Agglomeration Economies with Heterogeneous Firms in China*

Evidence from Firm-Level Data in 2007

Asei Ito

Abstract

This paper examines the relationship between agglomeration and firm-level performance in China. The author estimates firm-level production functions, total-factor productivity, and return on assets to compare the performance of heterogeneous firms within and outside the agglomeration. Agglomeration studies distinguish between localization—regions characterized by a particular economic activity—and urbanization, which corresponds to the development of overall economic activity. This paper investigates both types. Results suggest that negative localization economies exist in labor-intensive industries and positive urbanization economies are observed in capital-intensive industries. From the perspective of examining heterogeneous firms, negative localization effect is explained by the negative selection effect—fewer high-productivity firms are clustered in the apparel and fiber industries—and is explained by the negative agglomeration economy in the miscellaneous goods industry. The positive urbanization effect is explained by the agglomeration effect, i.e., the benefits of agglomeration serve all firms in the auto and communication equipment industries. Note that the productivity of Chinese firms in "Marshallian" localization clusters is not higher than that of firms located outside clusters. This finding challenges the theoretical prediction, prompting the need for further research.

Keywords

China; agglomeration economy; firm heterogeneity; TFP; ROA

* The author thanks Hiroyuki Kato, Shiro Hioki, Toshiyuki Matsuura, Masashi Hoshino, Yoshihiro Hashiguchi, Daisuke Fujii, Ryuhei Wakasugi, Masahiro Endoh, Ayako Obashi, Katsuji Nakagane, Kuanghui Chen, Tetsuya Komagata, Tomoo Marukawa, Ke Ding, Zhengping Fu, Yili Zhang, Yong Fang, Banri Ito, two anonymous referees, the conference and seminar participants at Kobe University, Keio University, East China University of Science and Technology, and the University of Tokyo. The author gratefully acknowledges the data set provided by the Grant-in-Aid for Scientific Research (A) "Empirical Analysis on Rural Urbanization in China" (Project code: 20252007, Term: 2008-2011, Chief: Hiroyuki Kato) and funding from the Konosuke Matsushita Memorial Foundation.

I Introduction

Since mainland China became "the workshop of the world," a growing number of scholars have investigated the geographical concentration of the economic activities of China. Despite some scholars having unintentionally discovered "Marshallian" industrial districts in China's coastal regions in the late 1980s,[1] recent studies have made more theoretical and detailed empirical contributions from a global perspective.[2] The background fact is the well-known concentration of production in China, as shown in Table 1. In 2008, 30% to 90% of the manufactured goods in China were produced in the top three of 31 provincial districts in the mainland. Most recent studies emphasize the upgrading of clusters and their implications for policymakers by drawing upon research into China's 30-year history of industrial clustering. Liu, ed. (2008), for example, covers a range of topics concerning regional policies and upgrading within clusters.

Table 1

Recent studies on geographical and spatial economics have developed sophisticated quantitative approaches toward agglomeration economies.[3] At least two important trends have emerged in literature: the development of indexes for measuring agglomeration, such as that by Duranton and Overman (2005), and the development of a theoretical background with heterogeneous firms, such as that by Baldwin and Okubo (2006) and Combes et al. (2009). However, only a few current papers concerning clustering in China have adopted these approaches, and almost none have focused on firm-level performance within agglomerations. Questions concerning industrial agglomeration in China remain wide open.

Using a simple method, this study clarifies the positive and negative contributions of agglomerations toward firm-level performance. The author estimates firm-level production functions, total-factor productivity (TFP), and return on assets (ROA), and compares them for industries within and outside an agglomeration area. This paper does not use sophisticated measurements or methods, such as those pioneered by Duranton and Overman (2005); what the literature on the agglomeration in China lacks is not the means to identify or measure agglomerations but to evaluate firm-level performance within them. Again, much remains to be discovered about productivity and profitability within China's agglomerations.

The literature on agglomeration economics distinguishes two types of agglomeration. The first is

[1] For example, sociologist Xiaotong Fei conducted field work in Wenzhou, Southern Jiangsu Province, and in the Pearl River Delta in the 1980s.
[2] See Sonobe and Otsuka (2006), Ding (2007), Ruan and Zhang (2009), and Marukawa (2010).
[3] See Combes et al.. (2008), and Glaeser, ed. (2010).

specialized agglomeration, the so-called "Marshallian" industrial district, or the area noted for an abundance of a specific industry (say "cluster"). The second is urbanization, the so-called Jacob's agglomeration, pertaining to areas having a significant number and variety of industries (say "city", see Cohen and Paul (2009)).[4] This paper focuses on both types.

Results of this paper suggest that firms operating within "Marshallian" clusters, especially in labor-intensive industries, are less productive and less profitable. On the other hand, firms in capital-intensive industries enjoy positive productivity benefits from urbanization economies. The finding of low productivity within "Marshallian" clusters defies the theoretical prediction.

The paper is organized as follows. Section 2 briefly reviews quantitative approaches to clustering and emphasizes the advantages of micro data. Section 3 details the estimation strategy, data, and process of generating variables. Section 4 reports a sequence of results. Section 5 discusses interpretations of those results. Section 6 concludes the paper.

Ⅱ Related Literature

Ⅱ-1 Basic Model of Estimating Agglomeration Economy

Cohen and Paul (2009) present a simple model to capture the agglomeration economy by using the production function:

$$Y_i = g(A_i)f(\mathbf{X}_i) \quad . \tag{1}$$

For firm *i*, where Y_i denotes the aggregate output, the vector $\mathbf{X}_i$ includes the levels of inputs traditionally specified in production functions (labor, capital, materials), and $g(A_i)$ represents shifts in the production function due to external factors underlying agglomeration economies. A_i denotes proxies for the degree of agglomeration, such as measures of density or scale (typically specified as the city size or density of employees). If a positive external economy exists, the coefficient of A_i should be positive and statistically significant.[5]

Previous studies feature a wide combination of proxies and productivity. Table 2 shows some proxies adopted in the literature using TFP. For example, Antonietti and Cainelli (2007) estimated firm-level TFP by adopting the Levinsohn-Petrin semi-parametric approach, and adopted the specialization index (LQ), the inverse of an Herfindahl index, and population density as proxies of agglomeration.

Table 2

[4] Cohen and Paul (2009) summarized "Marshallian" channels of agglomeration economies: labor market pooling, input sharing, and knowledge spillovers (pp.102).

[5] This simple approach attempts to capture the static external economy from the production side. There are also dynamic and general equilibrium approaches. A dynamic approach is shown in Henderson (2003).

Ⅱ-2 Recent Results

Despite the two trends presiding in the literature on agglomeration such as mentioned in introduction, this paper insists on the importance of evaluating firm performance within agglomerations. This viewpoint takes considerable interest in the controversial debate occurring in the economics of agglomeration. Antonietti and Cainelli (2007) state that "the effects of spatial agglomeration are still a puzzling question" and there is no "clear-cut conclusion." In other words, many contributions to the literature resulted in controversial effects (see Table 3).[6]

Table 3

Although the effects of agglomeration remain controversial, a greater number of recent studies have employed micro data. As Martin et al. (2008) theoretically investigated agglomeration economies affect firm-level TFP and not aggregate employment growth. It is noteworthy that many investigations using micro data have reported similar results. As Table 3 shows, many papers based on micro data suggest the presence of positive effects within localization-type agglomerations (clusters). Extending similar estimations to China by using micro data is valuable both to studies on the Chinese economy and to the economics of agglomeration.

Ⅱ-3 Focusing on the Firm Heterogeneity in Agglomeration

In addition, using micro data can shed light on a little-discussed topic—firm heterogeneity. Although the heterogeneity of firms is a long-established concept in management, it became an important issue in international economics only after Melitz (2003). Melitz investigates the question, "Within a developed country and within an industry, why do some firms export and others not?" Diversity of productivity among firms is his answer.[7]

Previous papers important for this study are Combes et al. (2009) and Arimoto et al. (2010). From the perspective of analyzing the urbanization economy with firm heterogeneity, Combes et al. (2009) identified and explained two channels for upgrading average productivity within agglomeration: the selection effect and the agglomeration effect. From the perspective of scholars interested in localization economies, Arimoto et al. (2010) provide a simpler empirical strategy, by employing summary statistics for productivity distributions.

---

[6] Cingano and Schivardi (2003) state a similar opinion.
[7] Baldwin and Okubo (2006) is the first paper that considers agglomeration economies with heterogeneous firms.

Agglomeration economics predicts that firms operating within clusters and cities are more productive. Combes et al. (2009) separates two previously undistinguished channels of productivity enhancement through clustering. The first is the agglomeration effect, which implies that all firms within agglomerations enjoy additional productivity from the effects of agglomeration. The second is the selection effect, which implies that more low-productivity firms within agglomerations go out of business because of concentrated competition. Using micro data, these two internal effects within agglomerations can be distinguished.

