Chulalongkorn Journal of Economics 23, 2011: 39-76

Industrial Agglomeration and Labor Productivity: Evidence from Manufacturing Establishments in Thailand

Phakpoom Tippakoon
Faculty of Economics, Thammasat University, Bangkok,Thailand
Corresponding author: phakpoom1977@gmail.com

Abstract

This paper aims to answer three research questions: do agglomeration economies help improve manufacturing establishments' labor productivity?; what form of agglomeration economies (urbanization or localization) is more conducive to enhancing labor productivity?; and at what sectoral and spatial scopes that agglomeration is most relevant for productivity improvement? To answer these research questions, I use a two-stage least square regression to analyze Thailand's industrial census data for the year 2007. The results from such analysis reveal that industrial agglomeration helps improve establishments' labor productivity. However, the form of agglomeration matters. Localization economies are more conducive to such productivity improvement than urbanization economies. This happens only when a broader-range and complementary activities are spatially agglomerated. In other words, sectoral scope of agglomeration matter.

Keywords: Industrial Agglomeration, Labor Productivity, Urbanization Economies, Localization Economies

1. Introduction

The relationship between spatial agglomeration and establishments' labor productivity is a fundamental issue in industrial agglomeration literature. This issue has been subject to theoretical discussion and empirical investigation over three decades. Empirical evidence regarding the effects of agglomeration economies on productivity from various countries has been increasingly added to the body of literature.[1] However, despite such richness in the body of literature, there are still some controversial and debated issues. First, there exists a theoretical debate about the effects of localization economies versus urbanization economies, and empirical studies still provide contrasting evidence on this issue (Panne, 2004; Rosenthal and Strange, 2004). Second, while the notion that agglomeration economies generate productivity growth is widely accepted, the knowledge about spatial and sectoral scopes in which agglomeration economies take place is not yet well established. Until recently, researchers have made little effort to examine the effects of agglomeration economies at different spatial and sectoral scopes in order to find the scope at which the effects of agglomeration economies on productivity are most vigorous.[2]

In this paper, I take these issues into consideration. First, I empirically investigate the productivity effects of localization economies and those of urbanization economies. Specifically, I test the effects of these two forms of agglomeration separately to see which form is conducive to the increase in manufacturing establishments' labor productivity. Second, taking the issue of spatial and sectoral scopes into account, I examine the effects of industrial agglomeration on establishments' labor productivity at different industrial and spatial units. This is to see whether different spatial and sectoral scopes of agglomeration have different effects on establishments' labor productivity. All in all, the ultimate goal of this paper is to answer the following research questions: do agglomeration economies help improve establishments' labor productivity?; which form of agglomeration (urbanization or localization) is more

---

[1] Empirical studies which examine a direct relationship between agglomeration and productivity are Sveikauskas (1975), Segal (1976), Moomaw (1981), Ciccone and Hall (1996), Capello (1999), Ciccone (2002), Rigby and Essletzbichler (2002), Henderson (2003), Madsen et al. (2003), Cingano and Schivardi (2004), Koo (2005), Liu et al. (2005), Baldwin et al. (2008), Cainelli (2008), and Brown and Rigby (2009). In some other studies, due to the lack of reliable data to directly measure productivity, productivity is indirectly measured by such indicators as employment growth (Glaeser et al., 1992; Henderson et al., 1995), wage premium or wage growth (Glaeser and Mare, 2001; Wheaton and Lewis, 2002; Glaeser and Resseger, 2009), or new-enterprise startups (Rosenthal and Strange, 2003), assuming that productivity is related with those measures (i.e. higher productive regions (firms) tend to exhibit higher employment, wage, and start-up rate than less productive regions (firms)). See Rosenthal and Strange (2004) for a comprehensive literature review.

[2] Notable studies include Rosenthal and Strange (2003), Baldwin et al. (2008), and Brown and Rigby (2009).

conducive for productivity improvement?; and at what sectoral and spatial scopes that agglomeration is most relevant for productivity improvement?

This paper is structured as follows. Section 2 presents some theoretical models used for analyzing the relationships between industrial agglomeration and labor productivity, and discusses some hypotheses based on theoretical and empirical literature. Section 3 provides some discussions on data and variable construction. Section 4 discusses relevant methodological issues. Section 5 presents and discusses the results of regression analysis. Finally, Section 6 concludes the paper.

2. Model and Hypotheses

To estimate the effects of industrial agglomeration on the labor productivity of manufacturing establishments, I begin with a standard production function:

$$Y_{ijr} = A_{ijr}K_{ijr}^{\alpha}L_{ijr}^{\beta} \tag{1}$$

where Y_{ijr}, K_{ijr}, and L_{ijr} are respectively, value-added, capital stock, and labor force of establishment *i* embedded in industry *j* and region *r*. The term A_{ijr} denotes the state of technology of the establishment, which is assumed to be influenced by agglomeration economies (i.e. localization and urbanization economies) as well as establishment-specific characteristics (Henderson, 2003; Moretti, 2004; Martin et al., 2008) and can be modeled as:

$$A_{ijr} = (LE)_{jr}^{\gamma}(UE)_{r}^{\delta}X_{ijr}^{\lambda} \tag{2}$$

where LE_{jr} is localization economies generated from the agglomeration of industry *j* in region *r*; UE_r is urbanization economies generated from the agglomeration of all industries in region *r*; X_{ijr} denotes a set of factors which may influence the establishment's state of technology.[3] Thus, equation (2) assumes that establishment i's state of technology not only depends on its specific assets, X_{ijr}, but also on its immediate environment in terms of localization and urbanization economies (Martin et al., 2008). Obviously, the equation assumes that productivity effects of industrial agglomeration are generated from two sources – agglomeration of establishments in the same industry and agglomeration of establishments in different industries.

Next, we can divide equation (1) by L_{ijr} to give a labor productivity function:

$$y_{ijr} = \frac{Y_{ijr}}{L_{ijr}} = A_{ijr}K_{ijr}^{\alpha}L_{ijr}^{\beta-1} \tag{3}$$

---

3 X_{ijr} can be thought of in terms of establishment's specific assets such as its participation in international trade, foreign investment, and investment in research and development.

where the lower case y_{ijr} denotes establishment *i*'s value-added per employee which is a measure of labor productivity used in this study.

To specify a testable econometric model, equations (2) and (3) are transformed into a linear function using natural logarithm. This process results in the following equations:

$$ln\ A_{ijr} = \gamma lnLE_{jr} + \delta lnUE_{r} + \lambda x_{ijr} \tag{4}$$

and

$$ln\ y_{ijr} = ln\ A_{ijr} + \alpha ln\ K_{ijr} + (\beta\text{-}1)ln\ L_{ijr} \tag{5}$$

where lower case x_{ijr} denotes the log of X_{ijr} which is taken as control variables in this study. By substituting equation (4) into equation (5), an extended equation is produced as follows:

$$ln\ y_{ijr} = \alpha ln\ K_{ijr} + (\beta\text{-}1)ln\ L_{ijr} + \gamma lnLE_{jr} + \delta lnUE_{r} + \lambda x_{ijr} \tag{6}$$

Equation (6) considers manufacturing establishments' labor productivity as a function of their capital investment, employment of labor, and other establishment-specific assets as well as localization and urbanization economies generated from, respectively, their co-location with other establishments in the *same* industry and their co-location with other establishments from *different* industries. I will now provide some theoretical discussion and draw hypotheses regarding the productivity effects of localization and urbanization economies as well as control variables as follows.

How do we expect the coefficients of localization economies (γ) and urbanization economies (δ)? Regarding localization economies, it follows from Marshall's (1920) observation that productivity can be enhanced when sectorally related firms are spatially agglomerated. According to him, such agglomeration generates *pecuniary externalities* because specific goods and services provided by specialized suppliers and workers with specific skills are always available and can be acquired at relatively low costs. Additionally, knowledge and information spillovers can occur easily in clusters where firms undertake related activities or share some basic understanding of specific industrial production, which allows for the transfer of industrial specific knowledge. Moreover, the agglomeration of related firms can also facilitate face-to-face interactions, and thus allows for the transfer of tacit knowledge which cannot be easily transferred by codification methods (Lissoni, 2001; Dahl and Pedersen, 2004). The positive effects of localization economies are proven by some empirical studies such as Nakamura (1985) and Martin et al. (2008).[4] However, it can be argued that benefits from own-industry agglomeration may be offset by the costs associated with an increased competition between firms in the same sector: when sectorally related firms are clustered, the degree of competition for labor and inputs may increase because similar firms need similar production factors (Lall et al., 2004). Thus, as localization economies may generate both benefits as well as costs, it is possible to expect either positive or negative effect of this variable on labor productivity.

The urbanization economies thesis differs from that of localization economies since it sees the spatial agglomeration of firms from different industries as key to productivity enhancement (Panne 2004: 595). The reason for this is that a city with diversified industrial structure can facilitate a transmission of knowledge and ideas across different lines of work. Firms embedded in such environment can benefit from the exchanges of different information, knowledge, and ideas that are new to them and are vital for creativity, innovation, and productivity (Jacobs 1969). Some empirical studies (e.g. Sveikauskas, 1975; Tabuchi, 1986; Glaeser et al., 1992; Cicone and Hall, 1996; Tabuchi and Yoshida, 1999) show that urbanization economies have significant effects on productivity improvement.[5] However, it can be argued that as the city grows larger, the benefits from urbanization economies may be offset by the costs of agglomeration (e.g. increased wage rates, land rents, and commuting time) which may result in a decline in firms' productivity (Carlino, 1979; Lall et al., 2004; Baldwin et al., 2008). Thus, in this study, either positive or negative effect of urbanization economies can be expected.

[4] Nakamura (1985) uses cross-section data of Japanese cities in 1979 and estimates the effects of agglomeration economies on productivity. He shows that heavy industries receive more productivity benefits from localization economies than from urbanization economies. Recent work by Martin et al. (2008) uses French firm-level panel data to estimate the effects of localization economies (defined as the employment of neighboring firms in the same industry), controlling for unobserved firm, industry, and regional heterogeneities. They find that a 10% increase of employment in neighboring firms of the same industry increases a firm's productivity by approximately 0.4-0.5%.

[5] Sveikauskas (1975) uses population size as a measure for urbanization and finds that a doubling of population size is associated with a 5.98% increase in labor productivity. Tabuchi (1986) uses Japanese city data and finds a doubling of population density – a measure of urbanization – increases labor productivity by about 4.3%. Glaeser et al. (1992) intentionally test localization economies against urbanization economies and find that urban diversity, not specialization, encourages employment growth - a proxy for productivity growth. Cicone and Hall (1996) establish that the relationship between employment density and productivity does exist. They empirically show that a doubling of employment density increases average labor productivity by 6%. Tabuchi and Yoshida (1999) examine the effects of agglomeration economies on consumption and production sides, using Japanese city-based data in 1992. They find that doubling city size increases the nominal wage by approximately 10% and argue that such increase is associated with an increase in productivity.

What about the effects of control variables, x_{ijr}, on establishments' labor productivity? In this study, control variables include structural and establishment-specific factors, namely localized competition ($COMP_{jr}$), export (EXP_{ijr}), import (IMP_{ijr}), foreign investment (FDI_{ijr}), organizational structure ($SING_{ijr}$), and investment in research and development (RND_{ijr}).

