VIVES

2011

DISCUSSION PAPER

22

Markups and Agglomeration: Price Competition versus Externalities

Liqiu Zhao

liqiu.zhao@econ.kuleuven.be

Markups and Agglomeration: Price Competition versus Externalities[☆]

Liqiu Zhao*

*Vives Center for Regional Economics, Faculty of Economics and Business, Katholieke
Universiteit Leuven, Naamsestraat 61, 3000, Leuven, Belgium*

Abstract

Agglomeration can affect markups through two potential channels: agglomerated regions toughen competition (price competition effect) and firms are more productive on average in agglomerated regions (agglomeration externalities and firm selection effect). However, the literature is inconclusive on which force dominates. This paper models these two channels by introducing agglomeration economies to the model of Melitz and Ottaviano (2008). Under parameters from the empirical studies, I demonstrate that the price competition effect tends to dominate the others, i.e., firms in more agglomerated regions charge lower markups. Using a unique Chinese firm-level data from 2002 to 2004, I investigate the effect of spatial agglomeration on markups of firms. By addressing the potential endogeneity problems using instrumental-variable method, I find that in China an increase in the number of own-industry firms in the same region has a negative causal effect on markups of firms and a positive effect on productivity. But firms in agglomerated regions have higher output and profit.

Keywords: Markups, Price competition, Agglomeration externalities

JEL: L11, R12

[☆]I would like to thank Joep Konings, Gianmarco Ottaviano, Damiaan Persyn, Stijn Vanormelingen, Linke Zhu, and seminar and conference participants at K.U.Leuven, CEPR-JIE conference/school in Tel Aviv 2011. I gratefully acknowledge the financial support from the research council (LICOS excellence center) and VIVES. All errors are my own.

*Tel.: +32 16 324 227; fax: +32 16 326 599

Email address: liqiu.zhao@econ.kuleuven.be (Liqiu Zhao)

1. Introduction

There is a long tradition in applied industrial organization of estimating the influence of market structure, various competition and trade policies on market power, typically measured by markups. A number of studies have found evidences of significant association between markups and characteristics of industry structure, such as industrial concentration (Collins and Preston, 1969) and barriers to entry (Bain, 1956). Some economists are also interested in the impact of trade policies (Harrison, 1994; Konings and Vandenbussche, 2005) and competition laws (Kee and Hoekman, 2007; Konings et al., 2001) on markups. However, most of the studies ignore the spatial dimension in the research on markups. As noted by some economists, competition is somewhat localized (see Gabszewicz et al., 1986; Kaldor, 1935; Pinkse et al., 2002). In other words, each firm is likely to compete more intensely with its neighbors than with distant firms due to transportation costs and search costs. So it would be important and interesting to introduce the spatial dimension in the analysis of markups. Then the markups of firms depend not only on the total number of firms in the market but also on the spatial distribution of firms. This paper investigates the effect of geographical agglomeration on the markups of firms.

There are several potential channels through which agglomeration can affect markups. First, more own-industry firms within a certain region lead to tougher competition which lowers prices and thus markups (price competition effect). Second, the increased competition in an agglomerated market makes it harder for inefficient producers to operate profitably and thus drives out relatively inefficient (high-cost) producers from the market. So surviving firms in the market tend to have lower marginal costs and higher markups (selection effect). Third, firms benefit from agglomeration externalities through labor market pooling, input sharing, and knowledge spillovers as suggested by Marshall (1920), firmly supported by an abundance of empirical findings (see Melo et al., 2009; Rosenthal and Strange, 2004, for reviews and summaries of existing findings)¹, which leads to a lower marginal cost and higher

¹Productivity gains stemming from agglomeration externalities are geographically bounded as well. Some evidences suggest that agglomeration economies tend to attenuate rapidly across geographic space, for example, Rosenthal and Strange (2003) show that agglomeration economies attenuate rapidly over the first few miles, Duranton and Overman (2005) find that geographic localization mostly takes place at a relatively small

markups (agglomeration externalities). Therefore the impact of geographical agglomeration on markups is ambiguous, depending on the relative magnitude of agglomeration externalities, selection effect and price competition effect ².

