology within this industry and no significant results are found for current and past agglomeration (Table 5 in Appendix). For food products manufacturers, positive agglomeration effects on output could be due to their reliance on larger manufacturers for intermediate inputs. Alternatively, agglomeration economies can also result from process innovations that does not require purchasing of new technology. For manufacturers of fabricated metals products, no significant effects of agglomeration on outputs or new technology introduction are found. These results are presented in Table 6 and 7 in the Appendix.

6.2 Robustness Checks

For all regressions in this section, I experimented with multiple measures of agglomeration as mentioned in Section 5. Significant results are found only when agglomeration is measured using total number of firms. Agglomeration measured with own-industry employment does not have any effects. Furthermore, between different measurements using the number of firms (absolute number, density, and share of total firms in the local economy), there is no difference in the signs and statistical significances of the coefficients.

Within the provinces represented in the sample, a large number of firms come from Ha Noi and Ho Chi Minh city, which are the two largest urban centers in the country. Due to the difference in geographical settings between these two cities and the rest of the provinces, for each model presented here, I also ran separate regressions on the subsample of only firms from these two cities and the subsample of firms excluding those from these cities. Due to the small number of observations, using only firms from Ha Noi and Ho Chi Minh city does not lead to any significant results. The results on the subsample of firms outside of these cities are consistent with the results on the overall sample presented above.

6.3. Limitations and Future Directions

a) Competition as an alternative mechanism

So far, this paper focuses on the knowledge spillover mechanism of agglomeration. However, another mechanism that can explain the results found is competition. Similar to this paper, Chhair and Newman (2014) find that outputs of Cambodian informal firms decrease with agglomeration. They suggest that agglomeration increases the number of firms competing for local demand, decreasing the prices of goods and overall revenue for informal firms. Evidence from Ethiopian firms provided by Siba et al. (2012) shows that agglomeration decreases output prices but increases productivity. These two effects cancel each other out so that the effects on value of total outputs measured in the local currency is zero. Similarly, competition provides a plausible explanation for the results on total output in this paper. Informal firms' outputs suffer from lower prices as a result of increasing competition from formal firms. Formal firms, while affected by lower prices, enjoy agglomeration economies that boost their productivity. Thus their overall output is not negatively affected by agglomeration. Increasing local competition can also affect innovation decisions. Aghion et al. (2005) suggests that the relationship between innovation and competition resemble an inverted U. When the level of competition is too low, there is no incentive for monopolies to innovate. However, when the level of competition is too high, the expected returns on innovation is too low for firms to engage in the activity. Thus, it could be the case that local competition between small and medium enterprises in Vietnamese province is beyond the optimal point and increasing competition reduces the incentive to innovate.

Moreno-Monroy (2012) suggests that firms in developed economies often cluster for supply-side reason. In contrast, smaller firms in developing economies cluster due to demand factor. Agglomeration arises in large cities with a large local demand for goods and services. This makes studying the interaction between agglomeration economies and competition even more relevant in this context. Thus, an extension of this paper should consider how firms are affected by both mechanisms simultaneously.

b) Other limitations

While using firm fixed effects can partly address the endogeneity problems by controlling for firm specific, time-invariant characteristics that could affect their location choices, it does not solve the issue of selection bias. According to Martin et al (2011), a negative or positive shock in the province or in the industry can cause both likelihood innovation and agglomeration to change. While a negative shock can cause the SME to be less likely to invest in a new technology, it also causes other firms in the province to reduce employment, affecting the level of industrial agglomeration. Thus, agglomeration effects estimated are likely to be positively biased still.

Another limitation of this analysis was briefly mentioned in Section 3. There are no good theoretical reasons to believe that knowledge spillovers are geographically bounded. One can imagine firms located near a province border to interact more with agents from the

neighboring province than to another firm located on the other side of the same province. This is particularly important given the geographical natures of province boundaries in Vietnam. Province areas are irregularly shaped and very rarely do they have straight borders. There is also high variation in the sizes of provinces. A solution to this issue would be using data with exact locations of firms so that continuous distance between firms can be measured.