The first control variable is localized competition ($COMP_{jr}$) (i.e. the degree of competition in regional industry). The effect of localized competition on establishments' productivity is not yet clear and is still subject to on-going debates. On one hand, some scholars (e.g., Arrow, 1962; Romer, 1986) argue that monopolistic structure of regional industry is necessary for technological improvement. They maintain that knowledge spillovers are non-rival market externalities whose positive effects overflow to neighboring firms through one firm's innovation. Lack of property rights protection for innovative activities and of appropriate compensation to the innovators will reduce a firm's incentives to innovate, and consequently will slow down technological development. This theory predicts that technological development in regional industry will be faster if local industrial structure exhibits monopolistic behavior because it allows firms to appropriate the economic value accruing from their innovative activities (Feldman and Audretsch, 1999; Combes, 2000). On the other hand, it is argued that local industrial competitive structure is more conducive to knowledge spillovers and technological development than is monopolistic structure (Jacobs 1969, 1984; Porter, 1990). According to Porter (1990), firms embedded in a competitive environment are forced to innovate otherwise they will not be able to compete with their innovative neighbors. Fierce competition will lead to an improvement in existing technologies and to a rapid adoption of new technologies, which are necessary for industrial growth (Glaeser et al., 1992; Gao, 2004). Thus, in this analysis, we can expect the coefficient of $COMP_{jr}$ to be either positive or negative.

Some establishment-specific characteristics are also relevant for productivity improvement. It is argued that export can improve firms' productivity through a learning-by-exporting process: firms participating in the export market can learn from their international buyers and their competitors; consequently, knowledge flows from buyers and competitors help to improve the post-entry performance of export starters (Flyges and Wagner, 2008; Aw and Hwang, 1995; Greenaway and Kneller, 2004).[6] Imports can also help firms to improve their productivity because, by importing new intermediate products from foreign markets, firms can expand their domestic product scope through the introduction of new product varieties, which generates dynamic gains from trade (Goldberg et al., 2008). Furthermore, importing more advanced intermediate inputs allows firms to learn new technologies which, consequently, help enhance their

[6] See Wagner (2007) for a very extensive review of empirical literature on the relationship between export and firm-level productivity.

technological capabilities (Vogel and Wagner, 2008). Thus, in the current study, I expect that establishments that export (EXP_{ijr}) or import (IMP_{ijr}) will be more productive than those that do not.

Foreign direct investment (FDI) literature maintains that foreign ownership is a significant factor for technological progress and productivity growth (Caves, 2007).[7] It argues that in order to compete with local firms (that have better knowledge and information about local markets), foreign firms rely on "proprietary assets" (i.e. superior managerial and technological capabilities) which are intangible and are more likely to be efficiently transferred through internalization and expansion abroad. Foreign affiliates are said to benefit from the transfer of these assets from their parent firms and thus are more likely than local firms to be more efficient (Benfratello and Sembenelli, 2006). Empirical literature also notes the importance of technological advantages of parent firms as a key to the better performance of foreign affiliates (Siripaisalpipat and Hoshino, 2000). Based on these arguments, therefore, it is possible to expect that establishments which have foreign investment (FDI_{ijr}) will be more productive than those which do not have foreign investment.

Establishments' organizational structure is another factor which may affect labor productivity. Establishments embedded in the multi-establishment firm structure can benefit from knowledge spillovers circulated among establishments of a given firm (Martin et al., 2008). Learning from other establishments in the same firm structure can help them to enhance their technological capability (Henderson, 2003). Consequently, the productivity tends to be higher for such establishments categorized as branch, affiliated company, or subsidiary than those independent establishments. In the context of this study, independent establishments ($INDEP_{ijr}$) are expected to be less productive than branch or affiliated companies.

It has been established that investment in research and development (R&D) by firms can help improve their productivity (Grilliches, 1986). Economists believe that firms' own investment in R&D can directly enhance their technological capability in generating new knowledge, information, and products. Moreover, R&D investment can also enhance firms' absorptive capacity, i.e. ability to identify, assimilate, and exploit knowledge from the environment (Cohen and Levinthal, 1989). Thus, we can expect that establishments that invest in R&D (RND_{ijr}) will exhibit higher labor productivity than those that do not.

---

[7] Literature on FDI also suggests that foreign investment has spillover effects (Kohpaiboon, 2006; Yokota, 2008). According to this literature, FDI inflows bring new technologies and know-how to the host country and its technologies spill over to domestic firms through three channels: demonstration, linkages, and labor mobility (Kohpaiboon, 2009). However, it is argued that vertical linkages of foreign and domestic firms are found to be the most efficient form of FDI technology spillovers (Rodriguez-Clare, 1996; Javorcik, 2004).

Finally, I include dummy variables for industrial category (IND_j) and for region (REG_r) in the equation. The inclusion of IND_j is to control for unobserved industrial effects which may influence establishments' labor productivity (e.g. macro-economic policies on the industry, technological progress at the industrial level, and industry life-cycles). REG_r is also included in the equation to capture unobserved regional characteristics which may affect establishments' productivity (e.g. regional policies, infrastructure, and resource endowment).

Combining all of the above variables, we can form a full econometric model to be tested as follows (theoretical expected sign of coefficient is in parenthesis):

$$\begin{aligned} ln\, y_{ijr} = {} & \alpha ln\, K_{ijr} + (\beta\text{-}1) ln\, L_{ijr} + \gamma lnLE_{jr} + \delta lnUE_r + \lambda_1 COMP_{jr} + \lambda_2 EXP_{ijr} \\ & + \lambda_3 IMP_{ijr} + \lambda_4 FDI_{ijr} + \lambda_5 INDEP_{ijr} + \lambda_6 RND_{ijr} + \pi_1 IND_j + \pi_2 REG_r + \varepsilon_{ijr} \end{aligned} \tag{7}$$

where

$ln\, y_{ijr}$	=	Labor productivity of establishment *i* in industry *j* and region *r*
lnK_{ijr}	=	Fixed assets of establishment *i* in industry *j* and region *r*
$ln\, L_{ijr}$	=	Number of workers of establishment *i* in industry *j* and region *r*
$lnLE_{jr}$(+/–)	=	Localization economies of industry *j* and region *r*
$lnUE_r$(+/–)	=	Urbanization economies of region *r*
$COMP_{jr}$(+/–)	=	Localized competition of industry *j* and region *r*
EXP_{ijr}(+)	=	Dummy for export for establishment *i* in industry *j* and region *r*
IMP_{jr}(+)	=	Dummy for import for establishment *i* in industry *j* and region *r*
FDI_{ijr} (+)	=	Dummy for foreign share in establishment *i* industry *j* and region *r*
$INDEP_{ijr}$(–)	=	Dummy for independent establishment for establishment *i* industry *j* and region *r*
RND_{ijr}(+)	=	Dummy for establishment's laboratory unit for establishment *i* industry *j* and region *r*
IND_j	=	Dummies for industrial category (IND_j = 1 for industrial category *j* and 0 for other categories
REG_r	=	Dummy for region (REG_r = 1 for region *r* and 0 for other regions)
ε_{ijr}	=	A stochastic error term containing other factors which affect lny_{ijr}

3. Data and Variable Construction

3.1 Data Source

This analysis relies mainly on two industrial census data, i.e. industrial censuses 1997 and 2007, provided by the National Statistical Office of Thailand (NSO). These two data sets contain the population of manufacturing establishments that existed in

1996 and 2006. The numbers of establishments in the 1997 and 2007 data sets are 32,489 and 73,931, respectively. The advantage of these data sets lies in the fact that they represent the population of manufacturing establishments of all sizes in both years. Thus, there is no problem of selection bias in favoring a particular group of establishments. However, to protect the confidentiality of private information, NSO does not provide names and addresses of establishments. Thus, it is impossible to trace the presence of the same establishments in both years.[8] Having realized such limitation, I decided to use the cross-section analysis based on the 2007 data set. However, the 1997 data set is also utilized by selecting some variables and using them as instrumental variables in the two-stage least square regression (2SLS) (see section 4).

At the establishment level, some balance-sheet data (e.g., employment, capital, exports, sales, intermediate costs, and wage) is available. Information on establishment location at various regional levels (district, province, subregion, and region), industry classification (at 2-, 3- and 4-digit levels), and establishment structure (foreign investment and legal status) is also provided. This information is sufficient to construct variables of our interest.

3.2 Variable Construction

a. Dependent variable: *labor productivity*

The dependent variable is manufacturing establishment's labor productivity which is defined as the logarithm of establishment's value-added (in Thai baht) divided by the number of workers employed. Thus, lny_{ijr} is constructed as follows:

$$ln\, y_{ijr} = \ln\left(\frac{(Value\ added)_{ijr}}{(Number\ of\ workers\ employed)_{ijr}}\right)$$

where value-added is calculated by taking the difference between establishment's sales and its intermediate costs (in Thai baht); and workers here refer to all fulltime workers who are employed in both production and non-production processes.

b. Localization economies variable

As discussed above, localization economies are theoretically defined as spatial agglomeration of manufacturing establishments operating in the same sector. Thus, for establishment *i* embedded in sector *j* and region *r*, the localization economies variable is defined as:

[8] With this limitation, we cannot organize the data into a panel data set and conduct statistical analysis to control for firms' unobserved heterogeneity which is a common methodological problem in the productivity analysis at the firm level (see Combes et al., 2008a; Matin et al., 2008)

$$lnLE_{jr} = ln(EST_{jr})$$

where EST_{jr} denotes the number of establishments in sector *j* and region *r*. In particular, $lnLE_{jr}$ is constructed by taking natural logarithm of the number of manufacturing establishments operating in the same sector and located in the same region. $lnLE_{jr}$ is measured at the regional industry level.

c. Urbanization economies variable

To capture the diversity of industrial structure of a region – theoretical definition of urbanization economies – a Herfindahl Index (*HI*) is used and is defined as follows:

$$HI_r = \sum_j \left(\frac{EST_{jr}}{EST_r}\right)^2$$

where EST_r is the number of manufacturing establishments (all sectors) in region *r*, and EST_{jr} is as previously defined. This index measures the degree to which industrial structure of region *r* is diversified. *HI* takes the continuous value between zero and one: "$HI = 0$" means that industrial structure of the region is perfectly diversified, while "$HI = 1$" means that industrial structure in the region is occupied by a single industry.

To interpret urbanization economies variable ($lnUE_r$) in terms of elasticity (as with localization economies variable), I take a natural logarithm of a reverse *HI* as follows:

$$lnUE_r = \ln\left(\frac{1}{HI_r}\right)$$

$lnUE_r$ now is a continuous variable taking a value between zero and infinity (Martin et al., 2008). The degree of industrial diversity increases as the value of $lnUE_r$ increases. Unlike the localization economies variable, the urbanization economies variable is measured at the regional level. It signifies the extent to which *overall* industrial structure of region *r* is diversified.

As noted above, the effects of agglomeration may transmit across industrial and spatial scopes. To deal with this issue, I measure localization and urbanization variables at different spatial and sectoral scopes. For spatial scope, province and subregion are taken as measurement units. The province is an administrative entity. Our data set contain all 76 provinces in Thailand. The subregion is a group of contiguous provinces which is classified by the National Economic and Social Development Board (NESDB).[9] Based on this classification, there are 18 subregions. However, in this analysis, Bangkok and its five vicinity provinces are separated from the NESDB's original classification and used to form another subregion (called the Bangkok Metropolitan Region: BMR). This is because this group of provinces is the largest industrial agglomeration area in the country and is different from other groups of provinces. Therefore, in total we have 19 subregions. For sectoral scope, I use three levels of industrial classification (i.e. 2-digit, 3-digit, and 4-digit industries) as industrial units to measure localization and urbanization economies. Hence, for subscripts *j* and *r* in $lnLE_{jr}$ and $lnUE_r$, *j* has three units and *r* has two units of measurement. When we intersect three industrial units with two regional units, we get six entities in which manufacturing establishments are embedded: 2-digit provincial industry, 2-digit subregional industry, 3-digit provincial industry, 3-digit subregional industry, 4-digit provincial industry, and 4-digit subregional industry (Figure 1). When the effects of localization and urbanization economies are measured, they are measured at all of these six entities.

