Industrial agglomeration and firm size: Evidence from China

Dongya Li [a], Yi Lu [b,*], Mingqin Wu [c]

[a] *University of Hong Kong, Hong Kong*
[b] *National University of Singapore, Singapore*
[c] *South China Normal University, China*

ARTICLE INFO

ABSTRACT

This paper, by using annual surveys of manufacturing firms from 1998 to 2005 in China, first documents a positive correlation between industrial agglomeration and firm size, which is previously found in developed economies. Next, by using the instrumental variable estimations, we identify that industrial agglomeration has a positive and statistically significant causal impact on firm size. Finally, we find that firms are more likely to become larger by locating with a number of larger firms than with a larger number of firms.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Industrial activities are unevenly distributed across space, e.g., *manufacturing belt* in the United States (Fritz, 1943), *blue banana belt* in the European Union (Delamaide, 1994), and *Pacific coast industrial belt* in Japan (Kitamura and Yada, 1977).[1] The agglomeration of industrial activities has significant impacts on firm behavior and firm performance such as productivity (e.g., Ciccone and Hall, 1996; Henderson, 2003), organization of production processes (e.g., Holmes, 1999; Li and Lu, 2009), and innovation (e.g., Feldman and Audretsch, 1999; Carlino et al., 2007).

In two seminal papers, Kim (1995) and Holmes and Stevens (2002), by using plant-level data in the United States, find a positive correlation between industrial agglomeration and plant size both across and within manufacturing industries. Subsequent studies further confirm this finding by using datasets from other developed economies, e.g., Ireland (Barrios et al., 2006) and Italy (Lafourcade and Mion, 2007). An interesting research question is that whether the same pattern exists also in developing economies, where economic environments differ a lot from their counterparts in developed economies. And more

importantly, does the positive linkage between industrial agglomeration and firm size implies that firms become larger by locating in concentrated industrial areas or reflects the self-selection by larger firms into these areas? And how does industrial agglomeration affect firm size? Answers to these questions have important implications for both academic researches and government policies.[2] However, very few empirical studies have examined these issues. In this paper, we fill the void by investigating empirically the impact of industrial agglomeration on firm size in China.

China presents a good setting to study this topic. Before 1978, China adopted a central planning system, and nearly all economic activities including location choice and production scale were determined by the central government, which was largely influenced by political considerations. Confronted with the poor economic performance, however, China's central government started to reform its economy by gradually introducing private ownership and market competition in the late 1978. The economic reform not only induced a massive entry of privately-owned enterprises and foreign multinationals, but also restored incentives and decision-making powers in state-owned enterprises. As a result, there have been significant changes in the distribution of

---

[1] For more detailed description of spatial distribution of economic activities in the United States and Canada see Holmes and Stevens (2004); in the European Union see Combes and Overman (2004); and in Japan and China see Fujita et al. (2004).

[2] For example, if industrial agglomeration has a causal impact on firm size, it not only contributes to the studies of agglomeration economies as proposed by Marshall (1890), Arrow (1962) and Romer (1986), but also suggests that government policies of setting up industrial zones to attract investments have their merits as firm size is found to be correlated with many firm performance indicators (i.e., productivity).

industrial activities. For example, Bai et al. (2004) find an upward trend of industrial agglomeration in the latter half of 1985–1997, while Lu and Tao (2009) show that the industrial agglomeration continues during the period of 1998–2005. The fast-changing economic environment in China allows us to examine the interaction between industrial agglomeration and firm behavior.

Our dataset comes from annual surveys of manufacturing firms conducted by the *National Bureau of Statistics of China* for the period of 1998 to 2005. Using Holmes and Stevens (2002)'s specification, we find that firm size is positively and statistically significantly correlated with industrial agglomeration, which is consistent with the findings in the literature. Meanwhile, in terms of magnitude, the estimated coefficient in China (0.411) is very similar to that in the United States (0.436) as reported by Holmes and Stevens (2002).

To further investigate the relationship between industrial agglomeration and firm size, we follow Henderson (2003)'s estimation framework. Both the pooled OLS estimation and the panel fixed-effect estimation show that industrial agglomeration, measured by a firm's total neighboring employment within the same 4-digit industry and same city, has a positive and statistically significant impact on firm size. To identify whether industrial agglomeration has a causal impact on firm size, we use the instrumental variable estimation à la Li and Lu (2009). The instrumental variable estimation results substantiate our early findings, showing that industrial agglomeration causes firms to become large in production scale. We next include two measures of urbanization economies as in Holmes (1999) in the regression analysis to ensure that our results are not driven by urbanization economies (Jacob, 1969). It is found that our findings are robust to inclusion of these two additional measures, though urbanization economies also cast a positive impact on firm size.

To investigate how industrial agglomeration affects firm size, we decompose our measure of industrial agglomeration into two parts: the number and the average size of a firm's neighboring firms. It is found that both the number and the average size of a firm's neighboring firms have positive and statistically significant impacts on the firm's size, whereas the former has a smaller impact than the latter. These results suggest that a firm is more likely to become larger by locating with a number of larger firms than with a larger number of firms.

We interpret our findings as the evidence of localization economies. [3] To lend further support on localization economies, we conduct a regression of firm productivity on industrial agglomeration and find a positive and statistically significant estimated coefficient, which is consistent with the findings in the literature. However, there are other possible interpretations of our findings, such as product cycle theory (Dumais et al., 2002). To distinguish our interpretation from product cycle theory, we exclude newly-established firms in the regression analysis. Note that the product cycle theory implies that for old firms, those located in low agglomerated industrial areas should not have much difference in terms of size compared with their counterparts in high agglomerated industrial areas. The regression results without new firms show that industrial agglomeration still has a positive and statistically significant impact on firm size, and the magnitude of this impact is even increased compared with the regression results using the whole sample. These results indicate that the product cycle theory is not applicable in our case.