In this paper, I first model these channels by extending the model of Melitz and Ottaviano (2008) which incorporates endogenous markups to allow for the externalities between firms within a region. Through the simulations of the model with key parameters from empirical studies, I find that the competition effect dominates the effect of agglomeration externalities, i.e., markups decrease with agglomeration. Using a unique firm level data from China, I find a significantly negative effect of agglomeration on markups and a significantly positive effect of agglomeration on firm level productivity. The empirical results confirm that the competition effect dominates the effect of agglomeration externalities in China. However, firms tend to have higher output and profit in agglomerated regions.

The contribution of this paper is threefold. Firstly, this paper models three channels through which agglomeration can affect markups and demonstrates that price competition effect tends to dominate under reasonable parameter values. Secondly, to the best of my knowledge this is the first study that examines the effect of geographical agglomeration on markups using Chinese firm level data. As argued in Krugman (2011), Chinese economic geography fits well into the New Economic Geography framework as most of the growth in manufacturing takes place in emerging countries, like China. It is of interest to test the theory applying Chinese firm level data. Finally, I investigate whether geographical agglomeration has a causal effect on the markups of firms using instrumental-variable estimation. In particular, I use the length of highways per square kilometer and the number of firms at provincial level in 1965 to instrument China's geographic concentration of industries in 2003 and 2004. The IV estimates are consistent with the basic results and the effect is stronger than what is captured by the OLS estimates.

The research is related with a number of theoretical studies in the literature. The "New Economic Geography" produces models involving both monopolistic competition and increasing returns to scale (Helpman, 1998; Krug-

scale, less than fifty kilometers.

²Distinguishing the selection effect from agglomeration externalities empirically is beyond the scope of this paper, more information can be found in Combes et al. (2009). I will use the term agglomeration externalities in the rest of this paper.

man, 1991b; Venables, 1996), summarized in Fujita et al. (1999). However, in their model monopolistic competition à la Dixit-Stiglitz framework ignores strategic interactions between firms and firms charge a constant markup. Ottaviano et al. (2002) introduce the pro-competition effect in a monopolistic competition model by using a linear demand system, which incorporates endogenous markups. Applying this linear demand system in a heterogeneous firm model, Melitz and Ottaviano (2008) show that bigger market induces tougher competition, leading to higher average productivity and lower average markups, but they do not take into account of the agglomeration externalities. This paper is also related with some empirical studies in the literature. Firms in larger markets are found to have lower markups but larger output in service and retail industries (Bresnahan and Reiss, 1991; Campbell and Hopenhayn, 2005). While using data from manufacturing industry and more specifically from ready-mixed concrete sector³, Syverson (2004, 2007) examines the impact of plant density on price and productivity respectively. He finds that firms in the market with higher plant densities have lower price but higher productivity. His two studies suggest that the change of markups defined as price over marginal cost is ambiguous as plant density increases. It is of academic and policy interest to identify the impact of firm density on markups. This study is one of the few studies that probe into it. Another relevant study is Combes and Lafourcade (2011) and they find that markups are greater in both core and peripheral regions but lower in-between which should produce further concentration in the core region. So the outcome in their paper is a short-run equilibrium.

The rest of the paper is organized as follows. The next section constructs a theoretical framework that formalizes the intuition above. Section 3 introduces the empirical strategy. Section 4 describes the data and variables. Section 5 and 6 present the empirical results, and Section 7 offers some concluding remarks.

2. Model

The Dixit-Stiglitz Model takes away all forms of strategic interaction among firms and assumes that the markup is constant, which is one of the major weaknesses. The model implies that the entry of new firms does not

³The spatial competition is probably stronger in service industries than in manufacturing industries suggested by Syverson (2007)

generate any pro-competitive effect, i.e., the markup is independent of the number of firms in the market. Ottaviano et al. (2002) introduce a linear demand model into economic geography, which incorporates endogenous markups and integrates weak interactions between firms. The model in this paper extends the work of Melitz and Ottaviano (2008), which applies the linear demand system, to allow for the externalities between firms within a region.