Since the available data only have location information at the province level, I am restricted to measuring local knowledge spillover at the largest scale possible, the province. However, I expect spillover at a smaller scale to dominates for firms in this sample. Smaller household enterprises might not have the resources to establish network beyond the immediate surrounding. Interactions, at least on a day-to-day basis, will be more frequent at the district levels. Inability to find any spillover at the province level does not necessary means that there are no knowledge spillovers. Furthermore, since there might be unequal distribution of formal and informal enterprises within the province, the finer the geographical unit, the more accurate the estimation of spillovers. Tran and La (2017) find agglomeration to have different effects on firms located in urban and rural districts within the same province, suggesting within-province heterogeneity in agglomeration economies.

Furthermore, this paper only considers one form of agglomeration. A theoretical model developed by Duranton and Puga (2001) suggests that urbanization economies are more beneficial to younger firms as they search for the optimal production methods within diverse industrial environments. Only firms at later stages in their product life cycles benefit from specialization. Given this, it is likely that SMEs will benefit from the clustering of diverse economic activities, rather than that of their own industries. In addition, I was only able to estimate the effects of formal sector agglomeration. Spillovers and agglomeration economies from clustering of household firms are not addressed.

7. Conclusions

Identifying the forces behind process innovation is important for SMEs development policy in Vietnam. Based on existing research on agglomeration and innovation in developed countries, cluster-based policies have become popular in many parts of the world. This paper sets out to determine whether Vietnamese SMEs, both in the formal and informal sectors, would benefit from policies that encourage clustering of firms in the same industries. Process innovation is measured using a binary variable indicating whether a firm has introduced a new technology. Since this measurement is likely to understate the level of innovation in SMEs, I also measure the impact of clustering on firms' outputs to determine whether agglomeration increases productivity. With a panel of firm-level data, the endogeneity issue is addressed using firm fixed-effects.

Results suggest that high concentration of own-industry formal enterprises decreases outputs of household enterprises while having no significant impact on formal enterprises. Similar to Tran and La (2017) and Newman and Chhair (2014), the results suggest that household firms are disadvantaged by the clustering of formal enterprises in the same industries. This could be due to increased competitive pressure, which reduces output prices. Formal enterprises are not affected as the effects of agglomeration economies could have canceled out the negative competitive pressure on these firms. Lagged agglomeration does not significantly impact firm's outputs, except in the case of household food products manufacturers. This suggests that agglomeration forces operate differently across industries and that dynamic effects might be present in certain sectors. More importantly, higher number of own-industry formal firms within the same province decreases the likelihood of new technology introduction. In particular, a 1% increase in the number of own-industry formal enterprises in the province decreases the odds of new technology introduction by 60% in both formal and household firms. It is possible that knowledge spillovers provide firm with more information on technologies with low expected returns, thus discouraging them to innovate. Another explanation is that knowledge spillovers within the current period is not sufficient to decrease uncertainties in profitability of adoption. The second explanation fits with the results for lagged effects of agglomeration. Specifically, formal enterprises' likelihood of new technology introduction is positive correlated with agglomeration from the year before. This suggests that there is a dynamic aspect of the knowledge spillover process. While information received within the current period causes firm to delay technology introduction, knowledge accumulated for over a year could be more beneficial to firms. Overall the results suggest that knowledge spillover might be at work for Vietnamese SMEs. However, spillovers from interaction between firms could potentially spread information on less profitable technologies, or increase uncertainty within firms. Thus, in order for cluster-based policies to be effective in spurring innovation, policy makers should aim to improve the quality of information on new innovations. A long term approach is also necessary as accumulated knowledge is important to firms' innovation decisions.

These results are subjected to several caveats. First, while SMEs experience diseconomies from industrial localization, they could be benefiting from urbanization economies. Due to lack of comprehensive data on the informal sector, I was only able to estimate the effect of formal agglomeration. Second, the large geographical unit of analysis used might not be accurately capture agglomeration economies and the effects may be found at a smaller spatial unit. To expand on findings and address these limitations, future research should consider the effects of urbanization economies and agglomeration of informal enterprises. If data on firms' exact locations are not available, studies should examine agglomeration effects on different geographical scales. With longer panels of data, dynamic effects of agglomeration and accumulated knowledge spillovers can also be studied more closely. Finally, more theoretical and empirical work are needed to explore how competition interacts with agglomeration economies and how both mechanisms affect innovation decisions and productivity in small firms in developing countries.