This paper contributes to the literature by identifying a positive impact of localization economies on an important indicator of firm performance, firm size. Firm size is found to play an important role in underpinning the impacts of institutions on economic growth, such as economic institutions (Laeven and Woodruff, 2007) and financial development (Beck et al., 2008). Meanwhile, firm size is found to have significant impacts on a number of important operation decisions, such as investment in R&D (Cohen and Levin, 1989), financing decisions

(Barclay and Smith, 1995a, 1995b), managerial compensation (Jensen and Murphy, 1990), required rate of returns (Banz, 1981; Fama and French, 1992), short-run employment fluctuations (Evans, 1987; Dunne et al., 1989; Audretsch and Mahmood, 1995; Davis et al., 1996), and long-run growth (Fukuyama, 1995). In recognition of the importance of firm size, recently there has been a number of studies investigating what determines firm size. These include the development of financial intermediary (Beck et al., 2006) and the quality of legal institutions (Kumar et al., 2004; Beck et al., 2006; Laeven and Woodruff, 2007). While these papers study the impacts of various economic institutions, our focus here is to investigate whether and how industrial agglomeration affects firm size.

The paper is also related to the recent literature on heterogeneous firms. For example, exporters are found to be larger and more productive than non-exporters (Bernard and Jensen, 1995, 1999, 2004; Clerides et al., 1998; Melitz, 2003; Melitz and Ottaviano, 2008), and the big region is most attractive for the more productive firms (Baldwin and Okubo, 2006). This paper is to investigate the firm heterogeneity (in terms of size) across geographic space and identify whether and how industrial agglomeration affects firm size.

The remainder of the paper is structured as follows. Section 2 describes data, and Section 3 presents our empirical findings. The paper concludes with Section 4.

2. Data

Our dataset comes from annual surveys of manufacturing firms conducted by the *National Bureau of Statistics of China* for the period of 1998 to 2005. These annual surveys cover all state-owned enterprises, and those non-state-owned enterprises with annual sales of 5 million RMB (Chinese currency) or more. The dataset also provides detailed information on firms' identification, operations and performance, including firm location, industry code and employment, which are of special interest to this study. As shown in Table 1 shows the total number of manufacturing firms ranges from 161,548 in 1999 to 270,209 in 2006. After deleting the observations with no valid information about employment and industry (at 4-digit SIC level), the number varies from over 140,000 in the late 1990s to over 244,000 in 2005. The last row of Table 1 reports the number of observations that is used to calculate location quotient at the firm-level (introduced in the next section).

For our study, we need precise location and industry information of our sample firms. During the sample period, China's administrative boundaries and consequently its county, city or even region[4] codes have experienced some changes. Meanwhile, in 2003 a new classification system for industry codes was adopted in China to replace the old classification system that had been used from 1995 to 2002. To achieve consistency in the whole sample period (1998–2005), we convert the location codes and industry codes of all firms to those of year 1998. For more discussion on the details of adjustment, see Lu and Tao (2009).

Moreover, firms in China may have branches located in regions other than its domicile, which may raise the concern of the multi-plants issue. However, according to the Article 14 of *The Company Law of the People's Republic of China*, "To set up a branch, the company shall file a registration application with the company registration authority, and shall obtain the business license." Thus, if a firm has branches that engage in business operations in regions other than its domicile, the *National Bureau of Statistics of China* collects each branch as a different observation in our dataset. For example, Beijing Huiyuan Beverage and Food Group Co., Ltd. has six branches, in Jizhong (Hebei Province), Youyu (Shanxi Province), Luzhong (Shandong Province), Qiqihar (Heilongjiang Province), Chengdu (Sichuan Province), and Yanbian

---

[3] See Rosenthal and Strange (2004, 2006) for reviews about the effects of localization economies.

[4] Region here refers to 22 provinces, 4 province-level municipalities, and 5 minority autonomous regions in China.

Table 1
Description of the dataset.

Year	1998	1999	2000	2001	2002	2003	2004	2005
Total observations	164,638	161,548	162,297	170,724	181,071	195,815	270,209	265,522
Observations with valid employment and industry figures	146,344	143,202	144,730	154,420	164,852	179,815	246,738	244,490
Observations with valid Q_f^x	143,925	140,782	142,373	152,080	162,590	177,546	244,645	242,414

(Jilin Province); the dataset accordingly counts them as six different observations belonging to six different regions, in addition to their parent company located in Beijing.

3. Empirical results

3.1. Benchmark

To give a first draw about the relation between industrial agglomeration and firm size, we follow Holmes and Stevens (2002)'s specification. Specifically, the measure for industrial agglomeration (the location quotient, $Q_{i,\,c,\,t}^x$) and the measure for firm size (the size quotient, $Q_{i,\,c,\,t}^s$) at the location-level are given by:

$$
\begin{cases}
Q_{i,c,t}^x = \dfrac{x_{i,c,t}\,/\,x_{c,t}}{x_{i,t}\,/\,x_t} \\[2mm]
Q_{i,c,t}^s = \dfrac{x_{i,c,t}\,/\,n_{i,c,t}}{x_{i,t}\,/\,n_{i,t}}
\end{cases}
\tag{1}
$$

where i, c, t represent 4-digit industry, city,[5] and year, respectively, $x_{i,c,t}$, $x_{i,t}$, $x_{r,t}$ and x_t are the total employment in industry i in city c, the total employment in industry i, the total employment in city c, and the total employment at year t, respectively; and $n_{i,c,t}$ and $n_{i,t}$ are the number of firms in industry i in city c and the number of firms in industry i at year t, respectively. Intuitively, $Q_{i,c,t}^x$ is the ratio of industry i's employment share in city c at time t over its total employment share in the economy at time t. Hence, if $Q_{i,c,t}^x > 1$, it means compared with the national average, city c at time t has a higher employment share of industry i; or in other words, industry i is agglomerated in city c at time t.