Demand. Following Ottaviano et al. (2002) and Melitz and Ottaviano (2008), I assume all consumers share the same utility function as follows:

$$U = q_0^c + \alpha \int_{i \in \Omega} q_i^c di - \frac{1}{2} \gamma \int_{i \in \Omega} (q_i^c)^2 di - \frac{1}{2} \eta \left(\int_{i \in \Omega} q_i^c di \right)^2 \quad (1)$$

where q_0^c is the quantity of numeraire good consumed, q_i^c is the quantity of variety i . The demand parameters α , η , and γ are all positive. The parameters α and η index the substitution pattern between the differentiated varieties and the numeraire. The parameter γ indexes the degree of product differentiation between the varieties. Then the inverse demand for variety i is given by:

$$P_i = \alpha - \gamma q_i^c - \eta Q^c \quad (2)$$

where $Q^c = \int_{i \in \Omega} q_i^c di$ is the consumption level over all varieties. Integrating the inverse demand equation over all varieties Ω , solving for Q^c and substituting this back into equation (2), I can solve for a linear demand for variety i as

$$q_i = \frac{\alpha L}{\eta N + \gamma} - \frac{L}{\gamma} P_i + \frac{\eta N}{\eta N + \gamma} \frac{L}{\gamma} \bar{P} \quad (3)$$

where L is the number of consumers in the market. Let $\Omega^* \subset \Omega$ be the subset of varieties that are consumed ($q_i > 0$). N is the measure of consumed varieties in Ω^* and $\bar{P} = 1/N \int_{i \in \Omega^*} P_i di$ is the average price.

For any positive demand $q_i \geq 0$,

$$P_{max} = \frac{1}{\eta N + \gamma} (\gamma \alpha + \eta N \bar{P}) \equiv C_D \quad (4)$$

where P_{max} is the price at which demand is driven to 0. The notation C_D is referred to the cut-off of the zero-profit condition. Given the linear demand system, price elasticity of demand can be written as $e_i = \left[\frac{C_D}{P_i} - 1 \right]^{-1}$, varying with the price of variety i and the cut-off level in the market.

Production. Labor is the only factor of production and is inelastically supplied in a competitive market. The numeraire good is produced under constant returns to scale, using one unit of labor per unit of output, and its market is competitive. Each variety is produced by a single firm. I extend the model of Melitz and Ottaviano (2008) by modeling the total cost TC_i to allow for the externalities in the following way⁴:

$$TC_i = \frac{c_i}{N^\lambda} q_i \quad (5)$$

where c_i is firm specific productivity⁵, N is the number of firms within a region. Provided that the information owned by firms is different, the benefit of communication generally increases as the number of local own industry firms rises. In other words, the agglomeration externalities arise from the number of local own industry firms. The parameter λ measures the extent of externalities between firms within a region, for example, $\lambda = 0$ implies that there are no externalities, $\lambda = 0.05$ means that an increase in the number of firms in a region by 1% increases the output by 0.05%. Here I assume $0 \leq \lambda < 1$. Comparing with Melitz and Ottaviano (2008), firms exhibit two forms of heterogeneity: they differ in levels of efficiency and in the number of firms in the market. The first form of heterogeneity is the result of random draws as in Melitz and Ottaviano (2008) while the second form is the result of the agglomeration externalities. However, the exploration of the sources of externalities is well beyond the scope of this paper. A producer's profit-maximizing price and quantity are

$$P_i = \frac{\alpha\gamma}{2(\eta N + \gamma)} + \frac{\eta N}{2(\eta N + \gamma)} \bar{P} + \frac{c_i}{2N^\lambda} = \frac{1}{2} \left(C_D + \frac{c_i}{N^\lambda} \right) \quad (6)$$

$$q_i = \frac{\alpha L}{2(\eta N + \gamma)} + \frac{\eta N}{2(\eta N + \gamma)} \frac{L}{\gamma} \bar{P} - \frac{L}{\gamma} \frac{c_i}{N^\lambda} = \frac{1}{2} \frac{L}{\gamma} \left(C_D - \frac{c_i}{N^\lambda} \right) \quad (7)$$

The optimal price is determined by the cut-off C_D in the market and firm specific marginal cost depending on a random draw c_i and the number of firms in the market N . As expected, equation (6) shows that lower cost

⁴The total cost function can be derived from a Cobb-Douglas production function $F(l, N) = AN^\lambda l$ given exogenous wage (labor is supplied in a competitive market).