References

Audretsch, D. B., & Feldman, M. P. (2004). Knowledge spillovers and the geography of innovation. Handbook of regional and urban economics, 4, 2713-2739.

Ayyagari, M., Demirgü.-Kunt, A., & Maksimovic, V. (2011). Firm innovation in emerging markets: the role of finance, governance, and competition. Journal of Financial and Quantitative Analysis, 46(6), 1545-1580.

Chhair, S., & Newman, C. (2014). *Clustering, competition, and spillover effects: Evidence from Cambodia* (No. 2014/065). WIDER Working Paper.

Carlino, G., & Kerr, W. R. (2014). Agglomeration and innovation (No. w20367). National Bureau of Economic Research.

Comin, D. A., Dmitriev, M., & Rossi-Hansberg, E. (2012). *The spatial diffusion of technology* (No. w18534). National Bureau of Economic Research.

De Mel, S., McKenzie, D. J., & Woodruff, C. M. (2009). Innovative firms or innovative owners? Determinants of innovation in micro, small, and medium enterprises.

Duranton, G., & Puga, D. (2001). Nursery cities: Urban diversity, process innovation, and the life cycle of products. *American Economic Review*, *91*(5), 1454-1477.

Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies. In *Handbook of regional and urban economics* (Vol. 4, pp. 2063-2117). Elsevier.

Duranton, G. (2015). Growing through cities in developing countries. The World Bank Research Observer, 30(1), 39-73.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in cities. Journal of political economy, 100(6), 1126-1152.

Henderson, J. V. (2003). Marshall's scale economies. Journal of urban economics, 53(1), 1-28.

Hoppe, H. C. (2002). The timing of new technology adoption: theoretical models and empirical evidence. *The Manchester School*, *70*(1), 56-76.

Howard, E., Newman, C., Rand, J., & Tarp, F. (2014). *Productivity-enhancing manufacturing clusters: Evidence from Vietnam* (No. 2014/071). WIDER Working Paper.

Jacobs, J. (1962). The economy of cities. Vintage.

Kelley, M. R., & Helper, S. (1999). Firm size and capabilities, regional agglomeration, and the adoption of new technology. Economics of Innovation and New technology, 8(1-2), 79-103.

Kesidou, E., & Romijn, H. (2008). Do local knowledge spillovers matter for development? An empirical study of Uruguay's software cluster. World development, 36(10), 2004-2028.

Lall, S. (1992). Technological capabilities and industrialization. World development, 20(2), 165-186.

Le P. (2016). Financial assistance for small and medium enterprises to innovate. Science Journal of Hanoi National University 32(2) 75-82

Lee, B. S., Jang, S., & Hong, S. H. (2010). Marshall's scale economies and Jacobs' externality in Korea: The role of age, size and the legal form of organisation of establishments. *Urban Studies*, *47*(14), 3131-3156

Martin, P., Mayer, T., & Mayneris, F. (2011). Spatial concentration and plant-level productivity in France. Journal of Urban Economics, 69(2), 182-195.

McCaig, B., & Pavcnik, N. (2014). *Export markets and labor allocation in a low-income country* (No. w20455). National Bureau of Economic Research.

Moreno-Monroy, A. (2012). Critical commentary. Informality in space: Understanding agglomeration economies during economic development. Urban Studies, 49(10), 2019-2030.

OECD (2011) Defining innovation. (n.d.). Retrieved March 26, 2018, from https://www.oecd.org/site/innovationstrategy/defininginnovation.htm

Overman, H. G., & Venables, A. J. (2010). *Evolving city systems* (No. 2010, 26). Working paper//World Institute for Development Economics Research.

Smit, M. J., Abreu, M. A., & Groot, H. L. (2015). Micro-evidence on the determinants of innovation in the Netherlands: The relative importance of absorptive capacity and agglomeration externalities. Papers in Regional Science, 94(2), 249-272.

Tran, T. B., & La, H. A. (2017). Agglomeration Effects: Productivity of the Informal Sector in Vietnam. The Journal of Development Studies, 1-20.

Zhang, H. (2015). How does agglomeration promote the product innovation of Chinese firms?. China economic review, 35, 105-120.