II-4 Theoretical Prediction

This paper does not describe details of the theoretical model of Combes et al. (2009). Rather, it introduces the core elements of their model, especially the causes of the two effects that were noted from agglomeration. The fundamental causes pertinent to the Combes model are the degree of competition, given by Melitz and Ottaviano (2008), and the interaction of workers, given by Lucas and Rossi–Hansberg (2002). The former predicts that competition is tougher in bigger markets and, as a result, low-productivity firms that might have survived in a small market cannot compete through price and must exit the large market. This is "the selection effect," or a left-truncation in the productivity distribution. The latter predicts that productivity can be upgraded in the high-density agglomerations through wider exchange of ideas among workers, whereby every firm within the agglomeration boosts its productivity. This is "the agglomeration effect," or a rightward shift in the productivity distribution[8]

[8] The former—the selection effect via market competition—is described in Combes et al. (2009) as follows:

$$\frac{N_i}{4\gamma}\int_0^{\bar{h}_i}(\bar{h}_i - h)^2 g(h)dh + \sum_{j\neq i}\frac{N_j}{4\gamma}\int_o^{\bar{h}_j/\tau}\left(\bar{h}_j - \tau h\right)^2 g(h)dh = s \ .$$

The right-hand side of this equation indicates the sunk cost of entering area *i*, and the left-hand side indicates aggregate profit for a firm. Thus, the equation expresses the free-entry condition. N_i indicates the population size of area i where that firm is located, γ indicates the degree of differentiation of produced goods, $\bar{h}_i$ indicates the marginal cost cutoff for area *i*, and every firm randomly drawn the own unit labor requirement h (say, the inverse of productivity) from a distribution $g(h)$. Finally, τ indicates iceberg transportation cost. Put simply, the first term on the left-hand side captures the operational profit from local sales, and the second-term summation captures the operational profit from out-of-local sales. Trade cost τ on the second term exists because the firm must pay the trade cost when selling outside the area.

In this model, τ plays a significant role, in other words τ determines the market size which firms facing and market size determine the degree of competitive pressure (say, selection in agglomeration). The case of local product-market competition corresponds to $\tau = \infty$, no firm can make any profit from outside, which implies perfect local competition. In this situation, larger market size of area *i* faces more consumers, more price elastic demand, further enter, lower markup (which means profitability). As a result, some low-productivity firms that would have been able to survive in a small market cannot lower price any further, and must exit from the large market.

The latter, the agglomeration effect by workers interaction is more directly modeled, as follows:

$$\varphi_i \equiv ln\left[a(N_i + \delta\sum_{j\neq i}N_j)\right] - \ln(h) \ ,$$

where φ_i denotes the log productivity in area *i*, $\ln(h)$ denotes log unit labor requirement of area *i* (say, unit input), $\left[a(N_i + \delta\sum_{j\neq i}N_j)\right]$ indicates the output, which is determined as the sum of local market size and N_i, and outside market size $\sum_{j\neq i}N_j$. The case of perfect local interaction corresponds to δ=0, implying that each market size directly determines productivity in area *i*. By contraries, the case of global interaction corresponds to δ=1, implying that no productivity differences exist among areas.

Figure 1 illustrates the four possible situations discussed in Combes et al. (2009). The X-axis indicates firm-level TFP, and the Y-axis indicates density. Panel (c) represents the case of local market competition and local interaction. In this case, TFP distribution for firms within the agglomeration is rightward shifted and left-truncated (both agglomeration and selection effects impact firms). Panel (a) illustrates local market competition and global interaction; only selection effect is observed because there is no regional upgrading effect through worker interaction.

Figure 1

After providing theoretical predictions, Combes et al. (2009) empirically analyzed France from the perspective of urbanization, by using French census data for all French firms.[9] Results suggest that the right-shift is observed, but left-truncation is not.

Building on Combes et al. (2009) from the perspective of localization, Arimoto et al. (2010) analyzed Japan's silk-reeling industry by using a simpler empirical approach.[10] They provided a different theoretical model that concludes that similar productivity distributions exist within and outside clusters. In short, interpretations of causality and the research angle differ from Combes's approach, but their theoretical predictions concerning productivity distribution are same as Combes et al. (2009).

The methodology appearing in Arimoto et al. (2010) is useful in conducting empirical analysis because it provides a simple, direct method to compare productivity distribution within and outside clusters. More precisely, they adopted mean, IQR (Inter-quartile-range, mid-fifty), and each percentile points as a benchmark of comparing two distributions.[11]

The rightward shift of distributions in clusters has been observed, and the higher level of each percentile point can be compared with the non-cluster distribution. In addition, if regional competition is tougher within clusters then IQR should be smaller inside the clusters, and both percentile points 10 and 25 should be larger within clusters. Table 4 summarizes the theoretical predictions of these measurements.

---

[9] They defined an area with more than 200,000 population as a French "City"(agglomeration area).

[10] In Arimoto et al.(2010) model, sunk entering cost S and market accessibility C play significant roles. First, their model shows that the region having lower sunk entry cost induces more firms and grows as a cluster. Second, a larger number of entrants makes competition more intense in clusters and pulls up the lower limit of productivity needed to survive. Third, increase in the number of firms within a region slides the distribution of the firms' log productivity to the right through the same mechanism as Combes et al. (2009).

[11] The approach of Combes et al. (2009) requires estimating two integrated indexes that independently represent right-shift and left-truncation. The advantages of their method are the availability of comparisons among industries and clear criteria for judging whether each effect exists. However, it requires more complicated calculations of each percentile point. Although Combes et al. (2009) note that comparing detailed summary statistics of distributions is not sufficiently informative, the author believes that the results of existing summary statistics approaches also provide a clear message if several robustness checks are conducted.

Table 4

Besides productivity, this paper examines profitability within agglomerations, a subject that is not empirically analyzed in previous studies, although Combes et al. (2009) and Arimoto et al. (2010) include theoretical predictions of profitability in the concept of a mark-up. As mentioned, in each model, tougher local competition determines that firms with high productivity and high mark-ups enjoy smaller profit margins but larger sales.[12] As a result, it can be assumed that profitability is lower within agglomerations than outside agglomerations.

Ⅲ Empirical Strategy, Data and Variables

This section describes details of the empirical analysis. Our empirical estimation process is principally based on Henderson (2003) and Arimoto et al. (2010).

Ⅲ-1 Estimating Firm-Level Production Function

First, we examine the mean effect of clustering. Following Henderson(2003) and Arimoto et al. (2010), the author estimates a firm-level production function with a proxy of agglomeration by OLS using heteroscedasticity-consistent standard errors as follow,

$$\ln V_k = \alpha + \beta(\ln K_k) + \gamma(\ln L_k) + \delta\left(\ln Cluster_j\right) \;, \tag{2}$$

for firm k, where V_k denotes value-added, K_k denotes capital input, L_k denotes labor input, $Cluster_j$ denotes the proxy for agglomeration where firm k is located. If the estimated coefficient δ is positive, the proxy for agglomeration positively affects productivity. The number of local firms and employees within the industry are adopted as proxy variables for agglomeration in this step.[13] Regarding localization, the number of local firms and employees within the industry are aggregated using firm-level data discussed later, and the author estimates equation (2) at county and city levels.[14]

Next, to examine both localization and urbanization, agglomeration proxy indexes are generated as follows:

$$localization\ index_{ij} = \left(\frac{P_{ij}}{P_i} + \frac{E_{ij}}{E_i}\right)100/2, \tag{3}$$

[12] See Combes et al. (2009), Melitz and Ottaviano (2008).

[13] Local quotient (LQ) is also a famous index. LQ captures the relative specialization of industrial structure in region j. However the author has estimated LQ by using this data set, which often became larger than 1 with only a few numbers of plants or employees in that region because it only focuses on the ratio of industrial structure. Thus, the author does not use LQ.

[14] Mainland China currently has 31 province-level administrative units, 333 city-level units, and over 2800 county-level units.

$$urbanizaion\ index_{ij} = (\frac{P_j - P_{ij}}{P - P_i} + \frac{E_j - E_{ij}}{E - E_i})100/2, \quad (4)$$

where P_{ij} denotes the number of plants for industry *i* in region *j*, and $P_i = \sum_j P_{ij}$ denotes the total number of plants in industry *i*. Therefore, $\frac{P_{ij}}{P_i}$ presents region *j*s share of plants in industry *i*. E_{ij} denotes the number of industry *i* employees in region *j*, and $E_i = \sum_j E_{ij}$ denotes the total number of employees in industry *i*. Thus, $\frac{E_{ij}}{E_i}$ captures region *j*s share of employment in industry *i*. In brief, equation (3) presents a localization index measured as the arithmetic mean of both shares in terms of the number of firms and employees in region *j* and industry *i*. Equation (4) captures the degree of urbanization, which is measured by the region's share of firms and employees in the region's total, while deducting a particular industry from the region and the total.