Figure 1: Regional and Industrial Units used for Constructing $lnLE_{jr}$ and $lnUE_r$

		Industrial unit		
		2-digit	**3-digit**	**4-digit**
Regional unit	**Province**	2-digit provincial industry	3-digit provincial industry	4-digit provincial industry
	Subregion	2-digit subregional industry	3-digit subregional industry	4-digit subregional industry

Source: Author

At each spatial and sectoral entity, we have localization and urbanization variables defined as follows:

[9] According to the NESDB, the grouping of provinces is not done primarily for administrative reasons but for regional economic development. In a nutshell, this is based on the ideas that provinces with similar economic characteristics should have similar development strategies; resources necessary for economic development should be shared among those provinces; and development agencies in those provinces should be coordinated. This is one of the area-based or cluster-based development strategies recently endorsed by the NESDB.

(1) $lnLE_{jr}\_1$ = localization economies measured at 2-digit and provincial levels;
(2) $lnUE_{r}\_1$ = urbanization economies measured 2-digit and provincial levels;
(3) $lnLE_{jr}\_2$ = localization economies measured at 3-digit and provincial levels;
(4) $lnUE_{r}\_2$ = urbanization economies measured at 3-digit and provincial levels;
(5) $lnLE_{jr}\_3$ = localization economies measured at 4-digit and provincial levels;
(6) $lnUE_{r}\_3$ = urbanization economies measured at 4-digit and provincial levels;
(7) $lnLE_{jr}\_4$ = localization economies measured at 2-digit and subregional levels;
(8) $lnUE_{r}\_4$ = urbanization economies measured at 2-digit and subregional levels;
(9) $lnLE_{jr}\_5$ = localization economies measured at 3-digit and subregional levels;
(10) $lnUE_{r}\_5$ = urbanization economies measured at 3-digit and subregional levels;
(11) $lnLE_{jr}\_6$ = localization economies measured at 4-digit and subregional levels;
(12) $lnUE_{r}\_6$ = urbanization economies measured at 4-digit and subregional levels.

d. Localized competition

To measure the degree of localized competition (or competitive market structure of industry j in region r), I use the Herfindahl Index (HI) of market share concentration, which is defined as:

$$HI_{jr} = \sum_{i \in J_r} \left(\frac{S_{ijr}}{S_{jr}} \right)^2$$

where S_{ijr} is the sales of establishment i in industry j and region r; S_{jr} the total sales of all establishments in industry j and region r; and J_r the set of establishments belonging to industry j in region r. HI_{jr} is a summary measure of the market share of each establishment in the regional industry relative to the whole regional industry market. Its value ranges from 0 to 1. $HI_{jr} = 0$ when all establishments in a regional industry have the same market share; $HI_{jr} = 1$ when the whole market share of a regional industry is dominated by only one establishment. Based on the Herfindahl Index, the localized competition variable ($COMP_{jr}$) is constructed as follows:

$$COMP_{jr} = \ln\left(\frac{1}{HI_{jr}} \right).$$

The value of $COMP_{jr}$ ranges between zero and infinity. The increase in $COMP_{jr}$ signifies the increase in the degree of localized competition. This variable is measured at regional industry level and its coefficient can be interpreted in terms of elasticity.

e. Establishment-level variables

At the establishment level, I use seven variables. Each one is defined as follows. First, variable lnK_{ijr} is defined as the natural logarithm of the book value of establishment *i*'s fixed assets valued at the beginning of 2006 (in Thai baht). Second, variable lnL_{ijr} is the natural logarithm of the number of fulltime workers (both production and non-production workers) employed by the establishment. Third, variables EXP_{ijr} and IMP_{jr} are constructed as dummy variable taking the value of one if establishment *i* exports its products (or import products from abroad); otherwise they take the value of zero. Fourth, FDI_{ijr} is a dummy variable for foreign investment share in establishment *i*. This variable takes 1 if establishment *i* has foreign share (no matter how much the share is), and 0 if it has no foreign investment share. Fifth, variable $INDEP_{ijr}$ is a dummy variable for independent establishment coded 1 if an establishment is an independent establishment and 0 if it is a branch or affiliated company. Finally, variable RND_{ijr} is defined as a dummy variable indicating whether or not establishment *i* invests in R&D. It is coded as 1 if the establishment invests in R&D and coded as 0 if it does not.

f. Region and industrial category dummies (IND_j and REG_r)

Region and industrial category are constructed as multiple dummy variables. I construct region dummies for six of Thai regions (namely, the BMR, Centre, North, Northeast, South, and East), taking BMR as a base variable (BMR = 0). In the same way, I construct dummies for four industrial categories including *resource-based, labor-intensive, machinery*, and *Metal, chemical*, and *paper* industries, taking resource-based category as a base variable (resource-based = 0).[10] Note that both REG_r and IND_j are included to capture unobserved sectoral and regional factors which may affect establishments' labor productivity. Their coefficients are not in the interest of this study. Variables used in the regression analysis and their construction are summarized in Table 1.

[10] Following Yokota (2008), I divide 23 Thai manufacturing industries into 4 groups as follows: (1) *Resource-based* industry include food products and beverages (TSIC15), tobacco products (TSIC16), woods and products of wood (TSIC20), coke and refined petroleum products (TSIC23), rubber and plastic products (TSIC25), other non-metallic mineral (TSIC26); (2) *Labor-intensive industry* consists of textiles (TSIC17), wearing apparels and dressing (TSIC18), leather and leather products (TSIC19), publishing, printing and reproduction of records (TSIC22), basic metals (TSIC27), furniture (TSIC36), and recycling (TSIC37); (3) *Machinery industry* includes machinery and equipment n.e.c. (TSIC29), office, accounting and computing machineries (TSIC30), electrical machineries and apparatus (TSIC31), radio, television and communication equipments (TSIC32), medical, precision and optical instruments, watches and clocks (TSIC33), motor vehicles, trailers and semi-trailers (TSIC34), other transport equipments (TSIC35); and (4) *Metal, chemical, and paper industry* include paper and paper products (TSIC21), chemicals and chemical products (TSIC24), and fabricated metal products (TSIC28).

4. Empirical Strategy

4.1 Methodological Issues

In estimating the effects of agglomeration economies on establishment's productivity, researchers often encounter an *endogeneity* problem (Combes et al., 2008a). Econometrically, the problem of endogeneity arises when one (or more) explanatory variable(s) is/are correlated with the error term in the regression model (i.e., ε_{ijr} in equation (7)), causing the Ordinary Least Square (OLS) estimator(s) to be biased (Wooldridge, 2006). In the empirical research on the relationship between agglomeration economies and productivity, the problem of endogeneity is said to be generated by two factors: unobserved heterogeneity and simultaneity (Combes et al., 2008a; Martin et al., 2008).

a. Unobserved heterogeneity

The problem of unobserved heterogeneity arises when some characteristics of establishment, industry, and location which can be related to the productivity of establishment and to some other explanatory variables are omitted from the model for various reasons such as lack of data and measurement problem. In this case, those unobserved characteristics are put into the error terms ε_{ijr}, causing ε_{ijr} to be correlated with explanatory variables. Consequently, estimating the model using OLS regression can give biased and inconsistent estimators (Wooldridge, 2006).

In the context of this study, the unobserved heterogeneity problem can take place at establishment, location, or industry levels. At the establishment level, for example, such variables as entrepreneurial and management skills and labor ability, which are correlated with establishment's productivity, are put into the error terms, as they are not observable or measurable. It is possible to consider those unobserved characteristics as being correlated with industrial agglomeration variables in our model (LE_{jr} and UE_r). For instance, entrepreneurs, managers, and workers who are embedded in the industrial cluster may be able to learn from their neighbors, which can enhance their ability. In this case, the variables LE_{jr} and UE_r can be potentially correlated with the error term, ε_{ijr}; and consequently, parameters γ and δ can be biased and inconsistent.[11]

[11] Martin et al. (2008) also notes that if an entrepreneur is less risk-averse than others, he might tend to distort the labor-capital mix in a particular way, have different innovation strategies and also tend to seek more risky but more lucrative markets. As a result, parameters α and β can be biased as well.

Moreover, some location and industry factors such as local climate, transportation infrastructure, natural resources, and industrial (positive and negative) shocks can in many ways affect the value-added of manufacturing establishments. At the same time, a region endowed with well-developed physical and industrial infrastructures (e.g., specialized education institution, and investment promotion schemes) can be attractive for establishments as well. Thus, the correlations between these unobserved locational and industrial characteristics and variables LE_{jr} and UE_r may exist, causing parameters γ and δ to be biased and inconsistent (Combes et al., 2008a; Martin et al., 2008).

Table 1: Variables Used in the Regression Analysis

Variables	Variable Names	Variable Construction
1. Dependent variable: (lny_{ijr})	Labor productivity	ln(Value added/Number of workers employed)
2. Independent variables		
lnK_{ijr}	Capital	ln(Values of fixed assets)
$ln\ L_{ijr}$	Labor	ln(Number of workers)
$lnLE_{jr}$	Localization economies	ln(Number of establishments in the same industry and same region)
$lnUE_{r}$	Urbanization economies	ln[1/]
$COMP_{jr}$	Localized competition	ln[1/]
EXP_{ijr}	Export	Dummy (1 = export; 0 = not export)
IMP_{ijr}	Import	Dummy (1 = import ; 0 = not import)
FDI_{ijr}	Foreign investment share	Dummy (1 = having foreign investment share; 0 = no foreign investment share)
$INDEP_{ijr}$	Independent establishment	Dummy (1 = independent; 0 = otherwise)
RND_{ijr}	Research and Development	Dummy (1 = having laboratory unit; not having laboratory unit)
IND_{j}	Industrial category	Industrial category dummies (1 for industry *j* and 0 otherwise)
REG_{r}	Region	Region dummies (1 for region *r* and 0 otherwise)

Source: Author

b. Simultaneity

In an econometric sense, the problem of simultaneity occurs when one or more of the explanatory variables is/are jointly determined with the dependent variable (Wooldridge, 2006). In the case of this study, it can be considered that localization and urbanization economies variables may be jointly determined along with labor productivity. For instance, highly productive establishments may tend to be located in the industrial cluster, and through the learning process in the cluster, establishments may be able to improve their productivity. In this context, the relationship between industrial agglomeration and establishment's labor productivity is not unidirectional – they reinforce each other.

In the empirical studies examining the effects of agglomeration economies on firms' productivity, it is found that the simultaneity problem can occur through many channels. For example, Martin et al. (2008) note that the negative (or positive) shocks in the region or in the industry may cause firms to close (or open) or lay off (or hire) employees which in some way affect both firms' productivity and degree of agglomeration. Also, when productivity is measured in terms of labor wage, the reverse causality between agglomeration and wage is present. According to Combes et al. (2008b), more productive labor tends to be agglomerated in the larger, denser, and more skilled local labor market. Agglomeration of highly productive labor creates inter-regional wage differentials. This finding implies that firms that decide to locate in the industrially agglomerated areas in order to utilize high skilled labor are those that can afford to pay high wages, and tend to be highly productive firms.

4.2 Regression Method

As OLS estimator may potentially be biased and inconsistent in the presence of endogeneity, in empirical work it is common to address this problem by using the Two-Stage Least Square (2SLS) regression. This involves finding *instrumental variables* that are correlated with the endogenous explanatory variable(s) but not with the residuals (i.e. such variables are said to be exogenous) (Combes et al., 2008a). The first stage is to regress, based on OLS procedure, the suspected endogenous explanatory variable on instrumental variable(s) and all exogenous variables in the model to obtain the expected values. Then, regression analysis is run with the endogenous explanatory variable replaced by their expected values to obtain the 2SLS estimator. With the best instrumental variable (i.e. variable exhibiting a very strong correlation with endogenous variable and having no correlation with the error term), the 2SLS estimator is proven to be *asymptotically* unbiased and consistent.[12]

[12] However, using 2SLS regression also has a price. Due to the fact that 2SLS relies on the estimated values of endogenous variable, it usually generates larger standard errors than does the OLS. Very often, it results in insignificant estimators. For extensive discussions on procedures and properties of 2SLS regression, see Wooldridge (2002) and Wooldridge (2006).