Columns 1–2 of Table 2 exhibit values of $Q_{i,c,t}^x$ and $Q_{i,\,c,\,t}^s$ for the whole example and each year, respectively. It shows that $Q_{i,r}^x$ experiences a modest decline in the late 1990s and then an steady increase in the early 2000s, which seems to contrast with the findings by Lu and Tao (2009) that the industrial agglomeration increases in the same period. The difference could result from that the location quotient at the city level ($Q_{i,c,t}^x$) fails to take into account the large plant issue as pointed out by Ellison and Glaeser (1997). To partially address this issue, Holmes and Stevens (2002) propose another measure of the location quotient and the size quotient, which are at the firm-level. Specifically, the location quotient ($Q_{f,t}^x$) and the size quotient ($Q_{f,t}^s$) at the firm-level are given:

$$
\begin{cases}
Q_{f,t}^x = \dfrac{\left(x_{i,c,t}-x_{f,t}\right)\,/\,\left(x_r-x_{f,t}\right)}{\left(x_{i,t}-x_{f,t}\right)\,/\,\left(x_t-x_{f,t}\right)} \\[3mm]
Q_{f,t}^s = \dfrac{x_{f,t}}{\left(x_{i,t}-x_{f,t}\right)\,/\,\left(n_{i,t}-1\right)}
\end{cases}
\tag{2}
$$

where $x_{f,t}$ is the employment of firm f in industry i in city c at year t. As shown in Columns 4–5 of Table 2, $Q_{f,t}^x$ steadily increases from 4.086 in 1998 to 4.904 in 2005, which is consistent with Lu and Tao (2009)'s findings.

The relation between industrial agglomeration and firm size is captured by the correlation between the location quotient (i.e., $Q_{i,c,t}^x$ and

$Q_{f,t}^x$) and the size quotient (i.e., $Q_{i,c,t}^s$ and $Q_{f,t}^s$) (Holmes and Stevens, 2002):

$$
\begin{cases}
\beta_t^s = \dfrac{cov\left(q_{i,c,t}^s, q_{i,c,t}^x\right)}{var\left(q_{i,c,t}^x\right)} \\[4mm]
\beta_{f,t}^s = \dfrac{cov\left(q_{f,t}^s, q_{f,t}^x\right)}{var\left(q_{f,t}^x\right)}
\end{cases}
\tag{3}
$$

where lowercase q represents the natural logarithm of the uppercase counterpart. Columns 3 and 6 of Table 2 show the estimated coefficients for β_t^s and $\beta_{f,t}^s$ respectively. All the estimated coefficients are positive and statistically significant at the 1% level. Meanwhile, for the coefficients at the firm-level ($\beta_{f,t}^s$), the value increases from 0.044 in 1998 to 0.070 in 2005.

These results imply that the positive correlation between industrial agglomeration and firm size previously found in developed economies also hold in developing economies such as China. Moreover, over the period of 1998–2005, the correlation between industrial agglomeration and firm size is found to become stronger and stronger in China. In the following subsections, we further investigate whether this relation is robust to the control of unobserved industry and city characteristics as well as urbanization economies, whether industrial agglomeration has a causal impact on firm size, and how industrial agglomeration affects firm size.

3.2. Omitted variables issue

The above results regarding the positive correlation between industrial agglomeration and firm size could be biased due to some omitted variables such as industry and city characteristics.[6] To make sure the omitted variables do not bias our findings, we adopt the panel fixed-effect estimation à la Henderson (2003). More specifically, we estimate the following equation:

$$size_{fict} = \alpha_i + \beta \cdot agglomeration_{fict} + \gamma_t + \varepsilon_{fict} \tag{4}$$

where $size_{fict}$ is the logarithm of employment for firm f (that is located in industry i in city c) at year t; $agglomeration_{fict}$ measures the degree of industrial agglomeration, which is the logarithm of firm f's total neighboring employment in the same 4-digit industry i, same city c, and same year t; α_i is the firm dummy, capturing all time-invariant firm characteristics; γ_t is the time dummy, capturing all the effects affecting firms in the same year; and ε_{fict} is the error term. Standard error is clustered at the firm-level, to deal with the potential heteroskedasticity problem.

We first report the pooled OLS estimation results in Column 1 of Table 3, in which we replace firm dummy α_i in Eq. (4) with a full set of industry and province dummy.[7] It is found that industrial agglomeration has a positive and statistically significant estimated coefficient.

---

[5] City here is referred to as prefecture, rather than city proper.

[6] See Lu et al. (2009) for the discussion about the importance to deal with unobserved industry and region characteristics in estimating the agglomeration effect in China.

[7] We use 2-digit industry dummies instead of 4-digit industry dummies and province dummies instead of city dummies, because regressions with 1,274,501 observations and 517 4-digit industry dummies and/or 337 city dummies demand equipment that can do very intensive computing, to which we do not have access.

Table 2
Correlation between location quotient and size quotient.

	1	2	3	4	5	6
	Location level			Firm level		
	$Q_{i,r}^x$	$Q_{i,r}^s$	β^s	Q_f^x	Q_f^s	β_f^s
All	4.657	1.033	0.411	4.695	1.014	0.062
1998	4.896	1.032	0.443	4.086	1.015	0.044
1999	4.532	1.033	0.451	4.095	1.016	0.054
2000	4.629	1.031	0.437	4.508	1.015	0.055
2001	4.620	1.039	0.424	4.805	1.017	0.059
2002	4.680	1.028	0.413	4.911	1.014	0.057
2003	4.724	1.019	0.393	4.887	1.013	0.062
2004	4.517	1.052	0.383	4.952	1.010	0.075
2005	4.648	1.025	0.365	4.904	1.011	0.070

While the estimation thus far uses the whole sample, there is a concern that the inclusion of state-owned enterprises in the estimation may bias our estimation results. This is because state-owned enterprises enjoy favorable market positions and government policies, which may introduce a spurious correlation between firm size and industrial agglomeration. To address this problem, we re-do the analysis, focusing on a sub-sample of private firms (defined as firms with no equity held by the state). The estimation results are reported in Column 2 of Table 3. It is found that our finding regarding the positive impact of industrial agglomeration on firm size remains robust to the sub-sample of private firms.

Columns 3–4 of Table 3 present the panel fixed-effect estimation result corresponding to Eq. (4). The estimated coefficient of industrial agglomeration is still positive and statistically significant. These results are consistent with our previous findings that the industrial agglomeration has a positive impact on firm size, suggesting that our results are robust to the control for unobserved variables. Note that, however, the magnitude falls to 0.040 for the whole sample and 0.030 for the sub-sample of private firms, respectively. The drop of estimated coefficient with panel fixed-effect estimation could result from the control for time-invariant firm unobserved characteristics that may correlate with both the degree of industrial agglomeration and firm size. Meanwhile, it also could be possible that much of variations regarding industrial agglomeration and firm size lies among the inter-firm rather intra-firm differences.