Following Henderson (2003), the following equation is estimated by OLS using heteroscedasticity-consistent standard errors at county and city levels to capture both localization and urbanization separately:

$$\ln V_k = \alpha + \beta_1(\ln K_k) + \beta_2(\ln L_k) + \delta_1(localization\ index_{ij}) + \delta_2(urbanization\ index_{ij}).[15] \quad (5)$$

To avoid multicollinearity, a single-regression model is invoked. If the result is unstable, further estimations using the following equations are performed:

$$urbanization\ index_{ij} = \alpha + \delta(localization\ index_{ij}) + \varepsilon_{ij} \quad , \quad (6)$$

$$localization\ index_{ij} = \alpha + \delta(urbanization\ index_{ij}) + \varepsilon_{ij} \quad , \quad (7)$$

where each ε_{ij} is not influenced by the other variable. Therefore, ε_{ij} in equation (6) can be adopted as an urbanization-independent variable.

Ⅲ-2 Analyzing Firm-Level ROA

Firm-level ROA as a proxy of profitability is analyzed by comparing ROA for firms operating within and outside clusters and cities. Definitions of clusters and cities are discussed in Section 3-4.

Ⅲ-3 Distributional Analysis of TFP

To estimate firm-level TFP and to compare results for firms within and outside agglomerations, two types of TFP are adopted: the non-parametric index type and the parametric residual type. The former is calculated as follows:

$$\ln TFP_k = (lnV_k - \overline{lnV}) - \left[\frac{1}{2}(w_{K,k} + \overline{w_K})(lnK_k - \overline{lnK}) + \frac{1}{2}(w_{l,k} + \overline{w_l})(lnL_k - \overline{lnL})\right]. \quad (8)$$

For firm *k* in an industry, V_k denotes value-added, and $\overline{lnV}$ denotes the geometric average of

[15] As shown equation (4), to compute the $urbanization\ index_{ij}$, the portions of industry *i* are deducted from the region's and the county's sum to avoid the correlation between the localization index and the urbanization index; however, the correlations between $localization\ index_{ij}$ and $urbanization\ index_{ij}$ are 0.51 in apparel, 0.81 in miscellaneous, 0.22 in fiber, 0.28 in auto, 0.66 in communication equipment industry at the county level. Although the correlation between the two variables is high particularly in miscellaneous industry, the author also estimates single regression model and residual model, focusing only on the robust result.

value-added for the industry. K_k and $\overline{lnK}$ denote the capital input of firm k and the geometric average of the industry, respectively. L_k and $\overline{lnL}$ also represent the same definition of labor input. $w_{K,k}$ denotes the cost share of capital input, and $w_{l,k}$ denotes the cost share of labor input. This TFP index calculates a firm's productivity as the difference from an average firm in the industry.[16] Later, residual TFP is calculated as a residual term φ_k in a Cobb–Douglas production function by OLS using heteroscedasticity-consistent standard errors as follows:[17]

$$\mathrm{ln}V_k = \alpha + \beta(lnK_k) + \gamma(lnL_k) + \varphi_k. \tag{9}$$

To compare the log productivity distribution within and outside agglomerations, the county- and city-level distribution statistics for TFP are computed, and estimations using the following equations are performed by OLS using heteroscedasticity-consistent standard errors:

$$IQR_{ij} = \alpha + \beta\big(localization\ index_{ij}\big) + \gamma(urbanization\ index_{ij}), \tag{10}$$

$$PP_{xx,ij} = \alpha + \beta\big(localization\ index_{ij}\big) + \gamma(urbanization\ index_{ij}), \tag{11}$$

where IQR_{ij} denotes the inter-quartile-range of productivity distribution for industry i in region j, $PP_{xx,ij}$ denotes each percentile point of productivity distribution for industry i in region j. IQR is a measure of dispersion, is also called the mid-fifty, equal to the difference of the third and first quartiles. PP captures the basic form of a distribution, percentile points 10, 25, 75 and 90 are adopted in these estimations.

The benchmark has been given in Table 4. Thus, statistical tests can assist in deciding whether the selection effect or the agglomeration effect influence the average difference in TFP within and outside agglomerations. In this step, regions that have three or fewer firms are dropped from the empirical regression because they are not steadily distributed in terms of log TFP.

Ⅲ-4 Definition of Agglomeration Areas

Two types of firm-level dummy variables are adopted to distinguish agglomerations and non-agglomerations for both localization and urbanization. Before presenting definitions, it is essential to introduce definitions used in previous researches. Combes et al. (2009) define cities as areas exceeding 200,000 in population. Arimoto et al. (2010) classify areas with plant densities above the median value as clusters. These two definitions are inappropriate for China, for which it is difficult to decide appropriate population criteria, and where plants are concentrated in the coastal areas. The approach of Arimoto et al. (2010) results in the majority of plants becoming

---

16 For more detail, see Good, et al. (1996). This productivity estimation takes the non-parametric approach. Thus, we can avoid edogeneity, and it assumes constant returns to scale and perfect competition.

17 The edogeneity problem inherent in this estimation method has already been indicated. Although Olley-Pyke provide solutions for endogeneity, the author does not adopt their approach because it requires at least three years of balanced data, which necessitates the omission of approximately a half of un-balanced firm data. For this reason also, the author takes the non-parametric approach introduced above.

cluster-located firms. Thus, alternatives to these two definitions need to be explored.

The author takes a cumulative approach to identify agglomerations. Agglomeration areas are defined as areas that have the largest numbers for $localization\ index_{ij}$, or for which the $urbanization\ index_{ij}$ accounts for 25% or 50% of the total. For instance, before some data were dropped from the data set, in miscellaneous goods industry, 33,521 firms were located in 1,652 counties, and 25% of firms were located in 27 counties; 50% of the firms were located in 98 counties. In this case, the 27 counties are defined as "25% firm number clusters," and the 98 counties are defined as "50% firm number clusters" in the industry. To summarize the results, $localization\ index_{ij}$ and $urbanization\ index_{ij}$ are used to compute this variable.

Ⅲ-5 Data and Variables

The firm-level data used in this study are provided by *Bureau Van Dijk* which name is *Qin*. It contains more than 300,000 firms in mainland China. As Hoshino (2011) mentioned, observations of this data set are highly correlated with official census data for "over scale manufacturing firm" (cor = 0.9927 at the provincial level). In other words, there is no obvious geographical sampling bias. To exploit the potential of this data set, all proxies of agglomeration are computed before dropping selected firm data.[18]

This study uses cross-sectional data for 2007 and selects five manufacturing industries, shown in Table 5.[19] Criteria for selecting these industries were adequate sample size, existence of clusters, and capital-labor ratios. These criteria were chosen for three reasons: a sufficient number of samples is required for the empirical strategy in this study; there should be identifiable clusters in each industry; and coexistence of labor-intensive and capital-intensive industries is an obvious feature of Chinese manufacturing. The industries were selected across the spectrum of capital-labor ratios.

Table 5

Next, the process of generating variables is described. Firm-level value-added is calculated as the

[18] Although the dropped *Qin* data set still has high correlation with original *Qin* data and census data (cor = 0.9429 and 0.9515 at provincial level), based on logit and probit model estimations, agglomeration-located firms are more likely to be dropped because of the lack of some financial data, therefore the author generates the area proxy variables before dropping data to avoid the spatial bias. This approach is necessary and effective because only about a half of firm data contains full financial data, while most of them contain geographical and employment data.

[19] This paper picks up the miscellaneous goods industry because it often forms localized cluster and accounts for approximately 10 % of China's manufacturing exports. Typically, it contains leather products, wood products, furniture, decorations, daily plastic products, glass products, pottery products, kitchen goods, and arts and crafts products. More precisely, the author collected China SIC code 1923, 1924, 1929, 2032, 2039, 2040, 21--, 223-, 2319, 24--, 2679, 2950, 3060, 3080, 3081, 3082, 3089, 3145, 3146, 3153, 3422, 3424, 3479, 3482, 3489, 4130, 4142, and 42--.

sum of administrative expenses, operating expenses, financial expenses, accumulated depreciation, and operating profits. Administrative expenses include wages and welfare payments. Capital input is defined as net fixed assets. Labor input is represented by the number of workers. Labor cost share $w_{l,i}$ is calculated as administrative expenses divided by value-added, and $w_{k,i}$ is defined as $(1 - w_{l,i})$.