⁵Low c_i implies high productivity.

firms charge lower prices, however, they have higher price-cost margins⁶. Meanwhile, lower cost firms sell more products.

In addition, I assume that productivity draws $1/c$ follows a Pareto distribution with lower productivity bound $1/c_M$ and a shape parameter k , then the distribution of cost draws c is given by

$$G(c) = \left(\frac{c}{c_M} \right)^k \quad (8)$$

Free entry condition. As in Melitz and Ottaviano (2008), I assume that if producers choose to enter into the market, they have to pay sunk entry cost f_E and then learn their own marginal cost of production c_i drawn from the Pareto distribution. Given the free entry condition, the expected profit of firms in equilibrium is zero,

$$\int_0^{C_D N^\lambda} \pi(c) dG(c) - f_E = \frac{L}{4\gamma} \int_0^{C_D N^\lambda} \left(C_D - \frac{c}{N^\lambda} \right)^2 dG(c) - f_E = 0 \quad (9)$$

Then the cut-off level C_D is determined by

$$C_D = \left[\frac{2\gamma(k+2)(k+1)c_M^k f_E}{L(N^\lambda)^k} \right]^{\frac{1}{k+2}}. \quad (10)$$

The cut-off level C_D decreases with the number of firms in a market. Intensified competition in the agglomerated market drives out inefficient producers from the market, which decreases the cut-off level.

Under the assumption of Pareto distribution, the average firm-specific productivity of surviving firms can be written as: $\bar{c} = \int_0^{C_D N^\lambda} c dG(c) / G(C_D N^\lambda) = \frac{k}{k+1} C_D N^\lambda$. Plugging $\bar{P} = (C_D + \frac{\bar{c}}{N^\lambda}) / 2$ into equation (4), I get the cut-off from the zero-profit condition :

$$C_D = \frac{2(k+1)\gamma\alpha}{\eta N + 2(k+1)\gamma} \quad (11)$$

The free entry condition (FE), i.e., equation (10) and zero-profit condition (ZP), i.e., equation (11), are depicted in Figure 1 under reasonable param-

⁶Firm i charges monopolistic price-cost margin $\frac{P_i - MC_i}{P_i} = 1 - 2c_i / (C_D N^\lambda + c_i)$ which is a decreasing function of c_i .

eters⁷. The cut-off level C_D^* and the number of firms in the market N^* are determined by the market size L and other parameters in the model. While increasing the market size, the curve representing FE shifts downwards from short dashed line to long short dashed line. The cut-off level decreases, suggesting that spatial competition exerts stronger selection process and only more productive firms can survive in the larger market. While the number of firms in the market increases but less than the increase in the market size. Furthermore, a market with externalities can support more firms than that without externalities.

Figure 1: Market size, cut-off, and equilibrium number of firms

Figure 2 shows the markups as a function of the number of firms in two different cases given other parameters. In one case, $\lambda = 1/4$ (high degree of externalities), in the other case, $\lambda = 1/20$ (low degree of externalities). When the degree of externalities among firms in the market is high and the initial number of firms is low, an increase of the number of firms would lead

⁷I impose restrictions on the choice of parameter values, such that the marginal utility of varieties should be non-negative. The value for the shape parameter of the productivity distribution k is set following recent empirical evidences (see Corcos et al., 2009, for example). The parameters are set as follows: $\alpha = 10$, $\gamma = 2$, $\eta = 1$, $f_e = 15$, $k = 2$, and $c_M = 10$. The solid and dashed line in Figure 1 represent equation (11) and (10), respectively. The intersection of FE and ZP (see Figure 1) is the equilibrium.

to a dramatic decrease of marginal cost. If this effect outweighs the spatial competition effect, the price-cost margin of firm i that is $1 - 2c_i / (C_D N^\lambda + c_i)$ will increase. However, When the initial number of firms in the region is high, the increase of firm's productivity due to an increase of the number of firms is relatively low, the price-cost margin will decrease. In the case of low extent of externality, the impact of an increase of firms on marginal cost is weak, which is smaller than the spatial competition effect, then the price-cost margin will decrease with agglomeration. The empirical studies on agglomeration economies show that the estimates of λ range roughly from 0.04 to 0.10 (Rosenthal and Strange, 2004), depending on the sector and details of the estimation procedure. Therefore, I would expect that the effect of agglomeration on price-cost margins is likely to be negative in reality. With the same number of firms involved in a region, the markups in a market with externalities are larger than that without externalities.