Moreover, though the panel fixed-effect estimation is more efficient in controlling for time-invariant unobserved variables, it is more sensitive to measurement errors such as noise in annual variations in the data set, which may drive the estimated coefficient downward to zero (e.g., Griliches and Hausman, 1986; Angrist and Pischke, 2009; Hauk and Wacziarg, 2009). As a way of checking whether measurement errors are a severe problem in our estimation, we follow Griliches and Hausman (1986) in using an alternative way of estimating panel regression, that is, first-difference regression, and see whether the estimated coefficient from first-difference regression is still positive and statistically significant. More specifically, we take a long difference, that is, a 7-year first difference, transform of Eq. (4), i.e.,

$$\Delta size_{fic1998} = \beta \cdot \Delta agglomeration_{fic1998} + \gamma + \Delta\varepsilon_{fic1998}, \qquad (5)$$

where $\gamma \equiv \gamma_{2005} - \gamma_{1998}$; $\Delta size_{fic1998} \equiv size_{fic2005} - size_{fic1998}$; $\Delta agglomeration_{fic1998} \equiv agglomeration_{fic2005} - agglomeration_{fic1998}$; and $\Delta\varepsilon_{fic1998} \equiv \varepsilon_{fic2005} - \Delta\varepsilon_{fic1998}$. Regression results are reported in Columns 5–6 of Table 3. It is found that industrial agglomeration still has a positive and significant estimated coefficient, while the magnitude jumps to 0.094 for the whole sample and 0.075 for the sub-sample of private firms, respectively. The increase of estimated coefficient in the long-difference estimation is largely due to the sharp decrease in sample size, i.e., the number of observations falls from 1,274,501

(812,936) for the whole sample (sub-sample of private firms) in the panel fixed-effect regression to 34,781 (14,252) in the long-difference regression.

3.3. Endogeneity issue

One may cast the doubt on the results in Tables 3 and 4 that whether it is industrial agglomeration causing the increase of firm size because it could be possible that large firms choose to locate in agglomerated industrial areas.[8] To address these possible endogeneity issues and identify the causal impact of the industrial agglomeration on firm size, we use the instrumental variable estimation.

Following Li and Lu (2009), we use the historical population to instrument the degree of industrial agglomeration in contemporary China. More specifically, the instrument is the logarithm of population in each city in the year of 1986. The rationale for using this instrumental variable is based on two premises: the demand of a larger population attracts more manufacturers in each industry (see Krugman (1980) for a theoretical model, and Davis and Weinstein (2003) and Hanson (2005) for empirical supports), and the distribution of population persists over time (Davis and Weinstein, 2002).

As the instrument is time-invariant, it enters the panel instrumental variable estimation interacted with year dummies so that the instrument can be time-varying. Hence, the first-stage of the panel instrumental variable estimation is

$$agglomeration_{fict} = \alpha_i + \beta \cdot population_{c,1986} \times \gamma_t + \gamma_t + \varepsilon_{fict} \qquad (6)$$

The panel instrumental variable estimation results are reported in Columns 1–2 of Table 4, with Column 1 for the whole sample and Column 2 for the sub-sample of private firms. Panel B reports the first-stage estimation results. It is found that the instrument is positively and statistically significantly correlated with industrial agglomeration, consistent with the above argument. Meanwhile, the under-identification test confirms that our instrumental variable is relevant and the partial F-statistic suggests that the instrumental variable is also strong. Panel A reports the second-stage estimation results. Industrial agglomeration, after being instrumented, is still found to have a positive and statistically significant impact on firm size, which further supports our previous findings.

Note that the panel instrumental variable estimation has a much higher estimated coefficient than pooled OLS, panel fixed-effect and long-difference regression do. Meanwhile, the statistical significance of Hausman test implies that the panel instrumental variable estimation has a statistically different estimated coefficient from the corresponding panel fixed-effect estimation. Compared with the panel fixed-effect estimation, panel instrumental variable estimation can correct for the possible measurement errors problem and further control for time-varying unobserved characteristics. However, the potential problem with the panel instrumental variable estimation is that the instrument may not be entirely exogenous and given the possibility that the instrument may be positively correlated with the error term, the panel instrumental variable estimation could be over-estimated.

An alternative to carry out instrumental variable estimation in the panel setting is to first take a first-difference transform of Eq. (4) (so as to get rid of firm dummy α_i) and then adopt the two-stage-least-squares (TSLS) method to estimate the first-differenced equation using population in 1986 as the instrument. More specifically, in the long-

---

[8] Another possibility is that large firms attract industrial agglomeration. However, some recent studies find that small firms contribute more to external economies (e.g., Henderson, 2003; Faberman, 2007; Glaeser and Kerr, 2009; Rosenthal and Strange, 2003, 2009).

Table 3
Pooled OLS, panel fixed-effect and long-difference OLS estimation results.

	1	2	3	4	5	6
Estimation specification	OLS		Panel fixed-effect		Long-difference OLS	
Sample	Whole	Private firms	Whole	Private firms	Whole	Private firms
Agglomeration	0.098***	0.080***	0.040***	0.030***	0.094***	0.075***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.005)
Constant	3.291***	3.455***	4.591***	4.387***	− 0.073***	0.170***
	(0.018)	(0.020)	(0.008)	(0.010)	(0.005)	(0.008)
Year dummy	Yes	Yes	Yes	Yes	Yes	Yes
Firm dummy	–	Yes	Yes	Yes	–	–
Industry dummy	Yes	–	–	–	–	–
Province dummy	Yes	–	–	–	–	–
Number of observation	1,274,501	812,936	1,274,501	812,936	34,781	14,252
R-squared	0.1013	0.1060	0.0256	0.0039	0.0214	0.0155
p-value for F-Test	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: Standard errors, clustered at the firm-level, are reported in the parenthesis. *** indicates the statistical significance at the 1% level.
Overall R-squared is reported in Columns 3–4.

difference Eq. (5), we instrument $\Delta agglomeration_{fic1998}$ with $population_{c,1986}$.