To identify the geographical locations of firms, city name and postal code every firm data have were used. In this study, four municipality cities (Beijing, Chongqing, Shanghai, and Tianjin) are treated as a city-level region instead of a provincial-level region. Four-digit postal codes were employed to capture county-level location. Although a few four-digit postal codes do not correspond exactly to a county-level district, especially in urban centers, in most cases a postal code corresponds to county level.

Ⅳ Results of Estimation

Ⅳ-1 Results of Production Function Estimation

Before presenting the results of the regression analysis, Table 6 shows average firm-level labor productivity and non-parametric TFP within and outside agglomerations at county-level. These descriptive statistics are computed without regression. Results suggest that productivity in localized areas is lower than in non-agglomeration areas among labor-intensive industries such as apparel, miscellaneous goods, and fiber industry, and that there is no clear evidence of productivity in capital-intensive industries. In addition, although there is no clear evidence concerning productivity in urbanized areas, the average productivity in urbanized area is higher than that of in non-agglomeration areas in 14 cases of the total 20 cases.

Table 6

Next, it is examined by regression whether productivity within agglomerations is higher than outside agglomerations. Table 7 presents baseline results for localization economies that are measured by the number of firms and employees in the industry at county and city levels. Results suggest that negative localization economies exist in small magnitudes in all eight equations for the apparel and miscellaneous goods industries, with elasticities ranging from −0.027 to −0.0065. These results suggest that a 1% increase in an industry's number of firms or employees decreases firm output by 0.027% or less. On the other hand, positive localization economies are observed in all eight equations for the auto and communication equipment industries.

Table 7

Table 8 presents results of equation (5) at county and city levels. Although negative localization effect and positive urbanization effect are presented in model 3 and 7 in all five industries, some results in each industry are inconsistent with the single regression model. For instance, in miscellaneous goods industry, the estimated parameter of urbanization index in model 3 is inconsistent with the model 2 in both county- and city-level. Therefore, ε_{ij}, which is estimated by equations (6) or (7), was also used to assess the robustness in model 4, 8 in five industries. As results, the author found negative localization economy and positive urbanization economy in all industries as baseline; in particular, negative localization economy in the apparel, miscellaneous, and fiber industries, and positive urbanization economy in the auto and communication equipment industries are robust.

Table 8

Ⅳ-2 Results of ROA Estimation

Table 9 shows the results of ROA estimation by comparing within and outside the agglomerations at county-level. The definition of an agglomeration area is given in Section 3-4. Each ROA is found to be lower in agglomerations than non-agglomerations. It is conclusive that manufacturing firms in China earn lower profit margins within agglomerations.

Table 9

Ⅳ-3 Results of Distributional Analysis

Next, the shape of productivity within agglomerations is examined. As discussed in Section 2-4 and Table 4, ordinary selection affects only the lower tail of the log-productivity distribution, whereas agglomeration affects every point of the distribution.

Table 10 presents results of equation (10) and (11) in county-level, and Table 11 investigates the results of city-level estimation. In regard to the negative localization economies observed above, fewer probabilities of high-productivity firms are found among the apparel, miscellaneous goods, and fiber industries operating within localized clusters. All coefficients of PP_{75} PP_{90} are negative in these industries. For the miscellaneous goods industry, coefficients of PP_{10} and PP_{25} are also negative in both county- and city-level estimations. Hence, it can be concluded that the source of the negative localization economy in the miscellaneous goods industry is the negative agglomeration economy (a left-shift). Although there is no observable robust result for the apparel and fiber industries, most coefficients of PP_{10} and PP_{25} are positive, particularly at the city level, suggesting that ordinary selection occurs in these industries. Thus, the negative localization economy where the apparel and fiber industries operate arises from the negative selection effect (less high-productivity firms). In figure 2, the solid line indicates the 25 % cumulative localization index area, and the dashed line refers to non-cluster areas where the apparel industry is located.

On the other hand, the benefits of an urbanization economy are apparent for the auto and communication equipment industries. All coefficients of percentile points in both industries are positive, and most are statistically significant, at least at the 10% level. Hence, the cause of positive urbanization economy in capital-intensive industries is the agglomeration economy (rightward shift). In figure 3, the solid line indicates the 25 % cumulative urbanization index area, and the dashed line refers to non-city areas where the auto industry can be found. The "city" distribution is distinctly shifted rightward in this figure.

Table 10 (county level)

Table 11 (city level)

Figure 2 and 3

IV-4 Summary of Results

In this section, the four robust results from the investigation on firm-level data within agglomerations in China are investigated. First, the firms within a "Marshallian" localized cluster are less productive, particularly in labor-intensive industries. Second, a positive urbanization economy is found in the auto and communication equipment industries. Third, the average profitability is always lower within agglomerations than outside agglomerations. Fourth, as the shape of the log productivity distribution indicates, the negative localization economy is primarily explained by the negative selection effect in the apparel and fiber industries, and for the miscellaneous goods industry it is explained by the negative agglomeration economy. The positive urbanization economy in capital-intensive industries is caused by the agglomeration economy.

V Discussions

Results of this paper indicate negative localization economy, ordinary positive urbanization economy, and lower profitability within agglomerations in China. First, firms located in cities enjoy a wider exchange of ideas, perhaps elevating productivity. On the other hand, it is clear that low productivity within "Marshallian" localization clusters defies the theoretical prediction and make a sharp contrast with the results of previous works such as Arimoto et al.(2010). Thus, low productivity within China's clusters presents a new puzzling question. In addition, profitability within "city" and "cluster" is lower than that outside agglomeration areas.

Although the low productivity in localized clusters challenges the theoretical prediction, it is not surprising, as per a series of qualitative field research in China.[20] Much previous field researches have mentioned a large number of new entrants in China's cluster through the spillover of business information, and also that firms often attempt to imitate competitors' products and business models within clusters and then some leading companies attempt to "graduate" from the localized cluster to an urban cluster, cities such as Shanghai. It is not easy for small and medium enterprises in China's clusters to run their businesses outside the clusters, because typically they depend on information, infrastructure, and business linkage provided by the clustering. As a consequence, researchers have often observed intensive price competition, the low degree of product differentiation, and low profit margins within clusters.[21] Although the productivity of such firms is not high, their competitors are often firms in developing economies such as Brazil or Vietnam.[22] Hence, productivity lower than the national average might not matter within a cluster, and at the same time, a large amount of a specific industry have supported the local economic development.

Possibilities for additional interpretations arise. Recently, production costs have increased dramatically in China's coastal areas, and some firms have already moved inland or abroad. Thus, firms in labor-intensive industries might be more productive in inland China because of cheaper input costs. This might affect the productivity of firms in existing clusters, a phenomenon called "the industrial transformation hypothesis." In addition, theoretical predictions for agglomeration economies are primarily generated using the experience of developed economies, and often tested using data from these countries. Thus, testing the robustness of the theory requires paying much more attention to clusters in developing countries, and agglomeration theory should include the concepts of the developing stages of clusters and differences of industry.[23]

---

[20] See Sheng (2004), Marukawa (2001), Komagata (2005), Sonobe and Otsuka (2006), Marukawa(2009), and Wang (2009).

[21] For instance, regarding the features of China's localization clusters, Wang (2009) stated that "many clusters are at the bottom end of the global value chain, excessively depending on low price competition"(p.208), and Marukawa (2009) mentioned that "in Wenzhou, the speed of imitation is extremely fast, so the emergence of an industrial cluster can be observed in a short period. In other parts of China, the speed of emergence might be slower, but similar processes might be underway"(p.235).

[22] Bazan and Navas-Aleman (2003) mentioned that the major competitor of Brazil's well-known footwear cluster, Sinos Valley, is mainland China.

[23] Sonobe and Otsuka (2006) discuss the developmental stage of clustering. Okubo and Forslid (2010) pay attention to sectorial differences with a perspective of heterogeneous firms.

VI Concluding Remarks

This paper using firm-level data examined the relationship between agglomeration and firm-level performance in China. Based on the theoretical predictions in Combes et al. (2009) and Arimoto et al. (2010), and on the empirical strategy of Henderson (2003), the author quantitatively tested whether productivity is higher within agglomerations with a perspective of firm heterogeneity. Studies of agglomeration distinguish between localization and urbanization, and this paper investigated both types.

First, this paper found that productivity in labor-intensive industries is lower at a statistically significant level in small magnitudes within localized clusters than outside clusters. The source of negative localization economy is the negative selection effect in the apparel and fiber industries, and the negative agglomeration economy in the miscellaneous goods industry. Second, the positive urbanization economy that benefits capital-intensive industries is explained by the agglomeration effect. Third, profitability is lower within both types of agglomerations than outside agglomerations. In particular, low productivity in a localized cluster does not fit the theoretical prediction. Solving this puzzle of low productivity requires further empirical analysis by both quantitative and qualitative approaches and reconsideration of regional policy.