Figure 2: Price-cost margins and number of firms: $\lambda = 1/4$ and $\lambda = 1/20$

3. Empirical Strategy

The methodology for estimating markups is based on Roeger (1995), which is an extension of Hall (1988, 1990). Since this approach only requires firm level data and is easy to estimate, it has been applied in many

studies investigating the impact of institutional change, trade liberalization on market power, for example, Levinsohn (1993) and Harrison (1994), Konings et al. (2005), Kee and Hoekman (2007), etc. I give a brief summary of the approach below, for detailed derivations I refer the reader to Roeger (1995).

Consider a linear homogeneous production function which allows for the externalities between firms $Q_{it} = A_{it}N_{it}^{j^r\lambda}F(K_{it}, L_{it}, M_{it})$, where i is the firm index, t is the time index, K_{it} , L_{it} and M_{it} are capital, labor and material inputs for firm i , respectively. A_{it} represents productivity efficiency and $N_{it}^{j^r}$ is the number of own-industry firms in region r . Under imperfect competition, Hall (1988) shows that the primal Solow residual can be decomposed into an imperfect competition term and a productivity term⁸.

$$\begin{aligned} SR_{it} &\equiv \Delta q_{it} - \alpha_{Lit}\Delta l_t - \alpha_{Mit}\Delta m_{it} - (1 - \alpha_{Lit} - \alpha_{Mit})\Delta k_{it} \\ &= B_{it}(\Delta q_{it} - \Delta k_{it}) + (1 - B_{it})\lambda\Delta n_{it}^{j^r} + (1 - B_{it})\Delta a_{it} \end{aligned} \quad (12)$$

where $\alpha_{Jit} = P_{Jt}J_{it}/P_{it}Q_{it}$ ($J = K, L, M$) denotes the cost shares of inputs in turnover. Δq_{it} , Δk_{it} , Δl_t and Δm_{it} are growth rates of output, capital, employment and material input, respectively. $\Delta n_{it}^{j^r}$ is the growth rate of the number of own industry firms in region r . Market power is captured by the price-cost margin $B_{it} \equiv (P_{it} - MC_{it})/P_{it}$, which is directly related to the markups via $\mu_{it} = 1/(1 - B_{it})$. In equation (12), the output, input factors and the cost shares of inputs can be observed from the data, thus equation (12) can be applied to estimate B_{it} . However, because of the potential correlation between unobserved productivity shocks and output (and input) growth, instrumental variables are required to estimate B_{it} . Roeger (1995) discusses this problem in some detail and suggests an alternative approach that does not rely on the use of instruments which are very difficult to find in micro data. Using the properties of the dual Solow residual under imperfect competition, Roeger derives a similar expression for the price based, or dual, Solow residual:

$$\begin{aligned} SRP_{it} &\equiv \alpha_{Lit}\Delta P_{Lit} + \alpha_{Mit}\Delta P_{Mit} + (1 - \alpha_{Lit} - \alpha_{Mit})\Delta P_{Kit} - \Delta P_{it} \\ &= -B_{it}(\Delta P_{it} - \Delta P_{Kit}) + (1 - B_{it})\lambda\Delta n_{it}^{j^r} + (1 - B_{it})\Delta a_{it} \end{aligned} \quad (13)$$

⁸As the externalities are allowed for, the Solow residual also depends on the growth in the number of own industry firms in the market.