The usage of instrumental variables differs among researchers, depending on the data researchers have in hand and on how variables are expected to meet requirements to be good instruments. Yet, a common practice found in many previous studies is to use a time-lagged endogenous variable. Again, there is no exact rule on the length of time an endogenous variable should be lagged. It is more likely to depend on the available data. For example, Combes et al. (2008b) examines the relationship between productivity (in terms of workers' earnings) and employment density. They address the endogeneity problem by using employment density with four-decade lagged time. Rice and Venables (2004) estimate the effects of agglomeration (measured by population size) on productivity and earnings in Great Britain's regions during the period 1995-2001. In their study, the number of regional population in 1851 is used as an instrumental variable. Moreover, to instrument the current level of population density in GB's regions, Anastassova (2006) even uses a longer lagged period (i.e. regional population density in 1801) than that is used in Rice and Venables (2004).

The usage of a lagged endogenous variable as an instrumental variable has some advantages. First, it ensures that the reverse causality will no longer be a problem. For example, past agglomeration may affect current levels of productivity, but not vice versa. Second, with a long-time lag, we can be sure somehow that correlation between the lagged variable and the error term will not be present (or will not be very strong). For instance, the level of agglomeration 50 years ago should have no correlation (or very weak correlation) with the firm's current unobserved ability. Therefore, in this analysis I decide to instrument industrial agglomeration variables (LE_{ij} and UE_r) using their lagged values. Specifically, the level of agglomerations in 2006 will be instrumented by the level of agglomerations in 1996.[13]

It is worth noting that, the 2SLS estimator is less efficient than the OLS estimator when the explanatory variables are exogenous. Therefore, it is important to test for the presence of endogeneity before proceeding to using instrumental variables (Wooldridge, 2002). In this study, an endogeneity test is conducted based on the *Hausman's test* procedures. The results from this procedure show that the endogeneity problem in our agglomeration variables is present, which justify our usage of 2SLS regression. (See Appendix 1 for the procedures and results of *Hausman's test*).

[13] Bivariate correlations between each LE_{jr} or UE_r variable and their ten-year lag range from 0.512 (in case of $LE_{jr}\_4_{2006}$ and $LE_{jr}\_4_{1996}$) to 0.923 (in case of $UE_r\_4_{2006}$ and $UE_r\_4_{1996}$). Thus, ten-year lags can be taken as instrumental variables because their correlations with our agglomeration variables are not weak at all.

5. Results

5.1 Descriptive Statistics and Bivariate Correlations

Before conducting regression analysis, the data set was explored in order to remove some problematic cases. The cases were removed if they (1) contain a missing value in any variable; (2) are duplicate cases; (3) are cases of establishment with no workers or no value added which make the dependent variable ln y_{ijr} mathematically undefined[14]; or (4) contain some suspicious values (e.g. zeros in sales, fixed asset value, or intermediate costs or very extreme values). Based on these criteria, 8,904 cases were removed from the data set; hence, 65,027 cases remained to be used for regression analysis. Descriptive statistics for key variables used in the regression and their bivariate correlations are given in Tables 2 and 3, respectively.

As can be seen from Table 3, each independent variable, in general, exhibits a highly significant correlation with a dependent variable (i.e. all pairs of bivariate correlation are significant at 1% level). Some variables have bivariate correlation signs which run against our expectations. For example, variables EXP_{ijr}, IMP_{ijr}, and FDI_{ijr} have negative correlations with dependent variable. However, this can be changed when we run multiple regression which takes all variables' effects into account simultaneously. It can also be noted from Table 3 that correlations between each pair of independent variables are not extremely high, thereby no serious multicollinearity problem.[15]

---

[14] Note that lny_{ijr} is defined as *ln(value add/workers)*. This variable is mathematically undefined if number of worker (or value added) is zero.

[15] A high correlation among industrial agglomeration variables should not cause multicollinearity problem as well, because all industrial agglomeration variables will not be put together in the same model.

Table 2: Descriptive Statistics for Key Variables used in Regression Analysis

	Minimum	Maximum	Mean	Std. Deviation
lny_{ijr}	0.23	23.45	10.99	3.42
lnK_{ijr}	6.21	23.61	13.93	2.49
lnL_{ijr}	0.00	9.17	2.02	1.78
$lnLE_{jr}\_1$	0.69	7.37	4.90	1.17
$lnUE_{r}\_1$	1.21	2.56	2.06	0.30
$lnLE_{jr}\_2$	0.69	7.37	4.24	1.22
$lnUE_{r}\_2$	2.01	3.20	2.71	0.23
$lnLE_{jr}\_3$	0.00	7.37	3.65	1.32
$lnUE_{r}\_3$	2.14	3.75	3.18	0.32
$lnLE_{jr}\_4$	0.69	8.00	6.19	1.14
$lnUE_{r}\_4$	1.51	2.58	2.14	0.32
$lnLE_{jr}\_5$	0.69	7.71	5.49	1.20
$lnUE_{r}\_5$	2.42	3.13	2.82	0.23
$lnLE_{jr}\_6$	0.00	7.63	4.83	1.34
$lnUE_{r}\_6$	2.66	3.76	3.33	0.28
$COMP_{jr}$	0.00	3.94	2.01	0.90
EXP_{ijr}	0.00	1.00	0.08	0.28
IMP_{ijr}	0.00	1.00	0.09	0.29
FDI_{ijr}	0.00	1.00	0.04	0.19
$INDEP_{ijr}$	0.00	1.00	0.93	0.25
RND_{ijr}	0.00	1.00	0.03	0.17
# Obs.	65,027			

Source: Author's calculation

Table 3: Correlation Matrix of Variables used in the Regression Analysis

	lny_{ijr}	lnK_{ijr}	lnL_{ijr}	$lnLE_{jr}\_1$	$lnUE_{r}\_1$	$lnLE_{jr}\_2$	$lnUE_{r}\_2$	$lnLE_{jr}\_3$	$lnUE_{r}\_3$	$lnLE_{jr}\_4$	$lnUE_{r}\_4$
lny_{ijr}	1										
lnK_{ijr}	.649[a]	1									
lnL_{ijr}	-.602[a]	-.089[a]	1								
$lnLE_{jr}\_1$	-.012[a]	.094[a]	.214[a]	1							
$lnUE_{r}\_1$	-.339[a]	-.119[a]	.457[a]	.380[a]	1						
$lnLE_{jr}\_2$	-.135[a]	-.004	.232[a]	.854[a]	.384[a]	1					
$lnUE_{r}\_2$	-.211[a]	-.074[a]	.294[a]	.112[a]	.798[a]	.097[a]	1				
$lnLE_{jr}\_3$	-.167[a]	-.042[a]	.225[a]	.706[a]	.336[a]	.839[a]	.059[a]	1			
$lnUE_{r}\_3$	-.099[a]	-.036[a]	.149[a]	-.090[a]	.479[a]	-.112[a]	.824[a]	-.156[a]	1		
$lnLE_{jr}\_4$	.017[a]	.110[a]	.189[a]	.872[a]	.359[a]	.715[a]	.154[a]	.574[a]	-.061[a]	1	
$lnUE_{r}\_4$	-.347[a]	-.117[a]	.476[a]	.418[a]	.907[a]	.421[a]	.635[a]	.374[a]	.338[a]	.449[a]	1
$lnLE_{jr}\_5$	-.109[a]	.007	.202[a]	.749[a]	.364[a]	.870[a]	.135[a]	.715[a]	-.085[a]	.846[a]	.448[a]
$lnUE_{r}\_5$	-.292[a]	-.099[a]	.406[a]	.316[a]	.841[a]	.311[a]	.731[a]	.267[a]	.508[a]	.344[a]	.918[a]
$lnLE_{jr}\_6$	-.145[a]	-.034[a]	.196[a]	.607[a]	.321[a]	.720[a]	.097[a]	.883[a]	-.130[a]	.679[a]	.394[a]
$lnUE_{r}\_6$	-.168[a]	-.059[a]	.243[a]	.089[a]	.576[a]	.079[a]	.634[a]	.037[a]	.688[a]	.063[a]	.611[a]
$COMP_{jr}$	.104[a]	.156[a]	.117[a]	.593[a]	.207[a]	.470[a]	.051[a]	.377[a]	-.092[a]	.515[a]	.216[a]
EXP_{ijr}	-.294[a]	-.053[a]	.480[a]	.057[a]	.258[a]	.054[a]	.214[a]	.040[a]	.158[a]	.017[a]	.239[a]
IMP_{ijr}	-.300[a]	-.077[a]	.449[a]	.051[a]	.298[a]	.042[a]	.239[a]	.024[a]	.164[a]	.016[a]	.286[a]
FDI_{ijr}	-.219[a]	-.071[a]	.311[a]	-.009[b]	.208[a]	-.009[b]	.201[a]	-.010[b]	.158[a]	-.040[a]	.172[a]
$INDEP_{ijr}$	.223[a]	.037[a]	-.372[a]	-.027[a]	-.172[a]	-.020[a]	-.145[a]	.004	-.119[a]	-.019[a]	-.168[a]
RND_{ijr}	.208[a]	.250[a]	-.021[a]	.032[a]	-.028[a]	-0.002	-.017[a]	-.015[a]	-.008[b]	.034[a]	-.026[a]
# Obs.	65,027										

Table 3: Correlation Matrix of Variables used in the Regression Analysis (cont.)

	$lnLE_{jr}$_4	$lnUE_{r}$_4	$lnLE_{jr}$_5	$lnLE_{jr}$_5	$lnLE_{jr}$_6	$lnUE_{r}$_6	$COMP_{jr}$	EXP_{ijr}	IMP_{ijr}	FDI_{ijr}	$INDEP_{ijr}$	RND_{ijr}
$lnLE_{jr}$_4	1											
$lnUE_{r}$_4	.449[a]	1										
$lnLE_{jr}$_5	.846[a]	.448[a]	1									
$lnUE_{r}$_5	.344[a]	.918[a]	.334[a]	1								
$lnLE_{jr}$_6	.679[a]	.394[a]	.813[a]	.282[a]	1							
$lnUE_{r}$_6	.063[a]	.611[a]	.052[a]	.814[a]	.007	1						
$COMP_{jr}$	.515[a]	.216[a]	.397[a]	.162[a]	.315[a]	.020[a]	1					
EXP_{ijr}	.017[a]	.239[a]	.012[a]	.238[a]	.003	.202[a]	.039[a]	1				
IMP_{ijr}	.016[a]	.286[a]	.005	.276[a]	-.009[b]	.217[a]	.030[a]	.561[a]	1			
FDI_{ijr}	-.040[a]	.172[a]	-.042[a]	.174[a]	-.036[a]	.153[a]	0.00	.411[a]	.401[a]	1		
$INDEP_{ijr}$	-.019[a]	-.168[a]	-.011[a]	-.164[a]	.014[a]	-.140[a]	-.018[a]	-.308[a]	-.265[a]	-.160[a]	1	
RND_{ijr}	.034[a]	-.026[a]	0.00	-.022[a]	-.013[a]	-.012[a]	.050[a]	-.015[a]	-.016[a]	-.017[a]	.007	1
# Obs.	65,027											

Note: [a] and [b] denote a statistical significance at 1% and 5% (2-tailed), respectively.