References

Antonietti, R., Cainelli, G. (2007). The Role of Spatial Agglomeration in a Structural Model of Innovation, Productivity, and Export: A Firm-Level Analysis, PRIN WP series.

Arimoto, Y., et al. (2010). Agglomeration or Selection? The Case of the Japanese Silk-reeling Industry, 1909-1916, RIETI discussion paper No.10-E-003.

Baldwin, R.E., Okubo, T. (2006). Heterogeneous Firms, Agglomeration and Economic Geography: Spatial Selection and Sorting, *Journal of Economic Geography*, 6.

Bazan, L., Navas–Aleman, L. (2003). Upgrading in Global and National Value Chains: Recent Challenges and Opportunities for the Sinos Valley Footwear cluster, paper presented at the EADI's workshop, October.

Ciccone, A. (2002). Agglomeration-Effect in Europe, *European Economic Review,* 46.

Cingano, F., Schivardi, F. (2004). Identifying the Sources of Local Productivity Growth, *Journal of European Economic Association,* 2.

Cohen, J.P., Paul, M. (2009). Agglomeration, Productivity, and Regional Growth: Production Theory Approaches, in: R. Capello, P. Nijkamp (Eds.), *Handbook of Regional Growth and Development Theories*, Edward Elgar.

Combes, P-P. (2000). Economic Structure and Local Growth: France, 1984-1993, *Journal of Urban Economics*, 47.

Combes, P-P. et al. (2008). *Economic Geography: The Integration of Regions and Nations*, Princeton University Press, Princeton and Oxford.

Combes, P-P. et al. (2009). The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection, CEPR WP series, 7191.

Ding, K. (2007). Domestic Market-based Industrial Cluster Development in Modern China, IDE-JETRO, Discussion Paper No.88.

Duranton, G., Overman, H.G. (2005). Testing for Localization Using Micro-Geographic Data, *Review of Economic Studies*, 72.

Feser, E.(2001). A Flexible Test for Agglomeration Economies in Two US Manufacturing Industries, *Regional Science & Urban Economics*, 31.

Glaeser, E.L. et al. (1992). Cities and Growth, *Journal of Political Economy*, 100.

Glaeser, E.L. ed. (2010). *Agglomeration Economics*, The University of Chicago Press, Chicago and London.

Good, D. et al. (1996). Index Number and Factor Demand Approaches to the Estimation of Productivity, Economic Research Reports of New York University, 96-34.

Henderson, J.V. (2003). Marshall's Scale Economies, *Journal of Urban Economics*, 53.

Hoshino, M. (2011). TFP of Manufacturing Enterprises in Coastal and Inland China, 2004-2007: Evidence from Listed Enterprises Panel Data, Kato, H., Wu, B. ed. *A Study on Urbanization and Regional Economic Development*, East China University of Science and Technology Press, Shanghai, pp.165-177 (in Chinese).

Komagata, T. (20005). *Theory of SMEs in Transforming China*, Zeimukanrikyokai (in Japanese).

Lall, S., et al. (2004). Agglomeration Economies and Productivity in Indian Industry, *Journal of Development Economics*, 73.

Liu, S. ed. (2008). *Annual Report on China's Industrial Cluster Development (2007-2008)* (*Zhongguo chanyejiqun fazhan baogao*), China Development Press, Beijing (in Chinese).

Lucas R.E. Jr., Rossi–Hansberg, E. (2002). On the Internal Structure of Cities, *Econometrica*, 70.

Martin, P., et al. (2008). Spatial Concentration and Firm-Level Productivity in France, CEPR Discussion Paper No. DP6858 .

Marukawa, T. (2001). China's Industrial Cluster: Growing Process and Structure, in: M. Seki (Ed.), *Asian Industrial Clusters*, IDE-JETRO (in Japanese).

Marukawa, T. (2009). The Emergence of Industrial Clusters in Wenzhou, China, in: B. Ganne and Y. Lecler (Ed.), *Asian Industrial Clusters, Global Competitiveness and New Policy Initiatives*, World Scientific, Singapore, pp.213-237.

Marukawa, T. (2010). The Transfer of Cutlery Industry from Japan to China, Paper presented at NDRC-ADB workshop, September.

Melitz, M.J. (2003). The Impact of Trade on Intra-industry Reallocation and Aggregate Industry Productivity, *Econometrica,* 71.

Melitz, M.J., Ottaviano, G. (2008). Market Size, Trade, and Productivity, *Review of Economic Studies*, 75 .

Nakamura, R. (1985). Agglomeration Economies in Urban Manufacturing Industries: A Case of Japanese Cities, *Journal of Urban Economics*, 17.

Okubo, T., Forslid, R. (2010). Spatial Relocation with Heterogeneous Firms and Heterogeneous Sectors, RIETI Discussion Paper Series 10-E-056.

Ruan, J., Zhang, X. (2009). Finance and Cluster-Based Industrial Development in China, *Economic Development and Cultural Change,* 58.

Sheng, S., Zheng, Y. (2004). *Zhejiang Phenomenon*, Tsinghua University Press, Beijing (in Chinese).

Sonobe, T. Otsuka, K. (2006). *Cluster-Based Industrial Development: An East Asian Model*, UK: Palgrave Macmillan.

Wang, J. (2009). New Phenomena and Challenges of Clusters in China in the New Era of Globalisation, in: B. Ganne and Y. Lecler (Ed.), *Asian Industrial Clusters, Global Competitiveness* *and New Policy Initiatives*, World Scientific, Singapore, pp.195-212.

Table 1: Shares of top 3 and top 5 provinces in 2008(quantity)

	Beer	Cigarettes	Chemical Fiber	Yarn	Cloth	Machine-made Paper and Paperboards	Sulfuric Acid	
Top 3 share	28.2%	29.9%	82.6%	60.6%	46.2%	47.2%	34.5%	
Top 5 share	40.9%	41.2%	87.8%	73.9%	56.3%	70.8%	48.2%	
	Caustic Soda	Soda Ash	Chemical Fertilizer	Chemical Pesticide	Ethylene	Primary Plastic	Cement	
Top 3 share	40.7%	43.7%	32.5%	53.9%	51.9%	34.3%	26.3%	
Top 5 share	52.8%	61.8%	45.5%	73.8%	68.5%	52.0%	40.3%	
	Plate Glass	Pig Iron	Crude Steel	Rolled Steel	Metal-cutting Machine Tools	Large and Medium-sized Tractors	Motor Vehicles	Cars
Top 3 share	37.6%	42.7%	41.7%	41.0%	61.8%	74.7%	27.3%	43.7%
Top 5 share	55.6%	56.8%	54.6%	53.4%	79.9%	89.8%	43.7%	62.5%
	Household Washing Machines	Household Refrigerators	Air Conditioners	Mobile Telephones	Micro-Computers	Integrated Circuit	Color Television Sets	
Top 3 share	66.7%	53.3%	64.1%	78.4%	89.1%	82.1%	64.0%	
Top 5 share	84.4%	78.2%	76.2%	91.0%	98.8%	93.8%	79.3%	

Source: *China Statistical Yearbook 2009*.

Table 2: Measument of agglomeration on related litelature using TFP

employment base			number of firm base		distance base	population base	
Herfindahl and Gini index	LQ	employees	Herfindahl and Gini index	LQ	Moran's I	population density	population
Arimoto, et al(2010) Ciccone (2002)	Antonietti and Cainelli (2007) Lall, et al (2004)	Feser (2001)	–	Henderson (2002)	Maria, et al(2008)	Antonietti and Cainelli(2007) Combes, et al(2009) Nakamura (1985) Lall, et al (2004)	Feser (2001)

Table 3: Main findings of previous works

	There are positive effects inside agglomerations	There are negative or no effect inside agglomerations
MAR-Localization Type	Henderson(2003:USA,TFP,firm-level) Cingano and Schivardi(2003:Italy,TFP,firm-level) Martin, et al(2008:France,TFP,firm-level) Antonietti and Cainelli(2007:Italy,TFP,firm-level)	Glaeser, et al(1992:USA,employment,aggregate-level) Combes(2000:employment,aggregate-level)
Jacob's-Urbanization Type	Glaeser, et al(1992:USA,employment,aggregate-level)	Henderson(2003:USA,TFP,firm-level) Cingano and Schivardi(2003:Italy,TFP,firm-level) Martin, et al(2008:France,TFP,firm-level) Antonietti and Cainelli(2007:Italy,TFP,firm-level)

Note：Country name indicates the object, next item indicates measurement of economy, and the final item indicates the type of data set.