where ΔP_{it} , ΔP_{Lit} , ΔP_{Mit} and ΔP_{Kit} are the growth rates of product price, wage, material price and the rental price of capital, respectively. By subtracting (13) from (12), the term capturing unobserved productivity shocks $(1 - B_{it})\Delta a_{it}$ cancels out:

$$\begin{aligned} SR_{it} - SRP_{it} &= (\Delta q_{it} + \Delta P_{it}) - \alpha_{Mit}(\Delta m_{it} + \Delta P_{Mit}) - \alpha_{Lit}(\Delta l_{it} + \Delta P_{Lit}) \\ &\quad - (1 - \alpha_{Mit} - \alpha_{Lit})(\Delta k_{it} + \Delta P_{Kit}) = B_{it}[(\Delta q_{it} + \Delta P_{it}) - (\Delta k_{it} + \Delta P_{Kit})] \end{aligned} \quad (14)$$

Equation (14) can easily be estimated with firm level data since all bracketed terms refer to growth rates of nominal values of output and inputs and can be observed from the data. Particularly, I need the information on growth rate of sales, growth in cost of employees, growth in material costs, and growth in the value of capital which are computable in my dataset. For capital, I use the book value of the fixed tangible assets; for the rental price of capital, I use the one that is commonly used in literature and is based on Hall and Jorgenson (1967), i.e. $P_{Kit} = P_I(r - \pi + \delta_i)$, where P_I denotes price indices of investment in fixed assets at provincial level, r refers to the long-term official interest rates of loans of financial institute, π is the inflation rate at country level, and δ_i is the average depreciation rate at 2-digit industry level⁹.

To simplify the notation, I rewrite the left hand side of equation (14) as Δy_{it} and the right hand side as Δx_{it} . By adding a noise error term u_{it} , I obtain a very simple regression equation:

$$\Delta y_{it} = B_{it}\Delta x_{it} + u_{it} \quad (15)$$

To investigate the effect of agglomeration on firm's markups, I assume that markups depend on spatial agglomeration, regional, industrial and firm's characteristics, as well as year dummies:

$$B_{it} = \alpha + \beta Agg_{it}^{jr} + \gamma X_{it} + \tau_t$$

where Agg_{it}^{jr} denotes geographic concentration of industry j in region r at time t , β is the coefficient of interest. X_{it} is a vector of control variables. τ_t is a vector of year dummies. Substituting the above equation into equation (15), I get the empirical specification:

⁹The average depreciation rate is calculated from China Industry Economy Statistical Yearbook.

$$\Delta y_{it} = \alpha \Delta x_{it} + \beta Agg_{it}^{jr} \Delta x_{it} + \gamma X_{it} \Delta x_{it} + \tau_t \Delta x_{it} + u_{it} \quad (16)$$

The coefficient of the interaction term of agglomeration variable and Δx_{it} captures the effect of spatial agglomeration on price-cost margins. The empirical specification allows me to examine the impact on markups, but it does not allow me to separate the price competition effect from the effect of agglomeration externalities.

4. Data and Variables

4.1. Data

The data used in this paper are drawn from the Oriana CD-ROM (version 2007 January) provided by Bureau van Dijk, which contains financial company information for public and private companies in the Asia-Pacific regions. The companies included in the database are either publicly listed or satisfy at least one of the following size criteria: minimum number of employees is 150, or annual turnover (total assets) at least 10 million (20 million) USD¹⁰. The original dataset covers an unbalanced panel of 23,613 plants over the period 2002-2004. I restrict the data to the manufacturing sector, based on the US SIC 1987 classification (SIC 20-39). The number of observations is reduced to 11,512 plants due to missing values on some of the input factors (See Appendix A for a detailed description of the dataset and cleaning process).

Table 1 provides summary statistics for the main variables used. Firms have 1,146 employees, earn a revenue of 54.79 million dollars and face a labor cost of 2.38 million dollars on average. Around 6.0% of firms are from special economic zones. The average number of own-industry (3-digit SIC code) firms at county level (4-digit zip code) is 3.53, while the average number of own-industry firms at city level (3-digit zip code) is 7.84. The average sale of own industry firms (exclu. firm) is 260.79 million dollars at county level while it is 554.76 million dollars at city level.

Figure 3 shows the maps of the distribution of firms in textile and electrical and electronic equipment industries in China. Firms in both industries

¹⁰Because of the limitation of the dataset, I can only examine the effect of the number of large firms on markups of large firms in this paper.

Figure 3: Maps of the distribution of firms in two industries in China (2004): (a)Textile Industry and (b)Electrical and Electronic Equipment Industry.