Source: Author's calculation

5.2 Regression Results

The results of 2SLS regression are reported in Table 4. As mentioned before, I measure the effects of localization and urbanization economies at six entities (see Figure 2.1). Thus, six panels in Table 4 report the results of 2SLS regression with respect to each spatial and sectoral entity in which localization and urbanization economies are measured.[16] In each panel, four model specifications (denoted by (1), (2), (3), and (4)) are reported: the first specification excludes both regional and industry dummies; the second specification includes only region dummies; the third specification includes only industry dummies; and the last specification includes both region and industry dummies. The inclusion and exclusion of region and industry dummies are denoted by "Yes" and "No", respectively.

The first point to note concerns the global fit of the models. As can be seen, our model specifications explain the variations of dependent variable quite well: the values of R^2 for all of our model specifications range between 0.720 and 0.745.

Variables that capture localization economies (LE_{jr}) exhibit interesting patterns. Localization economies tend to have positive effects on establishments' labor productivity at a broader range of industrial aggregation, and negative effects at a narrower range of aggregations. For instance, in Panel a, where localization economies are measured at provincial and 2-digit industrial levels, variable $lnLE_{jr}\_1$ is positive and significant (although the inclusion of industry dummies somehow changes its level of significance). However, once we move to more disaggregate levels of industry (i.e. to 3- and 4-digit levels in Panel b and Panel c, respectively), the effects of $lnLE_{jr}\_2$ and $lnLE_{jr}\_3$ become negative. Similarly, localization economies measured at the subregional level is also positive only for 2-digit industrial agglomeration while being negative for 3- and 4-digit agglomeration, as evident in the positive coefficients of $lnLE_{jr}\_4$ in Panel d, and negative coefficients of $lnLE_{jr}\_5$ and *lnLEjr_5* in Panel e, and Panel f, respectively. These results show that spatial agglomeration of manufacturing establishments in the same 2-digit industry would result in an increase in establishments' labor productivity, whereas spatial agglomeration of establishments in the same 3- or 4-digit industry is likely to reduce

[16] Each panel in Table 4 reports the results as follows: (1) Panel a reports the results for model specifications in which localization and urbanization economies are measured at 2-digit provincial industry; (2) Panel b reports the results for model specifications in which localization and urbanization economies are measured at 3-digit provincial industry; (3) Panel c reports the results for model specifications in which localization and urbanization economies are measured at 4-digit provincial industry; (4) Panel d reports the results for model specifications in which localization and urbanization economies are measured at 2-digit subregional industry; (5) Panel e reports the results for model specifications in which localization and urbanization economies are measured at 3-digit subregional industry; and (6) Panel f reports the results for model specifications in which localization and urbanization economies are measured at 4-digit subregional industry.

productivity. In other words, the agglomeration of a broader-range industry is more helpful in improving labor productivity of manufacturing establishments than the agglomeration of a narrow-range industry.[17] Thus, sectoral scope of industrial agglomeration matters for production efficiency.

Although we find some evidence to support Marshall-typed industrial agglomeration here, it can be argued that own-industry agglomeration does not hold in all cases. As shown, agglomeration has a positive effect on labor productivity only when it is measured at a 2-digit industrial level, but has negative effects when it is measured at 3- and 4-digit levels. Thus, it is possible that when sectorally related establishments are agglomerated, this increases the degree of competition for inputs (as they rely on similar inputs) or competition in their final product markets (as they produce similar products) (Lall et al., 2004). This is likely to be the case for the agglomeration of narrow-range production activities (i.e., 3-digit or 4-digit industries) which require more specific inputs and compete in a specific line of products.

The effects of urbanization economies are, in general, negative and significant, which indicate that diversified industrial structure is not good for productivity improvement in any level of spatial agglomeration (i.e. province or subregion). In other words, the increase in the industrial diversity of the province (or subregion) decreases the labor productivity of manufacturing establishments located in that province (or subregion). As evidenced in every model specification, the effects of $lnUE_r$ variables are negative and highly significant.[18]

Negative effects of urbanization can be expected if an increased agglomeration results in higher congestion costs that outweigh agglomeration benefits (Carlino, 1979; Lall et al. 2004; Baldwin et al., 2008). However, a problem arises in that this study's measure of urbanization economies only captures regional industrial diversity, without directly capturing industrial density or congestion.[19] To determine whether negative effects of industrial diversity take place because of over-agglomeration or high-congestion costs, I divided the sample into four groups based on provincial

[17] If the television industry is taken as an example, these results imply that labor productivity tends to increase when manufacturers of electronic valves, tubes and other electronic equipments are co-located (in the same province or in neighboring provinces) with manufacturers of transmitters, line telephony, line telegraphy, and television receivers, whereas the spatial agglomeration of electronic equipments manufacturers alone tends to decrease their labor productivity.

[18] An exception is in model 3 of Panel e. which $lnUE_r\_5$ is positive. However, as the level of statistical significance is very weak (*p-value* ≈ 0.095, which exceeds conventional level of 0.05), we can take this as an insignificant case.

[19] In fact, most of empirical studies that test Jacobs's ideas of urbanization economies rarely make a clear distinction between these two phenomena, implicitly assuming that diversity and density are two parallel phenomena of urbanization. Thus, although congestion costs associated with increased density can be considered as a negative side of urbanization, the impacts of diversity is still unclear (Fu and Hong, 2010).

industry density (i.e. highest-density, high-density, low-density, and lowest-density subsamples),[20] and then ran 2SLS regressions for each group to see how $lnUE_r$ variables behave. The results are provided in Appendix 2. It is evident that negative effects of industrial diversity are predominant in the highest-density subsample[21]: the coefficients of $lnUE_r$ variables are negative in every model specification, and five of six specifications are statistically significant. However, when other subsamples are examined, the results differ. In the high-density subsample, the coefficients of $lnUE_r$ are all positive and significant in five of six model specifications. In two other subsamples, the picture is less clear: coefficients of $lnUE_r$ are negative, positive or statistically insignificant at different levels of spatial and sectoral aggregations. Based on these results, therefore, it can be argued that negative effects of urbanization economies are partially explained by increased congestion costs.[21]

Most control variables behave as expected. EXP_{ijr} are positive in every model specification, despite some variation in its statistical significance levels. Thus, manufacturing establishments that export their products are more likely to have higher labor productivity than those that do not.

Compared to branches or affiliated companies, independent establishments tend to be less productive, as evident in the negative coefficients of the $INDEP_{ijr}$ variable in every model specification where it is statistically significant. This evidence supports the notion that establishments embedded in multi-establishment firm structure tend to benefit from technological spillovers within the intra-firm network (Henderson, 2003).

---

[20] Provincial industry density is defined as the number of provincial manufacturing establishments divided by provincial area size. Each subsample consists of 19 provinces.

[21] The highest-density subsample mainly consists of manufacturing establishments in BMR area.

[22] The results of this study are similar to those of Fu and Hong (2010) which find that negative effects of industrial diversity on firms' productivity exist only in cities with a population size of larger than 500 thousand.

Table 4: 2SLS Regression Results

	Panel a.				Panel b.			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Intercept	2.47(0.09) [a]	1.73(0.13) [a]	2.82(0.09) [a]	2.57(0.13) [a]	2.88(0.13) [a]	2.75(0.18) [a]	2.94(0.13) [a]	3.77(0.18) [a]
lnK_{ijr}	0.78(0.00) [a]	0.78(0.00) [a]	0.76(0.00) [a]	0.76(0.00) [a]	0.78(0.00) [a]	0.78(0.00) [a]	0.75(0.00) [a]	0.75(0.00) [a]
$ln\ L_{ijr}$	-1.05(0.01) [a]	-1.04(0.01) [a]	-1.05(0.01) [a]	-1.04(0.01) [a]	-1.04(0.01) [a]	-1.03(0.01) [a]	-1.04(0.01) [a]	-1.03(0.01) [a]
$lnLE_{jr}\_1$	0.09(0.01) [a]	0.13(0.01) [a]	0.02(0.02)	0.01(0.02)				
$lnUE_{r}\_1$	-0.55(0.04) [a]	-0.38(0.05) [a]	-0.23(0.04) [a]	-0.22(0.05) [a]				
$lnLE_{jr}\_2$					-0.15(0.01) [a]	-0.14(0.01) [a]	-0.18(0.01) [a]	-0.21(0.01) [a]
$lnUE_{r}\_2$					-0.30(0.04) [a]	-0.31(0.05) [a]	-0.02(0.04)	-0.25(0.05) [a]
$COMP_{jr}$	0.26(0.01) [a]	0.26(0.01) [a]	0.29(0.01) [a]	0.29(0.01) [a]	0.39(0.01) [a]	0.40(0.01) [a]	0.37(0.01) [a]	0.39(0.01) [a]
EXP_{ijr}	0.10(0.03) [a]	0.05(0.03)	0.07(0.03) [b]	0.04(0.03)	0.09(0.03) [a]	0.05(0.03)	0.06(0.03) [c]	0.04(0.03)
IMP_{ijr}	-0.07(0.03) [a]	-0.07(0.03) [b]	-0.05(0.03)	-0.05(0.03) [c]	-0.14(0.03) [a]	-0.14(0.03) [a]	-0.08(0.03) [a]	-0.10(0.03) [a]
FDI_{ijr}	0.02(0.04)	-0.02(0.04)	-0.02(0.04)	-0.04(0.04)	-0.07(0.04)	-0.09(0.04) [b]	-0.08(0.04) [b]	-0.07(0.04) [c]
$INDEP_{ijr}$	-0.10(0.03) [a]	-0.08(0.03) [a]	-0.04(0.03)	-0.03(0.03)	-0.05(0.03) [c]	-0.03(0.03)	0.01(0.03)	0.02(0.03)
RND_{ijr}	0.99(0.03) [a]	0.98(0.03) [a]	0.98(0.03) [a]	0.97(0.03) [a]	0.98(0.03) [a]	0.98(0.03) [a]	0.97(0.03) [a]	0.96(0.03) [a]
REG_{r}	No	Yes	No	Yes	No	Yes	No	Yes
IND_{j}	No	No	Yes	Yes	No	No	Yes	Yes
R^2	0.720	0.729	0.740	0.740	0.724	0.730	0.735	0.745
F-Stat.	16579 [a]	11946 [a]	13138 [a]	10105 [a]	16755 [a]	12031 [a]	13291 [a]	10201 [a]
Obs (No.)	65,027	65,027	65,027	65,027	65,027	65,027	65,027	65,027

Note: (1) [a], [b], and [c] denote a statistical significance at 1%, 5%, and 10% levels, respectively; (2) The numbers in parentheses are Heteroscedasticity-robust standard errors.

Source: Author's estimation

Table 4: 2SLS Regression Results (cont.)