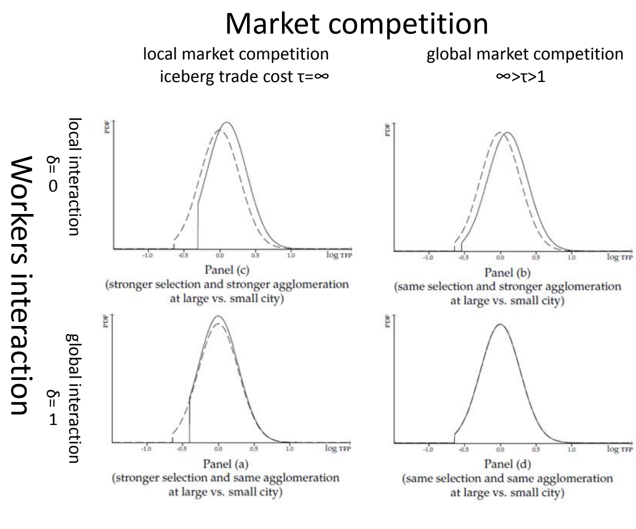

Figure 1: Theoretical prediction by Combes et al. of log productivity distribution

Note: Solid lines represent distribution of large cities (agglomeration area), dashed line represents small cities (non-agglomeration area)

Source: Combes et al. (2009)

Table4: Measures of productivity distribution in cluster relative to non-clusters

		mean	IQR	10 and 25 percentile	75 and 90 percentile
Case 1	Left truncate, selection effect	+	−	+	0
Case 2	Right shift, agglomeration effect	+	0	+	+
Case 3	Both effects	+	−	++	++
Case 4	Neither effect	0	0	0	0

Table 5: The five selected industries

	China SIC code	Observations	Capital per capita (thousand yuan)
Auto industry	372	4197	341.1
Communication equipment, computer, and other electronic devices industry	40	5446	220.5
Fiber industry	17	12902	171.6
Miscellaneous goods industry	see appendix	17174	104.4
Apparel industry	18	6807	62.5

Note: Capital per capita is calculated as net fix asset divided by number of labor.

Table 6: Descriptive statistics on firm-level productivity by non-parametric method at county-level

		Apparel industry		Miscellaneous goods industry		Fiber industry		Communication equipment industry		Auto industry	
		non-agglomeration	agglomeration	non-agglomeration	agglomeration	non-agglomeration	agglomeration	non-agglomeration	agglomeration	non-agglomeration	agglomeration
Labor productivity (thousand RMB)	50% localization	50.60603	49.35082	102.3394	78.76508	100.8978	92.19823	234.4307	186.2742	130.8426	162.1726
	25% localization	53.11876	41.80139	92.32166	79.59578	100.8948	83.22985	227.038	154.8525	146.6287	152.9629
Index TFP	50% localization	0.0027599	−0.0311027	0.0388359	−0.0750153	0.0020448	−0.0404811	−0.0207523	−0.0271711	−0.0894842	0.076706
	25% localization	0.0022992	−0.0634368	−0.0078423	−0.0750816	0.0379636	−0.1909462	−0.0360111	0.0107999	0.0315897	−0.0648986
Labor productivity (thousand RMB)	50% urbanization	49.25511	50.44424	81.19256	94.64435	85.36677	107.3774	254.7352	166.7406	125.0718	168.0974
	25% urbanization	49.32191	51.50469	83.27641	103.6781	93.38406	106.1078	229.6932	152.4856	125.8156	211.5269
Index TFP	50% urbanization	−0.033601	−0.0016496	0.0284189	−0.0741665	0.0326867	−0.071427	−0.0582232	0.0070088	−0.080206	0.0746084
	25% urbanization	−0.0477678	0.0717994	−0.0144976	−0.0656538	−0.0090358	−0.053417	−0.0523484	0.0516606	−0.1117162	0.3267652

Table 7: Basic result of localization economy estimated by absolute proxies

	Apparel industry				Miscellaneous goods industry				Fiber industry			
	model1	model2	model3	model4	model1	model2	model3	model4	model1	model2	model3	model4
ln(K)	0.5745	0.5751	0.5743	0.5748	0.5936	0.5942	0.5929	0.5933	0.6084	0.6089	0.6072	0.6077
	[55.65]***	[55.67]***	[55.58]***	[55.70]***	[108.25]***	[108.41]**	[108.08]**	[108.19]***	[89.44]***	[90.21]***	[89.39]***	[90.11]***
ln(L)	0.2963	0.2952	0.2992	0.2962	0.2877	0.2864	0.2925	0.2897	0.2855	0.2842	0.2872	0.287
	[22.16]***	[21.99]***	[22.19]***	[22.15]***	[43.45]***	[43.26]***	[43.80]***	[43.69]***	[32.26]***	[32.33]***	[33.11]***	[33.02]***
ln(no. of own industry firm in the county)	−0.0107				−0.0215				−0.0019			
	[−1.45]				[−5.06]***				[−0.47]			
ln(no. of own industry firm in the city)		−0.0096				−0.027				−0.006		
		[−1.19]				[−5.78]***				[−1.24]		
ln(no. of own industry employees in the county)			−0.0134				−0.0209				0.0056	
			[−2.00]**				[−5.93]***				[1.34]	
ln(no. of own industry employees in the city)				−0.0065				−0.0227				0.0031
				[−0.84]				[−5.46]***				[0.57]
constant	2.0672	2.0787	2.1347	2.0927	2.1105	2.1813	2.1917	2.2571	1.825	1.8547	1.7679	1.7819
	[25.31]***	[23.51]***	[22.34]***	[18.56]***	[46.56]***	[42.88]***	[42.80]***	[36.82]***	[35.56]***	[31.46]***	[29.37]***	[22.28]***
Adj-R-squared	0.6568	0.6568	0.6569	0.6567	0.7145	0.7147	0.7147	0.7146	0.7441	0.7441	0.7441	0.7441
observation	6807	6807	6807	6807	17174	17174	17174	17174	12902	12902	12902	12901

Note: T-value in brankets. *, **, and *** indicate p<0.1, p<0.05, and p<0.01.

	Communication equipment industry				Auto industry			
	model1	model2	model3	model4	model1	model2	model3	model4
ln(K)	0.6251	0.623	0.623	0.6225	0.6382	0.638	0.6368	0.6376
	[61.04]***	[61.14]***	[60.82]***	[60.91]***	[52.03]***	[51.99]***	[52.16]***	[51.88]***
ln(L)	0.2457	0.2489	0.2489	0.2507	0.3455	0.3457	0.3429	0.3458
	[19.35]***	[19.76]***	[19.16]***	[19.62]***	[20.67]***	[20.58]***	[20.40]***	[20.63]***
ln(no. of own industry firm in the county)	0.0315				0.005			
	[4.31]***				[0.56]			
ln(no. of own industry firm in the city)		0.0209				0.0043		
		[2.69]***				[0.52]		
ln(no. of own industry employees in the county)			0.0074				0.0198	
			[1.37]				[2.83]***	
ln(no. of own industry employees in the city)				0.005				0.0079
				[0.86]				[1.07]
constant	2.0511	2.0716	2.114	2.1274	1.4532	1.4498	1.3314	1.3968
	[27.42]***	[26.05]***	[25.80]***	[23.63]***	[17.36]***	[16.66]***	[14.50]***	[13.56]***
Adj-R-squared	0.7947	0.7942	0.794	0.794	0.8134	0.8134	0.8137	0.8134
observation	5446	5446	5446	5446	4197	4197	4197	4197

Note: T-value in brankets. *, **, and *** indicate p<0.1, p<0.05, and p<0.01.