Appendix

Table 1: List of 2-digit level manufacturing industries

VSIC code	*Description*
10	Manufacture of food products
11	Manufacture of beverages
12	Manufacture of tobacco products
13	Manufacture of textiles
14	Manufacture of wearing apparel
15	Manufacture of leather and related products
16	Manufacture of wood and products of wood and cork (except furniture)
17	Manufacture of paper and paper products
18	Printing and reproduction of recorded media
19	Manufacture of coke and refined petroleum products
20	Manufacture of chemicals and chemical products
21	Manufacture of pharmaceuticals, medicinal chemical and botanical product
22	Manufacture of rubber and plastics products
23	Manufacture of other non-metallic mineral products
24	Manufacture of basic metals
25	Manufacture of fabricated metal products (except machinery and equipment)
26	Manufacture of computer, electronic and optical products
27	Manufacture of electrical equipment
28	Manufacture of other machinery and equipment
29	Manufacture of motor vehicles, trailers and semi-trailers
30	Manufacture of other transport equipment
31	Manufacture of furniture
32	Other manufacturing
33	Repair and installation of machinery and equipment

Table 2: Distribution of Firms Across 2-digits industry

VSIC code	*2007*		*2009*		*2011*		*2013*		*2015*		*Total*	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
10	629	32.3	659	34.1	657	35.9	682	36	755	36.3	3,382	34.9
11	38	2	48	2.5	47	2.6	49	2.6	50	2.4	232	2.4
12	2	0.1	1	0.1	0	0	0	0	0	0	3	0
13	91	4.7	85	4.4	71	3.9	65	3.4	71	3.4	383	4
14	76	3.9	92	4.8	93	5.1	100	5.3	109	5.2	470	4.9
15	32	1.6	36	1.9	36	2	38	2	44	2.1	186	1.9
16	205	10.5	190	9.8	151	8.3	168	8.9	184	8.8	898	9.3
17	48	2.5	52	2.7	45	2.5	44	2.3	40	1.9	229	2.4
18	50	2.6	58	3	44	2.4	56	3	72	3.5	280	2.9
19	0	0	2	0.1	2	0.1	2	0.1	3	0.1	9	0.1
20	24	1.2	23	1.2	20	1.1	27	1.4	26	1.2	120	1.2
21	9	0.5	7	0.4	10	0.5	13	0.7	14	0.7	53	0.5
22	108	5.6	102	5.3	84	4.6	99	5.2	125	6	518	5.4
23	132	6.8	116	6	97	5.3	82	4.3	78	3.7	505	5.2
24	12	0.6	8	0.4	10	0.5	10	0.5	14	0.7	54	0.6
25	329	16.9	339	17.6	337	18.4	329	17.4	359	17.2	1,693	17.5
26	9	0.5	2	0.1	5	0.3	7	0.4	7	0.3	30	0.3
27	23	1.2	21	1.1	27	1.5	26	1.4	26	1.2	123	1.3
28	12	0.6	10	0.5	7	0.4	6	0.3	10	0.5	45	0.5
29	13	0.7	9	0.5	8	0.4	7	0.4	7	0.3	44	0.5
30	5	0.3	1	0.1	1	0.1	2	0.1	2	0.1	11	0.1
31	82	4.2	64	3.3	67	3.7	73	3.9	76	3.7	362	3.7
32	11	0.6	5	0.3	7	0.4	9	0.5	8	0.4	40	0.4
33	5	0.3	1	0.1	2	0.1	2	0.1	2	0.1	12	0.1
Total	1,945	100	1,931	100	1,828	100	1,896	100	2,082	100	9,682	100

Table 3: Distribution of Firms Across Provinces

Province	*2007*		*2009*		*2011*		*2013*		*2015*		*Total*	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Ha Noi	523	26.9	503	26	464	25.4	490	25.8	545	26.2	2,525	26.1
Phu Tho	197	10.1	208	10.8	196	10.7	202	10.7	204	9.8	1,007	10.4
Hai Phong	140	7.2	148	7.7	144	7.9	135	7.1	174	8.4	741	7.7
Quang Nam	263	13.5	263	13.6	254	13.9	251	13.2	268	12.9	1,299	13.4
Nghe An	108	5.6	105	5.4	110	6	121	6.4	126	6.1	570	5.9
Khanh Hoa	72	3.7	77	4	77	4.2	71	3.7	82	3.9	379	3.9
Lam Dong	68	3.5	56	2.9	61	3.3	71	3.7	71	3.4	327	3.4
Ho Chi Minh	469	24.1	467	24.2	425	23.2	451	23.8	506	24.3	2,318	23.9
Long An	105	5.4	104	5.4	97	5.3	104	5.5	106	5.1	516	5.3
Total	1,945	100	1,931	100	1,828	100	1,896	100	2,082	100	9,682	100