	Panel c.				Panel d.			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Intercept	2.54(0.13)[a]	2.42(0.18)[a]	2.75(0.13)[a]	3.30(0.18)[a]	2.24(0.08)[a]	0.58(0.16)[a]	2.67(0.08)[a]	1.78(0.17)[a]
lnK_{ijr}	0.78(0.00)[a]	0.78(0.00)[a]	0.75(0.00)[a]	0.75(0.00)[a]	0.78(0.00)[a]	0.77(0.00)[a]	0.76(0.00)[a]	0.76(0.00)[a]
$ln\ L_{ijr}$	-1.04(0.01)[a]	-1.03(0.01)[a]	-1.04(0.01)[a]	-1.03(0.01)[a]	-1.05(0.01)[a]	-1.03(0.01)[a]	-1.05(0.01)[a]	-1.04(0.01)[a]
$lnLE_{jr}\_3$	-0.16(0.01)[a]	-0.14(0.01)[a]	-0.15(0.01)[a]	-0.17(0.01)[a]				
$lnUE_{r}\_3$	-0.15(0.03)[a]	-0.17(0.04)[a]	-0.01(0.03)	-0.16(0.04)[a]				
$lnLE_{jr}\_4$					0.20(0.01)[a]	0.26(0.01)[a]	0.08(0.01)[a]	0.14(0.01)[a]
$lnUE_{r}\_4$					-0.73(0.03)[a]	-0.28(0.05)[a]	-0.38(0.03)[a]	-0.22(0.05)[a]
$COMP_{jr}$	0.37(0.01)[a]	0.38(0.01)[a]	0.34(0.01)[a]	0.35(0.01)[a]	0.22(0.01)[a]	0.21(0.01)[a]	0.24(0.01)[a]	0.24(0.01)[a]
EXP_{ijr}	0.09(0.03)[a]	0.05(0.03)	0.05(0.03)	0.04(0.03)	0.11(0.03)[a]	0.07(0.03)[b]	0.08(0.03)[b]	0.05(0.03)
IMP_{ijr}	-0.17(0.03)[a]	-0.16(0.03)[a]	-0.10(0.03)[a]	-0.11(0.03)[a]	-0.03(0.03)	-0.03(0.03)	-0.03(0.03)	-0.03(0.03)
FDI_{ijr}	-0.08(0.04)[c]	-0.10(0.04)[b]	-0.07(0.04)[c]	-0.07(0.04)[c]	0.04(0.04)	0.00(0.04)	0.00(0.04)	-0.03(0.04)
$INDEP_{ijr}$	-0.03(0.03)	-0.02(0.03)	0.02(0.03)	0.03(0.03)	-0.10(0.03)	-0.09(0.03)[a]	-0.05(0.03)[c]	-0.05(0.03)
RND_{ijr}	0.98(0.04)[a]	0.97(0.03)[a]	0.97(0.03)[a]	0.96(0.03)[a]	0.99(0.03)[a]	0.98(0.03)[a]	0.98(0.03)[a]	0.97(0.03)[a]
REG_{r}	No	Yes	No	Yes	No	Yes	No	Yes
IND_{j}	No	No	Yes	Yes	No	No	Yes	Yes
R^2	0.725	0.730	0.735	0.735	0.720	0.731	0.734	0.735
F-Stat.	16797[a]	12051[a]	13275[a]	10185[a]	16668[a]	12042[a]	13171[a]	10154[a]
Obs.(No.)	65,027	65,027	65,027	65,027	65,027	65,027	65,027	65,027

Note: (1) [a], [b], and [c] denote a statistical significance at 1%, 5%, and 10% levels, respectively; (2) The numbers in parentheses are Heteroscedasticity-robust standard errors.

Source: Author's estimation

Table 4: 2SLS Regression Results (cont.)

	Panel e.				Panel f.			
	(1)	**(2)**	**(3)**	**(4)**	**(1)**	**(2)**	**(3)**	**(4)**
Intercept	2.93(0.13) [a]	2.76(0.22) [a]	2.88(0.12) [a]	4.32(0.22) [a]	2.72(0.14) [a]	2.88(0.22) [a]	2.88(0.13) [a]	3.80(0.22) [a]
lnK_{ijr}	0.78(0.00) [a]	0.78(0.00) [a]	0.76(0.00) [a]	0.75(0.00) [a]	0.78(0.00) [a]	0.78(0.00) [a]	0.76(0.00) [a]	0.76(0.00) [a]
$ln\ L_{ijr}$	-1.05(0.01) [a]	-1.03(0.01) [a]	-1.04(0.01) [a]	-1.04(0.01) [a]	-1.05(0.01) [a]	-1.03(0.01) [a]	-1.05(0.01) [a]	-1.04(0.01) [a]
$lnLE_{jr}\_5$	-0.08(0.01) [a]	-0.06(0.01) [a]	-0.16(0.01) [a]	-0.20(0.01) [a]				
$lnUE_{r}\_5$	-0.34(0.04) [a]	-0.36(0.06) [a]	0.08(0.05) [c]	-0.32(0.06) [a]				
$lnLE_{jr}\_6$					-0.11(0.01) [a]	-0.07(0.01) [a]	-0.11(0.01) [a]	-0.12(0.01) [a]
$lnUE_{r}\_6$					-0.20(0.04) [a]	-0.34(0.05) [a]	-0.03(0.04)	-0.29(0.05) [a]
$COMP_{jr}$	0.35(0.01) [a]	0.35(0.01) [a]	0.34(0.01) [a]	0.35(0.01) [a]	0.34(0.01) [a]	0.35(0.01) [a]	0.31(0.01) [a]	0.31(0.01) [a]
EXP_{ijr}	0.08(0.03) [b]	0.05(0.03) [a]	0.04(0.03)	0.03(0.03)	0.08(0.03) [b]	0.06(0.03) [c]	0.05(0.03)	0.04(0.03)
IMP_{ijr}	-0.13(0.03) [a]	-0.12(0.03) [a]	-0.10(0.03) [a]	-0.11(0.03) [a]	-0.15(0.03) [a]	-0.13(0.03) [a]	-0.09(0.03) [a]	-0.10(0.03) [a]
FDI_{ijr}	-0.08(0.04) [b]	-0.09(0.04) [b]	-0.09(0.04) [b]	-0.08(0.04) [b]	-0.08(0.04) [c]	-0.09(0.04) [b]	-0.07(0.04)	-0.06(0.04)
$INDEP_{ijr}$	-0.07(0.03) [b]	-0.05(0.03) [c]	0.00(0.03)	0.02(0.03)	-0.05(0.03) [c]	-0.05(0.03)	0.00(0.03)	0.00(0.03)
RND_{ijr}	0.98(0.03) [a]	0.98(0.03) [a]	0.97(0.03) [a]	0.96(0.03) [a]	0.98(0.03) [a]	0.98(0.03) [a]	0.97(0.03) [a]	0.97(0.03) [a]
REG_{r}	No	Yes	No	Yes	No	Yes	No	Yes
IND_{j}	No	No	Yes	Yes	No	No	Yes	Yes
R^2	0.724	0.729	0.735	0.735	0.724	0.729	0.734	0.735
F-Stat.	16698 [a]	11981 [a]	13247 [a]	10164 [a]	16700 [a]	11985 [a]	13202 [a]	10135 [a]
Obs.(No.)	65,027	65,027	65,027	65,027	65,027	65,027	65,027	65,027

Note: (1) [a], [b], and [c] denote a statistical significance at 1%, 5%, and 10% levels, respectively; (2) The numbers in parentheses are Heteroscedasticity-robust standard errors.

Source: Author's estimation

RND_{ijr} is the most consistent and most robust variable in this analysis. Its coefficients vary between 0.96 and 0.99 in our model specifications and are highly significant at a 1% level in every model. This confirms the importance of establishments' R&D investment in generating and enhancing technological capability (Cohen and Levinthal, 1989).

$COMP_{jr}$ is another consistent and robust variable in this analysis. Its coefficients are always positive and vary between 0.21 and 0.40 in the model specifications. It is also statistically significant at a 1% level in every model. The positive coefficient of $COMP_{jr}$ can be interpreted that the more market share is equally distributed among establishments (i.e., no single establishment dominates the market), the higher the establishment's labor productivity will be. These results support Porter's (1990, 1998) argument that localized competition, rather than monopolistic local industrial structure, is a key factor to increasing growth and competitiveness of local industry.

Two variables –FDI_{ijr} and IMP_{ijr} – have coefficients that run counter to our expectation. The coefficients of the FDI_{ijr} are significantly negative (or positive with no statistical significance) which indicate that labor productivity of manufacturing establishments having foreign investment is lower (or not necessarily higher) than that of Thai-owned establishments. Though this result is not consistent with general expectations, it is not unfathomable in the case of Thailand. Previous studies that compare the labor productivity of foreign-owned and Thai-owned establishments (Ramstetter, 1994; Ramstetter, 2001) find little evidence to suggest that the former has higher labor productivity. The comparative analysis of labor productivity between foreign and Thai establishments by Ramstetter (2001) using NSO's industrial census 1996 and industrial survey 1998 find that productivity differentials are not observed as expected. Moreover, the econometric analysis in his study finds no evidence to suggest that foreign ownership will result in higher productivity. According to him, the variations in labor productivity differentials are more dependent on other factors such as industrial characteristics and scale economies (Ramstetter, 2001).[23] One possible interpretation is that Thai-owned establishments have been able to improve their production efficiency to the same level as (or higher level than) establishments which have foreign investment share. Of course, more investigation is needed to elaborate on this issue.

[23] To investigate the point made by Ramstetter (2001), I divided the sample of manufacturing establishments into four groups based on industrial categories (*resource-based*, *labor-intensive*, *machinery*, and *metal, chemical, and paper*) and re-estimated Model 7 to see the effects of FDI_{ijr} on establishments' labor productivity with respect to each industrial category (due to limited space, the results are not produced here). Positive coefficients are observed for *labor-intensive* and *machinery* industries but without statistical significance. For *resource-based* and *metal, chemical, and paper* industries, the coefficients are negative and statistically significant. These results are generally similar to those of Ramstetter (2001) (see Tippakoon, 2011).

The coefficients of IMP_{ijr} are negative in every model specification; and despite some variation in its statistical significance levels across the specifications, most of them are significant at the conventional 5% level. This indicates that importers of foreign intermediate inputs or products tend to be less productive than non-importers. This result may be in line with Augier et al.'s (2009) argument that the most efficient learning of new technologies embodied in imported products takes place when the importers have sufficient absorptive capabilities, which require a significant investment in human capital. Without such capabilities, the import of technologies may not result in enhancing establishments' productivity.

Before concluding this paper, the effects of localization and urbanization economies taking place at different spatial and sectoral settings are summarized based on the results in Table 4 (see Figure 2). At the provincial level, the agglomeration of sectorally related establishments from 2-digit industry yields a positive effect on manufacturing establishments' labor productivity. However, as we move to agglomerations at 3- and 4-digit levels, the effects become negative. This pattern is also exhibited when the subregion is used as a spatial unit of industrial agglomeration. For urbanization economies, it is found that their effects, measured in terms of agglomeration of establishments from various industries, are negative in any setting. Thus, depending on sectoral scope of agglomeration, localization economies, rather than urbanization economies, matter for establishments' productivity improvement.

Figure 2: Summary of Localization and Urbanization Effects

	Industrial unit		
Regional unit	**2-digit**	**3-digit**	**4-digit**
Province	$lnLE_{jr}\_1$: positive $lnUE_{r}\_1$: negative	$lnLE_{jr}\_2$: negative $lnUE_{r}\_2$: negative	$lnLE_{jr}\_3$: negative $lnUE_{r}\_3$: negative
Subregion	$lnLE_{jr}\_4$: positive $lnUE_{r}\_4$: negative	$lnLE_{jr}\_5$: negative $lnUE_{r}\_5$: negative	$lnLE_{jr}\_6$: negative $lnUE_{r}\_6$: negative

Note: Using information from Table 4 (panel a. to panel f.): (1) positive = coefficient(s) is/are positive and statistically significant at 5% level or smaller; and (2) negative = coefficient(s) is/are negative and statistically significant at 5% level or smaller.

Source: Author

6. Conclusion

This paper examined whether industrial agglomeration helps manufacturing establishments improve their labor productivity. To answer this question, I specified a production function which assumes that labor productivity is influenced by both establishment-specific as well as structural factors. To measure the productivity effects of industrial agglomeration, the effects of agglomeration that arise from localization economies (i.e. *spatial agglomeration of establishments operating in the same sector*) were separated from those generated by urbanization economies (i.e. *spatial agglomeration of establishments from different sectors*). It was also assumed that the effects of industrial agglomeration could vary with spatial and sectoral scopes of agglomeration. For empirical investigation, I applied 2SLS regression to analyze establishment-level data from the Thai manufacturing industrial census 2007.