Table 8: Firm-level localization and urbanization economies

	county-level				city-level			
	Apparel industry							
	model1	model2	model3	model4	model5	model6	model7	model8
ln(K)	0.5727	0.577	0.5748	0.5748	0.5755	0.5749	0.5755	0.5755
	[55.12]***	[55.76]***	[55.37]***	[55.37]***	[55.49]***	[55.47]***	[55.52]***	[55.52]***
ln(L)	0.3001	0.2949	0.3023	0.3023	0.2944	0.2954	0.2957	0.2957
	[22.21]***	[22.06]***	[22.35]***	[22.35]***	[21.87]***	[21.86]***	[21.84]***	[21.84]***
localization	-0.0372		-0.0784	-0.0374	-0.0065		-0.016	-0.0065
index	[-3.28]***		[-6.00]***	[-3.33]***	[-1.59]		[-2.14]**	[-1.59]
urbanization		0.0867	0.16			-0.0033	0.0184	
index		[3.47]***	[5.52]***			[-0.50]	[1.49]	
residual				0.16				0.0184
urbanization				[5.52]***				[1.49]
constant	2.0478	1.9809	1.9939	2.0179	2.0454	2.0319	2.0344	2.0387
	[27.46]***	[26.32]***	[26.47]***	[27.00]***	[27.41]***	[27.26]***	[27.28]***	[27.37]***
Adj-R-squared	0.6572	0.6574	0.6591	0.6591	0.6568	0.6567	0.6569	0.6569
observation	6807	6807	6807	6807	6807	6807	6807	6807

	Miscellaneous goods industry							
	model1	model2	model3	model4	model5	model6	model7	model8
ln(K)	0.5927	0.5939	0.588	0.588	0.5947	0.5957	0.5914	0.5914
	[108.04]**	[108.31]**	[106.04]**	[106.04]***	[108.30]**	[107.68]**	[105.90]**	[105.90]**
ln(L)	0.2912	0.2873	0.2978	0.2978	0.2878	0.2851	0.293	0.293
	[43.66]***	[43.46]***	[43.99]***	[43.99]***	[43.51]***	[42.71]***	[43.19]***	[43.19]***
localization	-0.0429		-0.1103	-0.0437	-0.0188		-0.0366	-0.0187
index	[-6.90]***		[-9.52]***	[-6.97]***	[-6.61]***		[-7.01]***	[-6.59]***
urbanization		-0.0097	0.1444			-0.0133	0.0277	
index		[-0.86]	[7.11]***			[-3.59]***	[4.07]***	
residual				0.1444				0.0277
urbanization				[7.11]***				[4.07]***
constant	2.0313	2.0193	2.0322	2.0419	2.0486	2.0317	2.0489	2.0521
	[50.09]***	[49.84]***	[50.33]***	[50.47]***	[50.05]***	[49.89]***	[50.14]***	[50.16]***
Adj-R-squared	0.7146	0.714	0.7154	0.7154	0.7147	0.7142	0.7149	0.7149
observation	17174	17174	17174	17174	17174	17174	17174	17174

	Fiber industry							
	model1	model2	model3	model4	model5	model6	model7	model8
ln(K)	0.6116	0.6072	0.6111	0.6111	0.6123	0.6083	0.6126	0.6126
	[89.88]***	[89.90]***	[89.93]***	[89.93]***	[90.68]***	[90.15]***	[90.58]***	[90.58]***
ln(L)	0.2822	0.2878	0.2835	0.2835	0.2781	0.2851	0.2784	0.2784
	[32.45]***	[33.36]***	[32.63]***	[32.63]***	[31.92]***	[33.00]***	[31.99]***	[31.99]***
localization	-0.0284		-0.0333	-0.0283	-0.0192		-0.0216	-0.0192
index	[-5.63]***		[-6.59]***	[-5.59]***	[-7.06]***		[-7.41]***	[-7.05]***
urbanization		0.0633	0.0866			-0.0081	0.0089	
index		[2.80]***	[3.83]***			[-1.65]*	[1.71]*	
residual				0.0866				0.0089
urbanization				[3.83]***				[1.71]*
constant	1.8235	1.8016	1.8049	1.8216	1.867	1.831	1.8569	1.8632
	[39.33]***	[38.63]***	[38.62]***	[39.31]***	[39.11]***	[37.99]***	[38.12]***	[38.80]***
Adj-R-squared	0.7445	0.7442	0.7448	0.7448	0.745	0.7441	0.745	0.745
observation	12902	12902	12902	12902	12902	12902	12902	12902

	Communication equipment industry							
	model1	model2	model3	model4	model5	model6	model7	model8
ln(K)	0.6186	0.6273	0.6165	0.6165	0.6224	0.6176	0.6137	0.6137
	[59.02]***	[60.93]***	[58.66]***	[58.66]***	[60.38]***	[60.56]***	[58.88]***	[58.88]***
ln(L)	0.2588	0.2493	0.2761	0.2761	0.2518	0.2585	0.2676	0.2676
	[19.80]***	[19.89]***	[20.95]***	[20.95]***	[19.53]***	[20.63]***	[20.33]***	[20.33]***
localization	-0.008		-0.038		0.0005		-0.006	
index	[-2.90]***		[-9.74]***		[0.29]		[-3.03]***	
urbanization		0.1177	0.2641	0.1169		0.0476	0.0589	0.0482
index		[6.15]***	[10.35]***	[6.47]***		[6.20]***	[6.81]***	[6.28]***
residual				-0.038				-0.006
localization				[-9.74]***				[-3.03]***
constant	2.1953	2.081	2.0446	2.0443	2.1761	2.094	2.0934	2.0826
	[31.30]***	[29.31]***	[28.93]***	[28.93]***	[31.16]***	[30.10]***	[30.14]***	[30.05]***
Adj-R-squared	0.7942	0.7956	0.799	0.799	0.7939	0.7955	0.7958	0.7958
observation	5446	5446	5446	5446	5446	5446	5446	5446

	Auto industry							
	model1	model2	model3	model4	model5	model6	model7	model8
ln(K)	0.6383	0.6352	0.635	0.635	0.638	0.6332	0.6309	0.6309
	[52.01]***	[51.68]***	[51.56]***	[51.56]***	[51.86]***	[50.98]***	[50.59]***	[50.59]***
ln(L)	0.3451	0.3504	0.351	0.351	0.3457	0.352	0.3531	0.3531
	[20.61]***	[20.85]***	[20.87]***	[20.87]***	[20.55]***	[20.82]***	[20.93]***	[20.93]***
localization	-0.0036		-0.0273		0.0024		-0.049	
index	[-0.21]		[-1.50]		[0.52]		[-5.94]***	
urbanization		0.1308	0.1444	0.13		0.0337	0.0901	0.0334
index		[5.31]***	[5.48]***	[5.25]***		[4.83]***	[7.33]***	[4.79]***
residual				-0.0273				-0.049
localization				[-1.50]				[-5.94]***
constant	1.4722	1.4369	1.4477	1.436	1.4649	1.4324	1.4763	1.4519
	[19.46]***	[19.57]***	[19.25]***	[19.61]***	[19.63]***	[19.59]***	[19.88]***	[19.77]***
Adj-R-squared	0.8134	0.8144	0.8145	0.8145	0.8134	0.8144	0.8159	0.8159
observation	4197	4197	4197	4197	4197	4197	4197	4197

Note: T-value in brankets. *, **, and *** indicate p<0.1, p<0.05, and p<0.01.

Table 9: Avarage ROA in agglomeration and non-agglomeration at county-level

	Industry	50% non-cluster	50% cluster	75% non-cluster	25% cluster
Localization	Apparel	15.1%	8.2%	12.9%	7.1%
	Miscellaneous goods	18.5%	7.7%	14.7%	6.3%
	Fiber	14.8%	8.8%	14.1%	4.8%
	Communication equipment	8.4%	5.6%	7.3%	5.6%
	Auto	13.4%	7.8%	12.1%	5.9%

	Industry	50% non-city	50% city	75% non-city	25% city
Urbanization	Apparel	16.9%	6.2%	13.5%	4.8%
	Miscellaneous goods	18.8%	6.8%	14.8%	4.9%
	Fiber	17.5%	6.0%	13.7%	5.7%
	Communication equipment	8.0%	5.9%	7.4%	5.6%
	Auto	13.9%	7.1%	11.2%	7.9%

Table 10: The marginal effects of localization and urbanization indices to the distribution of firms' TFP at county-level