The TSLS estimation results are reported in Columns 3–4 of Table 4, with Column 3 for the whole sample and Column 4 for the sub-sample of private firms. Panel B reports the first-stage results. It is found that our instrument is negatively and statistically significantly correlated our regressor of interest. Note that now the regressor of interest is a 7-year change or the growth rate of industrial agglomeration for this 7 years. The negative sign of the instrument reflects the catching up effect, that is, the more agglomerated an industry is, the slower growth in agglomeration it is. With respect to our central issue, industrial agglomeration is still found to be positive and statistically significant

in both the whole sample and the sub-sample of private firms. Note that compared with the panel instrumental variable estimation, the magnitude of the TSLS estimation now jumps to as high as 1.089 with an standard error of 0.242. The sharp increase in the estimated magnitude is accompanied by a sharp decrease in the sample, i.e., 33,221 compared to 1,088,762 in the panel instrumental variable estimation.

In summary, the instrumental variable estimation substantiates our previous findings that the industrial agglomeration has a positive and statistically significant causal impact on firm size. Given that each estimation method has its own trade-off and gives a different estimated coefficient, the economic magnitude of the impact of industrial

Table 4
Instrumental variable estimation results.

	1	2	3	4
Estimation method	Panel IV		TSLS	
Sample	Whole	Private firms	Whole	Private firms
	Panel A: second stage: dependent variable is size			
Agglomeration	0.224***	0.223***	1.089***	0.334***
	(0.074)	(0.051)	(0.242)	(0.063)
Year dummy	Yes	Yes	Yes	Yes
Firm dummy	Yes	Yes	–	–
Number of observation	1,088,762	663,418	33,221	13,686
	Panel B: first stage: dependent variable is agglomeration			
Population * year 1998	0.300**	0.438**		
	(0.150)	(0.217)		
Population * year 1999	0.307**	0.432**		
	(0.150)	(0.217)		
Population * year 2000	0.299**	0.411*		
	(0.150)	(0.217)		
Population * year 2001	0.304**	0.396*		
	(0.150)	(0.217)		
Population * year 2002	0.314**	0.398*		
	(0.150)	(0.217)		
Population * year 2003	0.313**	0.391*		
	(0.150)	(0.217)		
Population * year 2004	0.316**	0.387*		
	(0.150)	(0.217)		
Population * year 2005	0.295**	0.367*		
	(0.150)	(0.217)		
Population	–	–	− 0.032***	− 0.107***
	–	–	(0.007)	(0.011)
Year dummy	Yes	Yes	Yes	Yes
Firm dummy	Yes	Yes	–	–
Underidentification test	[359.257]***	[324.371]***	[23.12]***	[102.627]***
Partial F-statistic	[45.06]***	[40.80]***	[23.09]***	[103.46]***

Note: Standard errors, clustered at the firm-level, are reported in the parenthesis. *, ** and *** indicate the statistical significance at the 10%, 5% and 1% level, respectively.

Table 5
Urbanization economies.

Estimation specification	1	2	3	4	5
	Panel fixed-effect			Long-difference OLS	
Sample	Whole		Private firms	Whole	Private firms
Agglomeration	0.040***	0.030***	0.024***	0.063***	0.056***
	(0.001)	(0.001)	(0.001)	(0.004)	(0.005)
Urbanization 1	0.049***	0.027***	0.019***	0.048***	0.041***
	(0.002)	(0.001)	(0.002)	(0.005)	(0.008)
Urbanization 2		0.294***	0.174***	0.306***	0.184***
		(0.005)	(0.006)	(0.011)	(0.016)
Constant	4.128***	0.700***	2.049***	−0.140***	0.091***
	(0.017)	(0.067)	(0.081)	(0.006)	(0.010)
Year dummy	Yes	Yes	Yes	Yes	Yes
Firm dummy	Yes	Yes	Yes	–	–
Number of observation	1,253,927	1,253,927	800,761	34,160	14,071
R-squared	0.0136	0.0003	0.0009	0.0555	0.0319
p-value for F-Test	0.0000	0.0000	0.0000	0.0000	0.0000

Note: Standardized coefficients are reported. Standard errors, clustered at the firm-level, are reported in the parenthesis.
*** indicates the statistical significance at the 1% level. Overall R-squared is reported in Columns 1–3.

agglomeration on firm size is calculated in the following range: a one-standard-deviation increase in industrial agglomeration leads to a 0.06-standard-deviation (from the panel fixed-effect estimation) to 0.34-standard-deviation (from the panel instrumental variable estimation) increase in firm size for the whole sample of more than 1 million firms.

3.4. Urbanization economies

Since the seminal work by Glaeser et al. (1992), there emerges a large literature regarding the relative importance of localization economies and urbanization economies (e.g., Henderson et al., 1995; Henderson, 1997, 2003; Quigley, 1998; Maurel and Sedillot, 1999; Combes, 2000; Gao, 2004). To rule out the concern that our findings could be driven by urbanization economies, we further include two measures related to urbanization economies as in Holmes (1999) in our analysis. The first one is the firm's total neighboring employment in the same city and same 2-digit industry but different 4-digit industry (denoted by *urbanization*1), and the second one is the firm's total neighboring employment in the same city but different 2-digit industry (denoted by *urbanization*2).

We stepwisely include these two urbanization measures and report regression results in Table 5. When only *urbanization1* is included (Column 1), both the statistical significance and magnitude of the industrial agglomeration index do not change, though *urbanization1* also has a positive effect on firm size. When *urbanization2* is further included in the regression, the estimated coefficient for *agglomeration* drops from 0.040 to 0.030 but it is still positive and statistically significant. Similar results are found in the sub-sample of private firms (Column 3) and in the use of long-difference estimation (Columns 4–5).

These results suggest that although urbanization economies have positive impacts on firm size and explain some parts of the impact of

Table 6
Decomposition of localization economies.