Figure 1: Map of Provinces in the Sample

Table 4: Dynamic Effects on Ouput

Dependent variable: Log of Output	Formal Enterprises		Household Enterprises	
Log of own-industry formal firm				
Lagged by 1 years	0.0103		-0.00745*	
	(0.00950)		(0.00433)	
Lagged by 5 years		0.0214		-0.00852
		(0.0423)		(0.0213)
Log of total workforce	0.185***	0.208***	0.155***	0.136***
	(0.0146)	(0.0178)	(0.0106)	(0.0129)
Log of capital	0.0199***	0.0223**	0.0128***	0.0228***
	(0.00644)	(0.00866)	(0.00377)	(0.00492)
Log of cost intermediate inputs	0.748***	0.727***	0.762***	0.782***
	(0.00793)	(0.00943)	(0.00458)	(0.00589)
% workforce with professional degree	0.00133	0.00309	0.0110	-0.0173
	(0.00238)	(0.00265)	(0.0127)	(0.0135)
Received loan from bank	0.0293*	0.0155	-0.00196	-0.00534
	(0.0155)	(0.0169)	(0.00987)	(0.0121)
% Waged workers	0.000933	0.000780	0.00188***	0.00206***
	(0.000772)	(0.000749)	(0.000211)	(0.000246)
Received government assistance	0.0182	0.0335*	0.00990	0.00525
	(0.0155)	(0.0185)	(0.00964)	(0.0136)
Industry FE	YES	NO	YES	NO
Year FE	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES
Constant	1.435***	1.656***	1.334***	1.297***
	(0.105)	(0.230)	(0.0345)	(0.102)
Observations	2,931	1,969	6,287	3,835
R-squared	0.898	0.902	0.914	0.919
Number of firms	1,385	1,064	2,359	1,803

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

Table 5 : Effects on Odds-Ratio of New Technology Introduction of Food Products Manufacturers

New Technology Introduction	Formal Enterprises			Household Enterprise		
Log of own-industry formal firm						
No Lag	0.154			0.901		
	(0.206)			(0.668)		
Lagged by 1 years		1.479			0.995	
		(0.723)			(0.116)	
Lagged by 5 years			0.695			0.922
			(1.858)			(1.062)
Log of total workforce	1.136	1.138	2.600	1.208	1.211	3.347**
	(0.453)	(0.439)	(2.235)	(0.439)	(0.440)	(1.958)
Log of capital	1.131	1.104	1.513	1.094	1.091	1.552*
	(0.196)	(0.184)	(0.771)	(0.130)	(0.128)	(0.386)
Log of cost intermediate inputs	1.182	1.189	2.024	1.155	1.155	1.021
	(0.276)	(0.271)	(1.192)	(0.174)	(0.174)	(0.271)
% workforce with professional degree	0.971	0.984	0.995	3.350***	3.306***	1.624e+07
	(0.0582)	(0.0605)	(0.106)	(1.484)	(1.429)	(1.832e+10)
Received loan from bank	1.227	1.141	1.525	1.214	1.215	1.407
	(0.548)	(0.501)	(1.282)	(0.323)	(0.324)	(0.583)
% Waged workers	1.017	1.015	1.008	1.005	1.005	1.004
	(0.0181)	(0.0203)	(0.0346)	(0.00657)	(0.00660)	(0.00939)
Received government assistance	1.804	1.618	1.775	1.196	1.197	0.223**
	(0.719)	(0.625)	(1.434)	(0.308)	(0.309)	(0.131)
Year FE	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES
Observations	203	203	100	596	596	272
Number of firms	59	59	37	153	153	96

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

Table 6: Effects of Agglomeration on Outputs of Fabricated Metal Products Manufacturers