		residual TFP					index TFP				
		IQR	PP10	PP25	PP75	PP90	IQR	PP10	PP25	PP75	PP90
Apparel industry	County-level localization index	-0.0343	-0.0238	-0.0391	-0.0735	-0.1132	-0.0532	-0.0189	-0.0769	-0.1301	-0.1871
		[-0.83]	[-0.47]	[-0.89]	[-1.27]	[-1.50]	[-1.13]	[-0.42]	[-2.12]**	[-2.02]**	[-2.11]**
	County-level urbanization index	0.0239	0.0957	0.1663	0.1902	0.2908	0.011	0.0443	0.2011	0.212	0.3448
		[0.27]	[1.17]	[2.12]**	[1.73]*	[1.91]*	[0.12]	[0.64]	[2.66]***	[1.67]*	[2.04]**
	Observations	283	283	283	283	283	283	283	283	283	283
Miscellaneous goods industry	County-level localization index	-0.1277	-0.1099	-0.1503	-0.2781	-0.2606	-0.0907	-0.0205	-0.084	-0.1746	-0.1558
		[-2.78]***	[-2.26]**	[-3.27]***	[-3.90]***	[-3.35]***	[-1.47]	[-0.50]	[-1.81]*	[-1.88]*	[-1.70]*
	County-level urbanization index	0.0545	0.1472	0.1638	0.2182	0.2317	-0.1165	0.0176	-0.0286	-0.1451	-0.0848
		[1.04]	[1.85]*	[2.13]**	[2.35]**	[2.22]**	[-1.37]	[0.26]	[-0.40]	[-1.14]	[-0.64]
	Observations	599	599	599	599	599	599	599	599	599	599
Fiber industry	County-level localization index	-0.1401	0.0652	0.0147	-0.1253	-0.1587	-0.1793	0.0613	-0.0155	-0.1949	-0.238
		[-3.47]***	[1.70]*	[0.52]	[-2.91]***	[-3.08]***	[-3.99]***	[1.77]*	[-0.58]	[-3.67]***	[-3.67]***
	County-level urbanization index	0.0448	0.1958	0.1518	0.1966	0.2558	-0.0037	0.1409	0.0941	0.0904	0.1752
		[0.86]	[3.10]***	[2.63]***	[2.58]**	[2.33]**	[-0.04]	[2.44]**	[2.03]**	[0.80]	[1.61]
	Observations	512	512	512	512	512	512	512	512	512	512
Communication equipment industry	County-level localization index	-0.0123	-0.0095	-0.0152	-0.0275	-0.0054	-0.0223	-0.0381	-0.0345	-0.0568	-0.0454
		[-0.89]	[-0.61]	[-1.01]	[-1.07]	[-0.14]	[-1.47]	[-2.84]***	[-2.74]***	[-2.42]**	[-1.35]
	County-level urbanization index	0.1345	0.1523	0.1334	0.2679	0.2025	0.1614	0.2196	0.1509	0.3123	0.3119
		[1.81]*	[1.63]	[1.70]*	[2.23]**	[1.25]	[1.85]*	[2.46]**	[1.82]*	[2.23]**	[1.71]*
	Observations	205	205	205	205	205	205	205	205	205	205
Auto industry	County-level localization index	0.0249	0.0087	0.0039	0.0288	0.0375	0.1302	0.0393	0.0968	0.227	0.2568
		[0.68]	[0.31]	[0.14]	[0.73]	[0.72]	[1.70]*	[0.49]	[1.36]	[1.99]**	[1.82]*
	County-level urbanization index	-0.0183	0.1716	0.1437	0.1254	0.1082	0.2049	0.1175	0.1304	0.3353	0.4176
		[-0.33]	[2.72]***	[3.07]***	[1.73]*	[1.23]	[1.14]	[0.60]	[1.01]	[1.64]	[1.76]*
	Observations	255	255	255	255	255	255	255	255	255	255

Note: T-value in brankets. *, **, and *** indicate p<0.1, p<0.05, and p<0.01.

Table 11: The marginal effects of localization and urbanization indices to the distribution of firms' TFP at city-level

		residual TFP					index TFP				
		IQR	PP10	PP25	PP75	PP90	IQR	PP10	PP25	PP75	PP90
Apparel industry	City-level localization	-0.0481	0.039	0.0015	-0.0466	-0.0759	-0.0596	0.0501	-0.0041	-0.0637	-0.1127
		[-1.82]*	[1.14]	[0.05]	[-1.69]*	[-1.84]*	[-2.14]**	[1.72]*	[-0.15]	[-2.18]**	[-2.69]***
	City-level urbanization	0.0224	-0.0081	0.0343	0.0568	0.1325	0.0399	-0.0263	0.0315	0.0714	0.1751
		[0.49]	[-0.14]	[0.67]	[1.30]	[2.30]**	[0.85]	[-0.55]	[0.62]	[1.60]	[3.03]***
	Observations	116	116	116	116	116	116	116	116	116	116
Miscellaneous goods industry	City-level localization	-0.0543	-0.0278	-0.0395	-0.0938	-0.0956	-0.0174	-0.0037	-0.0127	-0.03	-0.0259
		[-2.76]***	[-0.90]	[-1.43]	[-3.04]***	[-2.67]***	[-0.68]	[-0.15]	[-0.64]	[-1.10]	[-0.67]
	City-level urbanization	-0.0295	0.1117	0.0923	0.0628	0.0311	-0.0818	0.0797	0.0485	-0.0333	-0.0749
		[-0.99]	[2.27]**	[2.19]**	[1.45]	[0.63]	[-2.13]**	[1.90]*	[1.59]	[-0.88]	[-1.33]
	Observations	201	201	201	201	201	201	201	201	201	201
Fiber industry	City-level localization	-0.0428	0.0577	0.0253	-0.0176	-0.0453	-0.0408	0.0629	0.0172	-0.0236	-0.059
		[-3.39]***	[2.64]***	[1.65]	[-0.94]	[-1.93]*	[-2.66]***	[2.92]***	[1.35]	[-1.06]	[-2.05]**
	City-level urbanization	-0.0024	0.0505	0.0488	0.0464	0.0658	-0.0214	0.0465	0.0481	0.0267	0.0664
		[-0.12]	[1.62]	[2.22]**	[1.84]*	[2.00]**	[-0.82]	[1.62]	[2.56]**	[0.91]	[1.60]
	Observations	192	192	192	192	192	192	192	192	192	192
Communication equipment industry	City-level localization	-0.0024	-0.0004	-0.0037	-0.0061	-0.0094	-0.0096	-0.014	-0.012	-0.0217	-0.0233
		[-0.45]	[-0.07]	[-0.74]	[-0.75]	[-0.89]	[-1.45]	[-1.88]*	[-1.72]*	[-1.69]*	[-1.49]
	City-level urbanization	0.0254	0.0514	0.0209	0.0463	0.07	0.0316	0.0724	0.027	0.0586	0.101
		[0.88]	[1.77]*	[0.69]	[0.98]	[1.23]	[1.01]	[2.44]**	[0.90]	[1.11]	[1.50]
	Observations	92	92	92	92	92	92	92	92	92	92
Auto industry	City-level localization	-0.0155	0.005	-0.0171	-0.0325	-0.0485	-0.0482	0.017	0.0102	-0.0379	-0.0283
		[-0.95]	[0.26]	[-1.43]	[-1.61]	[-1.66]*	[-1.31]	[0.41]	[0.24]	[-0.84]	[-0.54]
	City-level urbanization	-0.0099	0.061	0.0652	0.0553	0.0726	0.1298	0.0574	0.0509	0.1807	0.1569
		[-0.40]	[2.25]**	[3.09]***	[1.75]*	[1.65]	[2.08]**	[0.83]	[0.66]	[2.79]***	[2.25]**
	Observations	142	142	142	142	142	142	142	142	142	142

Note: T-value in brankets. *, **, and *** indicate p<0.1, p<0.05, and p<0.01.

Figure2: Localization economy in apparel industry

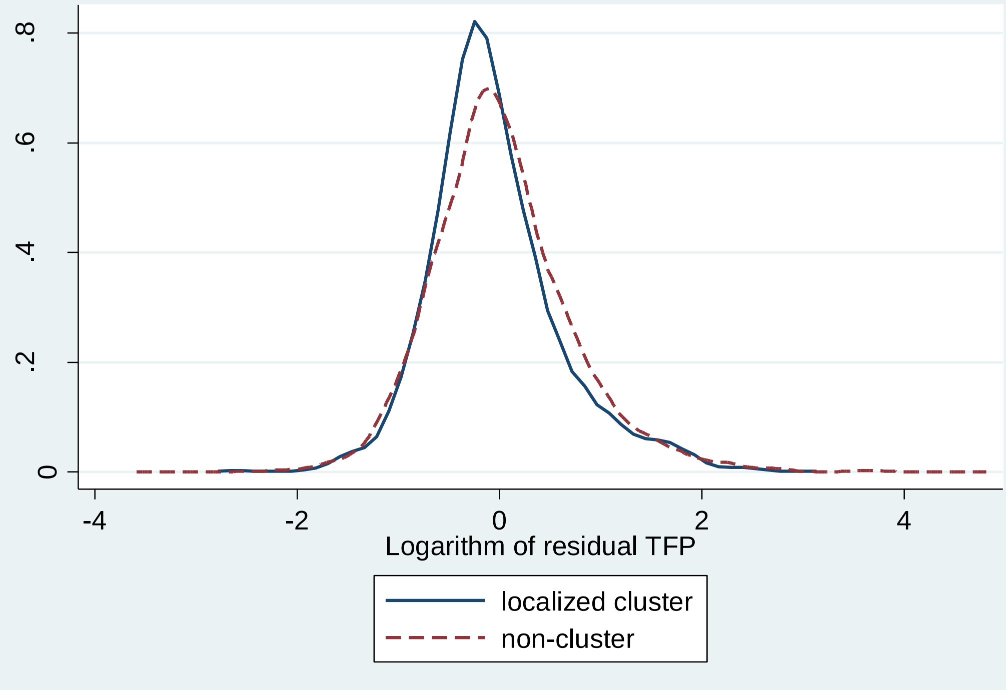

Figure 3: Urbanization economy in auto industry