Estimation specification	1	2	3	4	5
	Panel fixed-effect			Long-difference OLS	
Sample	Whole		Private firms	Whole	Private firms
Agglomeration—number	0.037***	0.026***	0.021***	0.042***	0.045***
	(0.002)	(0.002)	(0.002)	(0.006)	(0.009)
Agglomeration—average size	0.040***	0.032***	0.028***	0.079***	0.065***
	(0.002)	(0.001)	(0.001)	(0.005)	(0.007)
Urbanization 1—number	0.027***	0.018***	0.036***	0.031***	
	(0.002)	(0.003)	(0.008)	(0.011)	
Urbanization 1—average size	0.026***	0.021***	0.053***	0.049***	
	(0.001)	(0.002)	(0.005)	(0.008)	
Urbanization 2—number	0.274***	0.152***	0.297***	0.149***	
	(0.006)	(0.007)	(0.014)	(0.022)	
Urbanization 2—average size	0.128***	0.085***	0.115***	0.078***	
	(0.002)	(0.003)	(0.005)	(0.008)	
Constant	0.086***	0.033***	−0.184***	−0.048***	0.149***
	(0.002)	(0.002)	(0.003)	(0.008)	(0.013)
Year dummy	Yes	Yes	Yes	Yes	Yes
Firm dummy	Yes	Yes	Yes	Yes	Yes
Number of observation	1,274,501	1,253,927	800,761	34,160	14,071
R-squared	0.0479	0.0040	0.0081	0.0622	0.0370
p-value for F-Test	0.0000	0.0000	0.0000	0.0000	0.0000

Note: Standardized coefficients are reported. Standard errors, clustered at the firm-level, are reported in the parenthesis. *** indicates the statistical significance at the 1% level. Standardized coefficients are reported. Standard errors, clustered at the firm-level, are reported in the parenthesis. *** indicates the statistical significance at the 1% level. Overall R-squared is reported in Columns 1–3.

Table 7

Evidence of localization economies.

	1	2	3	4	5
Estimation specification	OLS	Panel fixed-effect	Panel + IV	Long-difference OLS	TSLS
Dependent variable	Logarithm of output				
Panel A:	Whole				
Agglomeration	0.003***	0.002***	0.164***	0.003*	0.232**
	(0.000)	(0.001)	(0.064)	(0.002)	(0.105)
Constant	0.794***	2.064***	0.113***	0.188***	
	(0.008)	(0.024)	–	(0.005)	(0.035)
Set of inputs	Yes	Yes	Yes	Yes	Yes
Year dummy	Yes	Yes	Yes	Yes	Yes
Industry dummy	Yes	–	–	–	–
Province dummy	Yes	–	–	–	–
Firm dummy	–	Yes	Yes	–	–
Number of observation	1,162,787	1,162,787	954,554	33,690	32,191
Panel B:	Private firms				
Agglomeration	0.000	0.001**	0.116***	0.001	0.167***
	(0.000)	(0.000)	(0.029)	(0.002)	(0.036)
Constant	0.897***	2.076***	0.121***	0.200***	
	(0.012)	(0.031)	–	(0.007)	(0.019)
Set of inputs	Yes	Yes	Yes	Yes	Yes
Year dummy	Yes	Yes	Yes	Yes	Yes
Industry dummy	Yes	–	–	–	–
Province dummy	Yes	–	–	–	–
Firm dummy	–	Yes	Yes	–	–
Number of observation	738,240	738,240	567,818	14,131	13,568

Note: Standard errors, clustered at the firm-level, are reported in the parenthesis. *, **, and *** indicates the statistical significance at the 1%, 5%, and 10% level, respectively.

industrial agglomeration on firm size, our main findings are not primarily driven by these urbanization economies.

3.5. Decomposition of agglomeration economies

As the last analysis, we study how industrial agglomeration affects firm size. To achieve this end, we decompose the degree of industrial agglomeration into two parts, the number and the average size of a firm's neighboring firms. To address the issue that the number and the average size of a firm's neighboring firms may have different units of measurement, we report the standardized coefficients. Regression results are reported in Table 6.

It is clear that both the number and the average size of a firm's neighboring firms have positive and statistically significant impacts on its size. However, in terms of magnitude, the number of the firm's neighboring firms have a smaller impact (i.e., the standardized coefficient is 0.026) than that of the average size of the firm's neighboring firms (i.e., the standardized coefficient is 0.032).[9] These results imply that a firm is more likely to become larger by locating with a number of larger firms than with a larger number of firms.

3.6. Discussion

In the previous sections, we identify a positive impact of industrial agglomeration on firm size. We interpret this finding as the evidence of localization economies, that is, industrial agglomeration leads to the increase of firm size. One mechanism of localization economies is that industrial agglomeration facilitates the sharing of inputs and increase the degree of vertical specialization among firms (see Holmes, 1999 for the evidence in the United States; Li and Lu, 2009 for the evidence in China), which in turn increases firms' operation scales (Stigler, 1951; Chen, 2005; Lu and Tao, 2008). Another mechanism of localization economies is that industrial agglomeration improves the matching between the employer and the employee (Marshall, 1890; Costa and

Kahn, 2001). As the employment turnover and the quantity of resources used to fill up the vacancy are found to be significantly higher among large firms (Barron et al., 1985; Davis et al., 1996), the reduction of labor searching costs in agglomerated industrial areas leads to a equilibrium with large firm size (Wheeler, 2008).

To further lend support on localization economies in our data, we conduct a regression à la Henderson (2003), that is, to look at the impact of industrial agglomeration on firm productivity.[10] More specifically, we estimate the following equation

$$y_{fict} = \alpha_i + \beta \cdot agglomeration_{fict} + X'_{fict}\lambda + \gamma_t + \varepsilon_{fict}, \qquad (7)$$

where y_{fict} is the logarithm of total output and X_{fict} is a set of inputs (labor, capital, and materials; all in the logarithm form). Regression results are reported in Table 7, with Panel A for the whole sample and Panel B for the sub-sample of private firms. It is found that industrial agglomeration has a positive and statistically significant impact on firm productivity, consistent with the findings in the literature.

However, there are other possible interpretations of our findings on the positive impact of industrial agglomeration on firm size. One possible interpretation of our results is the product cycle theory, that is, new firms are established in fringe locations while high agglomerated areas contain old and large firms (Dumais et al., 2002). If the product cycle theory is dominant, the regression without new firms should find smaller or even no significant impact of industrial agglomeration on firm sizes as the implication of the product cycle theory is that for old firms, those located in low agglomerated industrial areas should not have much difference in terms of size compared with their counterparts in high agglomerated industrial areas. In Columns 1–3 of Table 8, we first restrict the analysis to a sub-sample of firms established after 1998 (the starting year of our data set), and then further restrict the analysis to a sub-sample of firms with age above 8 years (the sample median) in

---

[9] The p-value of the F-test of this difference is 0.01.

[10] We also experiment with using logarithm of capital–labor ratio as a proxy for firm efficiency and find similar results (available upon request).