The results from 2SLS regression analysis revealed that localization economies do help improve establishments' labor productivity. However, it is found that positive effects of localization take place only for a spatial agglomeration of sectorally related establishments at the 2-digit industrial level. For spatial agglomeration at 3-digit and 4-digit levels, localization effects are negative. These results indicate that industrial agglomeration of manufacturing establishments operating in a broader range of production activities helps increase productivity. On the other hand, the agglomeration of establishments operating in a narrow range of activities tends to decrease productivity. As for the effects of urbanization economies, these are found to be negative in any spatial and sectoral settings. As I defined urbanization economies in terms of regional industrial diversity, negative coefficients of urbanization variables indicate that diversified industrial structure is not good for establishments' labor productivity. Further investigation on urbanization economies has revealed that negative effects of industrial diversity are more likely to be attributed to the congestion costs arisen when agglomeration expands further.

Thus, responding to the main research questions, we can conclude that industrial agglomeration helps improve manufacturing establishments' labor productivity. However, the form of agglomeration matters. Localization economies are more conducive to such productivity improvement than urbanization economies. This happens when a broader-range and complementary activities are spatially agglomerated.

References

Anastassova, L. (2006). *Productivity Differences and Agglomeration across Districts of Great Britain*. CEREI-GI Working Paper No. 289, Center for Economic Research and Graduate Education, Charles University, the Czech Republic.

Anderson, M. and Loof, H. (2009). *Agglomeration and Productivity: Evidence from Firm-level Data*. CESIS Working Paper No. 170, Centre of Excellence for Science and Innovation Studies (CESIS).

Arrow, K.J. (1962). The economic implication of learning by doing. *The Review of economic studies* 29, pp. 155-173.

Aw, B.Y. and Hwang, A.R. (1995). Productivity and the export market: a firm-level analysis. *Journal of Development Economic*s 47, pp. 313-332.

Becchetti, L., Panizza de, A., and Oropallo, F. (2007). Role of industrial district externalities in export and value-added performance: evidence from the population of Italian firms. *Regional Studies* 41(5), pp. 601-621.

Baldwin, J.R., Beckstead, D., Brown, M.W., and Rigby, D.L. (2008). Agglomeration and the geography of localization economies in Canada. *Regional Studies* 41(1), pp. 117-132.

Brown, W.M. and Rigby, D.L. (2009). *Marshallian Localization Economies: Where Do They Come From and To Whom Do they Flow?* CCPR Working Paper No.020, California Center for Population Research.

Cainelli, G. (2008). Spatial agglomeration, technological innovations, and firm productivity: evidence from Italian industrial districts. *Growth and Change* 39(3), pp. 414-435.

Capello, R. (1999). SME Clustering and factor productivity: a milieu production function model. *European Planning Studies* 7(6), pp. 719-735.

Caves, R. (2007). *Multinational Enterprises and Economic Analysis*. Cambridge University Press.

Ciccone, A. (2002). Agglomeration effects in Europe. *European Economic Review* 46, pp. 213-227.

Ciccone, A. and Hall, R.E. (1996). Productivity and the density of economic activity. *American Economic Review* 86, pp. 54-70.

Cingano, F. and Schivardi, F. (2004). Identifying the sources of local productivity growth. *Journal of European Economic Association* 2, pp. 720-742.

Cohen, W.M. and Lavinthal, D.A. (1989). Innovation and learning: the two faces of R & D. *The Economic Journal* 99(397), pp. 569-596.

Combes, P.-P. (2000). Economic structure and local growth: France, 1984-1993. *Journal of Urban Economics* 47(3), 329-355.

Combes, P.-P., Mayer, T., and Thisse, J.-F. (2008a). *Economic Geography: The Integration of Regions and Nations*. Princeton University Press.

Combes, P.-P., Duranton, G., and Gobillon, L. (2008b). Spatial wage disparities: sorting matters! *Journal of Urban Economics* 63(2), pp. 723-742.

Feldman, M.P. and Audretsch, D.B. (1999). Innovation in cities: science-based diversity, specialization and localized competition. *European Economic Review* 43, pp. 409-429.

Fryges, H. and Wagner, J. (2008). Exports and productivity growth: first evidence from a continuous treatment approach. *Reviews of World Economics* 144(4), pp. 695-722.

Fu, S. and Hong, J. (2010). Testing urbanization economies in manufacturing industries: urban diversity or urban size? *Journal of Regional Science* 51(3), pp. 585-603.

Gao, T. (2004). Regional industrial growth: evidence from Chinese industries. *Regional Science and Urban Economics* 34, pp.101-124.

Glaeser, E.L., Kallal, H.D., Scheinkman, J.A., and Shleifer, A. (1992). Growth in cities. *Journal of Political Economy* 100(6), pp. 1126-1152.

Glaeser, E.L. and Mare, D.C. (2001). Cities and Skills. *Journal of Labor Economics* 19(2), pp. 316-342.

Glaeser, E.L. and Resseger, M.G. (2009). *The Complementarity between Cities and Skills*. NBER Working Paper No. 15303, National Bureau of Economic Research.

Goldberg, P., Khandelwal, A., Pavcnik, N., and Topalova, K. (2008). *Imported Intermediate Inputs and Domestic Product Growth: Evidence from India*. NBER Working Paper No. 14416, National Bureau of Economic Research.

Greenaway, D. and Kneller, R. (2004). Exporting and productivity in the United Kingdom. *Oxford Review of Economic Policy* 20, pp. 358-371.

Grilliches, Z. (1986). Productivity, R and D, and basic research at the firm level in the 1970s. *The American Economic Review* 76(1), pp. 141-154.

Hall, A. (2004). Transport infrastructure, agglomeration economies, and firm birth: empirical evidence from Portugal. *Journal of Regional Science* 44(4), pp. 693-712.

Henderson, J.V., Kuncoro, A., and Turner, M. (1995). Industrial development in cities. *Journal of Political Economies* 103, pp. 1067-1085.

Henderson, J.V. (2003). Marshall's scale economies. *Journal of Urban Economics* 53, pp. 1-28.

Jacobs, J. (1969). *The Economy of Cities*. New York: Random House.

Jacob, J. (1984). *Cities and the Wealth of Nations*. New York: Random House.

Javorcik, B.S. (2004). Does foreign direct investment increase the productivity of domestic firms? In search of spillovers through backward linkages. *The American Economic Review* 94(3), pp. 605-627.

Kohpaiboon, A. (2009). *Vertical and Horizontal FDI Technology Spillovers: Evidence from Thai Manufacturing*. ERIA Discussion Paper No. 2009-08, Economic Research Institute for ASEAN and East Asia.

Koo, J. (2005). Agglomeration and spillovers in a simultaneous framework. *Annals of Regional Science* 39, pp. 35-47.

Lall, S.V., Shalizi, S., and Deichmann, U. (2004). Agglomeration economies and productivity in Indian industry. *Journal of Development Economics* 73, pp. 643-673.

Liu, M., Wang, W., Kuo, C. (2005). *Industrial cluster and Taiwanese IT firm's productivity: quantile regression*. Unpublished manuscript, Chung-Hua Institution for Economic Research.

Madsen, E.S., Smith, V., and Dilling-Hansen, M. (2003). *Industrial Clusters, Firm Location and Productivity: Some Empirical Evidence from Danish Firms*. Working Paper 03-26, Aarhus School of Business, University of Aarhus.

Mare and Timins (2006). *Geographic Concentration and Firm Productivity*. Motu Working Paper 06-08, Motu Economic and Public Policy Research.

Marshall, A. (1910). *Principle of Economics*. London: MacMillan.

Martin, P., Mayer, T., and Mayneris, F. (2008). *Spatial Concentration and Firm-level Productivity in France*. CEPR Working Paper No. 6858, Centre for Economic Policy Research.

Moomaw, R.L. (1981). Productivity and city size: a critique of the evidence. *Quarterly Journal of Economics* 96, pp. 675-688.

Moretti, E. (2004). Workers' education, spillovers, and productivity: evidence from plant-level production functions. *American Economic Review* 94, pp. 656-690.

Nakamura, R. (1985). Agglomeration economies in urban manufacturing industries: a case of Japanese cities. *Journal of Urban Economics* 17(1), pp. 108-124.

Panne van der, G. (2004). Agglomeration externalities: Marshall versus Jacobs. *Journal of Evolutionary Economics* 14, pp. 593-604.

Porter, M.E. (1990). *The Comparative Advantage of Nations*. New York: Free Press.

Rigby, D.L. and Essletzbichler, J. (2002). Agglomeration economies and productivity differences in US cities. *Journal of Economic Geography* 2, pp. 407-432.

Rodriguez-Clare, A. (1996). Multinationals, linkages and economic development. *The American Economic Review* 86(4), pp. 852-873.

Rice, P. and Venables, A.J. (2004). *Spatial Determinant of Productivity: Analysis of the Regions of Great Britain*. CEP Discussion Paper No. 642, London: Centre for Economic Performance.

Romer, P. M. (1986). Increasing returns and long-run growth. *The Journal of Political Economy* 98(5), pp. S71-S102.

Rosenthal, S.S. and Strange, W.C. (2001). The determinants of agglomeration. *Journal of Urban Economics* 50, pp. 191-229.

Rosenthal, S.S. and Strange, W.C. (2003). Geography, industrial organization, and agglomeration. *Review of Economics and Statistics* 85(2), pp. 377-393.

Rosenthal, S.S. and Strange, W.C. (2004). Evidence on the nature and sources of agglomeration economics, in Hendeson, J.V. and Thisse, J.F. (Ed.) *Handbook of Urban and Regional Economics – Volume 4*. Elsevier B.V, pp. 2119-2171.

Segal, D. (1976) Are there returns to scale in city size? *The Review of Economics and Statistics* 58(3), pp. 339-350.

Sveikauskas, L.A. (1975). The productivity of cities. *Quarterly Journal of Economics* 89(3), pp. 393-413.

Tabuchi, T. (1986). Urban agglomeration, capital augmenting technology, and labor market equilibrium. *Journal of Urban Economics* 20(2), pp. 211-228.

Tabuchi, T. and Yoshida, A. (1999). *Urban Agglomeration Economies in Consumption and Production*. CIRJE Discussion Paper No. 41, Center for International Research on Japanese Economy, University of Tokyo.

Tippakoon, P. (2011). *Clustering and Industrial Development: Evidence from Thailand*. Ph.D. Dissertation. Japan: Nagoya University.

Vogel, A. and Wagner, J. (2008). *Higher Productivity in Importing German Manufacturing Firms: Self-selection, Learning from Importing, or Both?* IZA Discussion Paper No. 3854, Bonn: Institute for the Study of Labor (IZA).

Wagner, J. (2007). Exports and productivity: a survey of the evidence from firm level data. *World Economy* 30, pp. 60-82.

Wheaton, W.C. and Lewis, M.J. (2002). Urban wages and labor market agglomeration. *Journal of Urban Economics* 51, pp. 542-562.

Wooldridge, J.M. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.

Wooldridge, J.M. (2006). *Introductory Econometrics: A Modern Approach*. Thomson South-Western.

Yokota, K. (2008). Spillovers and linkages between local and foreign plants, in Hiratsuka, D. and Kimura, F. (Ed.) *East Asia's Economic Integration: Progress and Benefits*, Palgrave McMillan, pp. 225-253.

Appendix 1: Hausman's Test Procedures for the Presence of Endogeneity

The *Hausman's test* can be performed in three steps as follows.