Dependent variable: Log of Output	Formal Enterprises			Household Enterprises		
Log of own-industry formal firm						
No Lag	-0.0173			0.0663		
	(0.187)			(0.0441)		
Lagged by 1 years		-0.0206			0.0104	
		(0.0222)			(0.00754)	
Lagged by 5 years			-0.0392			0.0314
			(0.111)			(0.0394)
Log of total workforce	0.188***	0.191***	0.287***	0.145***	0.143***	0.145***
	(0.0463)	(0.0464)	(0.0403)	(0.0212)	(0.0212)	(0.0217)
Log of capital	0.0322*	0.0325*	0.0360*	-0.00609	-0.00626	0.0113
	(0.0191)	(0.0190)	(0.0194)	(0.00700)	(0.00701)	(0.00758)
Log of cost intermediate inputs	0.799***	0.800***	0.751***	0.815***	0.813***	0.832***
	(0.0247)	(0.0246)	(0.0205)	(0.00885)	(0.00883)	(0.00924)
% workforce with professional degree	-0.00898	-0.00841	-0.00542	-0.0413	-0.0424	-0.0163
	(0.00780)	(0.00780)	(0.00712)	(0.0449)	(0.0449)	(0.0425)
Received loan from bank	0.00615	0.00643	-0.0331	-0.00248	-0.00380	-0.0122
	(0.0458)	(0.0457)	(0.0358)	(0.0173)	(0.0173)	(0.0183)
% Waged workers	0.00193	0.00193	0.000377	0.00149***	0.00149***	0.001000***
	(0.00198)	(0.00198)	(0.00126)	(0.000391)	(0.000391)	(0.000373)
Received government assistance	0.0427	0.0435	0.0706	0.0162	0.0138	0.0233
	(0.0486)	(0.0483)	(0.0454)	(0.0177)	(0.0177)	(0.0214)
Year FE	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES
Constant	0.924	0.926***	1.487**	0.870***	1.125***	0.944***
	(1.035)	(0.266)	(0.615)	(0.204)	(0.0627)	(0.180)
Observations	446	446	336	1,128	1,128	689
R-squared	0.899	0.899	0.939	0.948	0.948	0.968
Number of firms	211	211	182	418	418	318

Standard errors in parentheses

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

Table 7: Effects on Odds-Ratio of New Technology Introduction of Fabricated Metal Products Manufacturers

New Technology Introduction	Formal Enterprises			Household Enterprise		
Log of own-industry formal firm						
No Lag	0.961			0.180		
	(2.675)			(0.196)		
Lagged by 1 years		1.898			1.461	
		(1.886)			(0.633)	
Lagged by 5 years			2,565*			366.1**
			(11,341)			(1,094)
Log of total workforce	0.792	0.770	0.0167*	1.098	1.093	0.107
	(0.459)	(0.453)	(0.0370)	(0.669)	(0.657)	(0.252)
Log of capital	1.096	1.079	3.885	1.395*	1.353	0.806
	(0.240)	(0.233)	(3.675)	(0.277)	(0.262)	(0.609)
Log of cost intermediate inputs	1.096	1.089	0.750	0.707	0.728	1.394
	(0.320)	(0.317)	(0.531)	(0.186)	(0.190)	(1.734)
% workforce with professional degree	1.300***	1.290**	1.754**	2.376	2.184	5.167e+08
	(0.131)	(0.131)	(0.475)	(2.039)	(1.862)	(2.294e+12)
Received loan from bank	1.777	1.786	12.22*	3.228**	3.214**	3.376
	(1.165)	(1.153)	(16.88)	(1.620)	(1.620)	(4.593)
% Waged workers	0.971	0.971	0.972	0.999	1.001	0.975
	(0.0275)	(0.0274)	(0.0708)	(0.0121)	(0.0121)	(0.0288)
Received government assistance	4.466**	4.313**	9.416*	1.215	1.272	11.41
	(2.760)	(2.632)	(12.07)	(0.551)	(0.562)	(16.92)
Year FE	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES
Constant	156	156	74	225	225	74
Number of firms	43	43	26	60	60	26

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: the coefficients in columns 3 and 6 are extremely large. This is because only 26 firms are included in these analysis with observations from only 3 years. Within these 26 firms, most of them introduced a new technology in 2011 and not in the later years.