Table 8
Exclusion of new firms.

	1	2	3	4	5	6
Estimation specification	OLS	Panel fixed-effect	Panel + IV	OLS	Panel fixed-effect	Panel + IV
Sample	Firms established before 1998			Firms with age above 8 years		
Agglomeration	0.109***	0.044***	0.245**	0.117***	0.036***	0.240***
	(0.002)	(0.001)	(0.101)	(0.002)	(0.002)	(0.086)
Constant	3.105***	4.689***	–	3.281***	4.973***	–
	(0.024)	(0.010)	–	(0.028)	(0.012)	–
Year dummy	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummy	Yes	–	–	Yes	–	–
Province dummy	Yes	–	–	Yes	–	–
Firm dummy	–	Yes	Yes	–	Yes	Yes
Number of observation	813,166	813,166	709,269	572,821	572,821	484,097

Note: Standard errors, clustered at the firm-level, are reported in the parenthesis. ** and *** indicate the statistical significance at the 5% and 1% level, respectively.

Columns 4–6 of Table 8. It is found that the estimated coefficients of industrial agglomeration are always positive and statistically significant, and the magnitudes are in line with the regression results using whole sample in Table 3. These results imply that the product cycle theory is not applicable in our case.

4. Conclusion

Empirical studies using datasets from developed economies find that firm size is positively related to industrial agglomeration. In this paper, we attempt to investigate whether industrial agglomeration also has positive impacts on firm size in developing economies, where economic environments differ a lot from their counterparts in developed economies. Moreover, on top of the positive correlation, we aim at identifying whether industrial agglomeration has positive causal impacts on firm's size and how industrial agglomeration affects firm size.

Using annual surveys of manufacturing firms from 1998 to 2005 in China, we first document a positive correlation between industrial agglomeration and firm size using both Holmes and Stevens (2002) and Henderson (2003)'s estimation specification. Second, by using the instrumental variable estimation, we find that industrial agglomeration has a positive and statistically significant causal impact on firm size. Third, through decomposition of industrial agglomeration, we find that firms are more likely to become larger by locating with a number of larger firms than with a larger number of firms.

Acknowledgement

References

Angrist, J.D., Pischke, J., 2009. Mostly Harmless Econometrics. Princeton University Press.
Arrow, K.J., 1962. The economic implications of learning by doing. The Review of Economic Studies 29, 155–173.
Audretsch, D., Mahmood, T., 1995. New firm survival: new results using a hazard function. The Review of Economics and Statistics 77, 97–103.
Bai, C., Du, Y., Tao, Z., Tong, S., 2004. Protectionism and regional specialization: evidence from China's industries. Journal of International Economics 63, 397–417.
Baldwin, R.E., Okubo, T., 2006. Heterogeneous firms, agglomeration and economic geography: spatial selection and sorting. Journal of Economic Geography 6, 323–346.
Banz, R.W., 1981. The relationship between return and the market value of common stocks. Journal of Financial Economics 9, 3–18.
Barclay, M., Smith, C., 1995a. The maturity structure of corporate-debt. Journal of Finance 50, 609–631.
Barclay, M., Smith, C., 1995b. The priority structure of corporate-liabilities. Journal of Finance 50, 899–916.
Barrios, S., Bertinelli, L., Strobl, E., 2006. Geographic concentration and establishment scale: an extension using panel data. Journal of Regional Science 46, 733–746.
Barron, J., Bishop, J., Dunkelberg, W., 1985. Employer search: the interviewing and hiring of new employees. The Review of Economics and Statistics 67, 43–52.
Beck, T., Demirgüç-Kunt, A., Maksimovic, V., 2006. The influence of financial and legal institutions on firm size. Journal of Banking and Finance 30, 2995–3015.
Beck, T., Demirgüç-Kunt, A., Laeven, L., Levine, R., 2008. Finance, firm size, and growth. Journal of Money, Credit and Banking 40, 1379–1405.
Bernard, A.B., Jensen, J.B., 1995. Exporters, jobs, and wages in the U.S. manufacturing: 1976–1987. Brookings Papers on Economic Activity: Microeconomics, pp. 67–119.
Bernard, A.B., Jensen, J.B., 1999. Exceptional exporter performance: cause, effect, or both? Journal of International Economics 47, 1–25.
Bernard, A.B., Jensen, J.B., 2004. Why some firms export. The Review of Economics and Statistics 86, 561–569.
Carlino, G.A., Chatterjee, S., Hunt, R.M., 2007. Urban density and the rate of invention. Journal of Urban Economics 61, 389–419.
Chen, Y., 2005. Vertical disintegration. Journal of Economics and Management Strategy 14, 209–229.
Ciccone, A., Hall, R.E., 1996. Productivity and the density of economic activity. The American Economic Review 86, 54–70.
Clerides, S.K., Lach, S., Tybout, J.R., 1998. Is learning by exporting important? Micro-dynamic evidence from Colombia, Mexico, and Morocco. Quarterly Journal of Economics 113, 903–947.
Cohen, W., Levin, R., 1989. Empirical studies of innovation and market structure. In: Schmalensee, R., Willig, R. (Eds.), Handbook of Industrial Organization, vol. 2. Elsevier, Amsterdam.
Combes, P.P., 2000. Economic structure and local growth: France, 1984–1993. Journal of Urban Economics 47, 329–355.
Combes, P.P., Overman, H.G., 2004. The spatial distribution of economic activities in the European Union. In: Henderson, J.V., Thisse, J.F. (Eds.), Handbook of Regional and Urban Economics, vol. 4. Elsevier, Amsterdam.
Costa, D.L., Kahn, M.E., 2001. Power couples. Quarterly Journal of Economics 116, 1287–1315.
Davis, D.R., Weinstein, D.E., 2002. Bones, bombs, and break points: the geography of economic activity. The American Economic Review 92, 1269–1289.
Davis, D.R., Weinstein, D.E., 2003. Market access, economic geography and comparative advantage: an empirical test. Journal of International Economics 59, 1–23.
Davis, S., Haltiwanger, J., Schuh, S., 1996. Job Creation and Destruction. MIT Press, Cambridge, MA.
Delamaide, D., 1994. The New Superregions of Europe. Dutton, New York.
Dumais, G., Ellison, G., Glaeser, E.L., 2002. Geographic concentration as a dynamic process. The Review of Economics and Statistics 84, 193–204.
Dunne, T., Roberts, M., Samuelson, L., 1989. The growth and failure of U.S. manufacturing plants. Quarterly Journal of Economics 104, 671–698.
Ellison, G., Glaeser, E.L., 1997. Geographic concentration in U.S. manufacturing industries: a dartboard approach. Journal of Political Economy 105, 889–927.
Evans, D.S., 1987. Tests of alternative theories of firm growth. Journal of Political Economy 95, 657–674.
Faberman, R.J., 2007. The relationship between the establishment age distribution and urban growth. Working paper.
Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. Journal of Finance 47, 427–486.
Feldman, M.P., Audretsch, D.B., 1999. Innovation in cities: science-based diversity, specialization and localized competition. European Economic Review 43, 409–429.
Fritz, W.G., 1943. Natural Resource—Minerals. Industrial Location and Natural Resources. U.S. Government Printing Office, Washington, DC.
Fujita, M., Mori, T., Henderson, J.V., Kanemoto, Y., 2004. Spatial Distribution of Economic Activities in Japan and China. In: Henderson, J.V., Thisse, J.F. (Eds.), Handbook of Regional and Urban Economics, vol. 4. Elsevier, Amsterdam.
Fukuyama, F., 1995. Trust: The Social Virtues and the Creation of Prosperity. Free Press, New York.
Gao, T., 2004. Regional industrial growth: evidence from Chinese industries. Regional Science and Urban Economics 34, 101–124.
Glaeser, E.L., Kerr, W.R., 2009. Local Industrial Conditions and Entrepreneurship: How Much of the Spatial Distribution Can We Explain? Journal of Economics and Management Strategy 18, 623-663.