Step 1: The reduced form for each industrial agglomeration variable (which is suspected to be endogenous) is estimated by regressing each of them on all other variables in the structural model (including instrumental variables), and saving the residuals. Thus, each LE_{jr} and UE_r is regressed on other explanatory variables and their instrumental variables (LE_{jr} and UE_r with ten-year lag), then the residuals obtained from each regression are saved;

Step 2: The structural model (Equation 7 in the main text) is estimated with the residuals obtained from step 1 included;

Step3: The *F-test* is conducted for a joint statistical significance of residuals' coefficients based on the following procedure:

$$F = \left(\frac{RSS_r - RSS_u}{RSS_u}\right)\left(\frac{n-k-1}{m}\right)$$

where RSS_r = the sum of squared residuals from the restricted model and RSS_u = the sum of squared residuals from the unrestricted model; n = number of observations; k = number of parameters in the unrestricted model; and m = is the difference in degrees of freedom (df) between the restricted model (df_r) and unrestricted model (df_u) (i.e. m = $df_r - df_u$) (Wooldridge, 2006).

If the *F-statistic* is significant at a conventional 5% level, then the null hypothesis of no endogeneity is rejected. In other words, if the null hypothesis is rejected, LE_{jr} and UE_r variables are very much likely to be endogenous.

The results of *Hausman's test* procedures are shown in the following table. It is shown that, for all pairs of agglomeration variables, *F-statistics* are significant at the 1% level. Therefore, we say that our agglomeration variables are likely to be endogenous with the dependent variable.

	Endogenous Variables	**F-Statistics**	**Sig.**
Test 1	$lnLE_{jr}\_1$ and $lnUE_r\_1$	48.75	***
Test 2	$lnLE_{jr}\_2$ and $lnUE_r\_2$	24.63	***
Test 3	$lnLE_{jr}\_3$ and $lnUE_r\_3$	39.54	***
Test 4	$lnLE_{jr}\_4$ and $lnUE_r\_4$	32.15	***
Test 5	$lnLE_{jr}\_5$ and $lnUE_r\_5$	13.61	***
Test 6	$lnLE_{jr}\_6$ and $lnUE_r\_6$	31.28	***

Note: *** denotes 1% significance level.

Source: Author's calculation

Appendix 2: 2SLS Regressions Results for Provincial Density Subsamples

	(1) Highest-Density Group						(2) High-Density Group					
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Intercept	2.60(0.14)[a]	2.51(.20)[a]	2.59(.19)[a]	2.61(.13)[a]	2.97(.21)[a]	3.15(.25)[a]	1.18(.22)[a]	1.22(.35)[a]	1.24(.34)[a]	.28(.26)[a]	1.05(.34)[a]	1.08(.32)[a]
lnK_{ijr}	.78(.01)[a]	.77(.01)[a]	.77(.01)[a]	.77(.01)[a]	.78(.01)[a]	.77(.01)[a]	.76(.01)[a]	.76(.01)[a]	.76(.01)[a]	.76(.01)[a]	.76(.01)[a]	.76(.01)[a]
$ln\ L_{ijr}$	-1.03(.01)[a]	-1.03(.01)[a]	-1.03(.01)[a]	-1.03(.01)[a]	-1.03(.01)[a]	-1.03(.01)[a]	-1.07(.01)[a]	-1.06(.01)[a]	-1.06(.01)[a]	-1.07(.01)[a]	-1.06(.01)[a]	-1.06(.01)[a]
$lnLE_{jr}\_1$	-.12(.03)[a]						.12(.04)[a]					
$lnUE_{r}\_1$	-.21(.06)[a]						.04(.10)					
$lnLE_{jr}\_2$		-.24(.01)[a]						-.10(.02)[a]				
$lnUE_{r}\_2$		-.02(.07)						.26(.12)[b]				
$lnLE_{jr}\_3$			-.21(.01)[a]						-.06(.02)[a]			
$lnUE_{r}\_3$			-.10(.05)[b]						.16(.08)[b]			
$lnLE_{jr}\_4$				.09(.02)[a]						.18(.02)[a]		
$lnUE_{r}\_4$				-.59(.05)[a]						.30(.10)[a]		
$lnLE_{jr}\_5$					-.17(.02)[a]						-.06(.02)[a]	
$lnUE_{r}\_5$					-.13(.07)[c]						.30(.11)[a]	
$lnLE_{jr}\_6$						-.15(.01)[a]						-.03(.02)[c]
$lnUE_{r}\_6$						-.22(.07)[a]						.20(.08)[a]
$COMP_{jr}$	.33(.02)[a]	.41(.02)[a]	.37(.01)[a]	.21(.02)[a]	.32(.01)[a]	.30(.01)[a]	.37(.03)[a]	.50(.02)[a]	.48(.2)[a]	.35(.02)[a]	.49(.02)[a]	.47(.02)[a]
EXP_{ijr}	.08(.04)[c]	.07(.04)[c]	.07(.04)[c]	.08(.04)[b]	.05(.04)	.06(.04)	.15(.09)	.13(.09)	.13(.09)	.15(.09)	.12(.09)	.13(.09)
IMP_{ijr}	-.15(.04)[a]	-.20(.04)[a]	-.21(.04)[a]	-.10(.04)[a]	-.20(.04)[a]	-.20(.04)[a]	.21(.10)[b]	.15(.10)	.15(.10)	.23(.10)[b]	.15(.10)	.16(.10)[c]
FDI_{ijr}	-.06(.05)	-.12(.05)[a]	-.10(.05)[b]	.01(.05)	-.12(.05)[b]	-.10(.05)[c]	-.24(.15)	-.30(.15)[b]	-.27(.15)[c]	-.24(.15)	-.30(.15)[b]	-.27(.15)[c]
$SING_{ijr}$	-.05(.04)	-.01(.04)	.01(.04)	-.08(.04)[b]	-.03(.04)	-.02(.04)	-.11(.08)	-.10(.08)	-.08(.08)	-.11(.08)	-.10(.08)	-.08(.08)
RND_{ijr}	1.03(.06)[a]	1.02(.06)[a]	1.02(.06)[a]	1.04(.06)[a]	1.02(.06)[a]	1.02(.06)[a]	.95(.07)[a]	.94(.07)[a]	.94(.07)[a]	.95(.07)[a]	.94(.07)[a]	.94(.07)[a]
R^2	.707	.710	.710	.707	.708	.708	.714	.714	.714	.715	.714	.714
F-Stat.	6468[a]	6619[a]	6619[a]	6460[a]	6552[a]	6533[a]	3926[a]	3926[a]	6619[a]	3925[a]	3940[a]	3937[a]
Obs (No.)	27,805	27,805	27,805	27,805	27,805	27,805	16,250	16,250	16,250	16,250	16,250	16,250

Note: (1) [a], [b], and [c] denote a statistical significance at 1%, 5%, and 10% levels, respectively; (2) The numbers in parentheses are Heteroscedasticity-robust standard errors.

Source: Author's estimation

Appendix 2: 2SLS Regressions Results for Provincial Density Subsamples (cont.)

	(1) Low-Density Group						(2) Lowest-Density Group					
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Intercept	1.47(.25)^{a}	1.48(.33)^{a}	1.17(.35)^{a}	.08(.25)^{a}	1.31(.41)^{a}	1.19(.39)^{a}	.61(.31)^{b}	2.15(.49)^{a}	2.39(.43)^{a}	-.27(.30)	.96(.43)^{b}	.87(.40)^{b}
lnK_{ijr}	.81(.01)^{a}	.82(.01)^{a}	.82(.01)^{a}	.80(.01)^{a}	.82(.01)^{a}	.82(.01)^{a}	.75(.01)^{a}	.77(.01)^{a}	.77(.01)^{a}	.75(.01)^{a}	0.77(.01)^{a}	.77(.01)^{a}
$ln\ L_{ijr}$	-1.05(.01)^{a}	-1.06(.01)^{a}	-1.06(.01)^{a}	-1.06(.01)^{a}	-1.05(.01)^{a}	-1.05(.01)^{a}	-1.04(.01)^{a}	-1.02(.01)^{a}	-1.02(.01)^{a}	-1.02(.01)^{a}	-1.02(.01)^{a}	-1.02(.01)^{a}
$lnLE_{jr}\_1$	.19(.02)^{a}						.45(.05)^{a}					
$lnUE_{r}\_1$	-.34(.10)^{a}						-.21(.15)					
$lnLE_{jr}\_2$		.17(.03)^{a}						.03(.04)				
$lnUE_{r}\_2$		-.23(.10)^{b}						-.31(.27)				
$lnLE_{jr}\_3$			.11(.03)^{a}						.03(.04)			
$lnUE_{r}\_3$			.03(.08)						-.34(.12)^{a}			
$lnLE_{jr}\_4$				.40(.02)^{a}						.32(.03)^{a}		
$lnUE_{r}\_4$				.17(.09)^{c}						.20(.12)^{c}		
$lnLE_{jr}\_5$					.13(.02)^{a}						.04(.03)	
$lnUE_{r}\_5$					.20(.11)^{c}						.10(.03)^{a}	
$lnLE_{jr}\_6$						.05(.2)^{a}						.05(.03)^{b}
$lnUE_{r}\_6$						.01(.10)						.10(.09)
$COMP_{jr}$	.19(.02)^{a}	.24(.02)^{a}	.26(.02)^{a}	.11(.02)^{a}	.25(.02)^{a}	.27(.02)^{a}	.23(.04)^{a}	.49(.03)^{a}	.49(.03)^{a}	.34(.03)^{a}	.49(.03)^{a}	.50(.03)^{a}
EXP_{ijr}	-.03(.11)	-.05(.11)	-.06(.11)	-.02(.11)	-.05(.11)	-.06(.11)	.32(.13)^{b}	.30(.13)^{b}	.30(.13)^{b}	.31(.13)^{b}	.29(.13)^{b}	.28(.13)^{b}
IMP_{ijr}	.05(.12)	.06(.12)	.02(.12)	.13(.12)	.04(.12)	-.01(.12)	.03(.12)	-.04(.13)	-.03(.13)	.05(.13)	-.06(.13)	-.04(.13)
FDI_{ijr}	.48(.20)^{b}	.48(.20)^{b}	.42(.20)^{b}	.54(.20)^{a}	.45(.20)^{b}	.40(.20)^{b}	-.42(.20)^{b}	-.52(.20)^{a}	-.51(.20)^{a}	-.41(.20)^{b}	-.53(.20)^{b}	-.52(.20)^{a}
$SING_{ijr}$	-.19(.10)^{b}	-.20(.10)^{b}	-.19(.10)^{c}	-.17(.10)^{c}	-.17(.10)^{c}	-.16(.10)	.19(.11)^{c}	.22(.11)^{b}	.21(.11)^{b}	.25(.11)^{b}	.23(.11)^{b}	.22(.11)^{b}
RND_{ijr}	.95(.07)^{a}	.97(.07)^{a}	.98(.07)^{a}	.95(.07)^{a}	.97(.07)^{a}	.97(.07)^{a}	.93(.09)^{a}	.92(.09)^{a}	.91(.09)^{a}	.92(.09)^{a}	.92(.09)^{a}	.92(.09)^{a}
R^2	.698	.698	.697	.705	.697	.696	.694	.691	.692	.696	.691	.691
F-Stat.	2628^{a}	2617^{a}	2607^{a}	2705^{a}	2610^{a}	2601^{a}	1893^{a}	1874^{a}	1873^{a}	1903^{a}	1875^{a}	1875^{a}
Obs.(No.)	12,074	12,074	12,074	12,074	12,074	12,074	8,898	8,898	8,898	8,898	8,898	8,898

Note: (1) ^{a}, ^{b}, and ^{c} denote a statistical significance at 1%, 5%, and 10% levels, respectively; (2) The numbers in parentheses are Heteroscedasticity-robust standard errors.

Source: Author's estimation