Glaeser, E.L., Kallal, H.D., Scheinkman, J.A., Shleifer, A., 1992. Growth in cities. Journal of Political Economy 100, 1126–1152.

Griliches, Z., Hausman, J.A., 1986. Errors in variables in panel data. Journal of Econometrics 31, 93–118.

Hanson, G.H., 2005. Market potential, increasing returns and geographic concentration. Journal of International Economics 67, 1–25.

Hauk, W.R., Wacziarg, R., 2009. A Monte Carlo study of growth regressions. Journal of Economic Growth 14, 103–147.

Henderson, J.V., 1997. Externalities and industrial development. Journal of Urban Economics 42, 449–470.

Henderson, J.V., 2003. Marshall's scale economies. Journal of Urban Economics 53, 1–28.

Henderson, J.V., Kuncoro, A., Turner, M., 1995. Industrial-development in cities. Journal of Political Economy 103, 1067–1090.

Holmes, T., 1999. Localization of industry and vertical disintegration. The Review of Economics and Statistics 81, 314–325.

Holmes, T., Stevens, J.J., 2002. Geographic concentration and establishment scale. The Review of Economics and Statistics 84, 682–690.

Holmes, T., Stevens, J.J., 2004. Spatial distribution of economic activities in North America. In: Henderson, J.V., Thisse, J.F. (Eds.), Handbook of Regional and Urban Economics, vol. 4. Elsevier, Amsterdam.

Jacob, J., 1969. The Economy of Cities. Vintage, New York.

Jensen, M., Murphy, K.J., 1990. Performance pay and top-management incentives. Journal of Political Economy 98, 225–264.

Kim, S., 1995. Expansion of markets and the geographic concentration of economic activities: the trends in the U.S. regional manufacturing structure, 1860–1987. Quarterly Journal of Economics 110, 881–908.

Kitamura, Y., Yada, T., 1977. Nippon-Kougyo No Chiiki-Kouzou (in Japanese). Daimeido-Hakkou, Tokyo.

Krugman, P., 1980. Scale economies, product differentiation, and the pattern of trade. The American Economic Review 70, 950–959.

Kumar, K., Rajan, R., Zingales, L., 2004. What determines firm size? Working paper.

Laeven, L., Woodruff, C., 2007. The quality of the legal system and firm size. The Review of Economics and Statistics 89, 601–614.

Lafourcade, M., Mion, G., 2007. Concentration, agglomeration, and the size of plants. Regional Science and Urban Economics 37, 46–68.

Li, B., Lu, Y., 2009. Geographic concentration and vertical disintegration: evidence from China. Journal of Urban Economics 65, 294–304.

Lu, Y., Tao, Z., 2008. Vertical integration and firm performance. Working paper.

Lu, J., Tao, Z., 2009. Trends and determinants of China's industrial agglomeration. Journal of Urban Economics 65, 167–180.

Lu, Y., Ni, J., Tao, Z., 2009. City-industry growth in China. Working paper.

Marshall, A., 1890. Principles of Economics. Macmillan, London.

Maurel, F., Sedillot, B., 1999. A measure of the geographic concentration in French manufacturing industries. Regional Science and Urban Economics 29, 575–604.

Quigley, J.M., 1998. Urban diversity and economic growth. Journal of Economic Perspectives 12, 127–138.

Romer, P.M., 1986. Increasing returns and long-run growth. Journal of Political Economy 94, 1002–1037.

Rosenthal, S.S., Strange, W.C., 2003. Geography, industrial organization, and agglomeration. The Review of Economics and Statistics 85, 377–393.

Rosenthal, S.S., Strange, W.C., 2004. Evidence on the nature and sources of agglomeration economies. In: Henderson, J.V., Thisse, J.F. (Eds.), Handbook of Regional and Urban Economics, vol. 4. Elsevier, Amsterdam.

Rosenthal, S.S., Strange, W.C., 2006. The micro-empirics of agglomeration economies. In: Arnott, D., McMillen, P. (Eds.), A Companion to Urban Economics. Wiley-Blackwell, Malden, MA.

Rosenthal, S.S., Strange, W.C., 2009. Small establishments/big effects: agglomeration, industrial organization, and entrepreneurship. In: Glaeser, E. (Ed.), The Economics of Agglomeration. National Bureau of Economic Research.

Stigler, G.J., 1951. The division of labor is limited by the extent of the market. Journal of Political Economy 59, 185–193.

Wheeler, C.H., 2008. Worker turnover, industry localization, and producer size. Journal of Economic Behavior and Organization 66, 322–334.