Table 1: Descriptive Statistics of Main Variables

Variable	Observations	Mean	Standard Deviation
value-added	10,715	9075.40	45931.78
Turnover	13,377	54786.17	220603.8
Employment	13,374	1146.05	2692.23
Tangible fixed assets	13,377	1780.87	12451.14
Material costs	13,377	36918.56	163731.5
Wage bill	13,377	2378.05	9497.50
SEZ dummy	13,377	0.062	0.24
Private dummy	13,360	0.22	0.41
Foreign dummy	13,360	0.40	0.49
Collective dummy	13,360	0.074	0.26
Other dummy	13,360	0.19	0.39
No. of own industry firms at county level	13,377	3.53	5.65
No. of own industry firms at city level	13,377	7.84	11.49
Total sales of own industry firms at county level	13,377	260786.5	997119
Total sales of own industry firms at city level	13,377	554763.9	1927248

Note: Value-added, turnover, tangible fixed asset, material cost and wage bill are expressed in thousands of U.S. dollar.

are mainly concentrated in the coastal area, especially in the Pearl River Delta, Yangtze River Delta and in areas near Beijing and Tianjin.

Figure 4 shows the kernel probability density estimates of the price-cost margins calculated following Collins and Preston (1969) for the sample, which offers preliminary evidence that price-cost margins are lower in core region (above median number of firms at city level) than in periphery region (below the median number of firms at city level)¹¹. Two distributions are plotted which correspond to firms that are located in core region and periphery region. The distribution of core region firms has higher weight than the periphery region distribution at low PCM levels, but has less weight at high PCM levels. The average PCM in core region is 0.312, as compared to 0.290 in periphery region (a t-test for equality is rejected).

¹¹Price-cost margin is calculated by $\frac{\text{turnover} - \text{material cost} - \text{cost of employees}}{\text{turnover}}$.

Figure 4: PCM kernel estimates(2004): firms in core and periphery regions

4.2. Variables

For agglomeration variables, as in Henderson (2003), I constructed county level (4-digit zip code) measures of own industry (3-digit SIC) sales, the number of own industry firms. For a firm i in industry j in region r at time t , the agglomeration variable based on the number of firms is defined as:

$$\text{Agg}_{it}^{jr} = \ln(\text{Nr. of firms}_t^{jr})$$

The agglomeration variable capturing the total sale is:

$$\text{Agg}_{it}^{jr} = \ln(\text{sales}_t^{jr} - \text{sales}_{it}^{jr} + 1)$$

where own firm sale is subtracted to avoid possible endogeneity problems. And a constant 1 is added to ensure the existence of natural logarithm.

I use Herfindahl index to capture the industrial concentration, defined as follows:

$$\text{HHI} = \sum_{i \in j} \left(\frac{\text{sales}_{ijt}}{\text{sales}_{jt}} \right)^2$$

where sales_{ijt} is the sales for firm i in industry j at time t , and sales_{jt} is the total sales for industry j at time t . Finally, firm size is measured by the logarithm of sales, $\text{size}_{it} = \ln(\text{sales}_{it})$.

5. Results

5.1. Basic Results

I first look at the estimated sectoral price-cost margins to examine whether there is substantial variation in price-cost margins across sectors. Figure 5 shows the estimates of price-cost margins for each 2-digit SIC sectors in China. The estimated price-cost margins range from 0.10-0.28. And it is not surprising to find that Chemicals and Allied Products, Petroleum Refining, Tobacco among the high price-cost margin sectors.

Figure 5: Price-cost margin per SIC 2-digit sector

Next, I turn to the pooled data and test how the price-cost margins vary with geographic concentration by estimating equation (16). The results are reported in Table 2. Column (1) in Table 2 shows that the average price-cost margins for manufacturing firms in China is 0.20 suggesting that the average markup is 1.25. To examine the relation between agglomeration and price-cost margins, I start by including only the number of own-industry firms at county level (in addition to year dummies) in Column (2) in Table 2. Additional controls are added sequentially in Column (3) to (4). The estimated coefficient on the number of own-industry firms at county level is negative and statistically significant at the 5% level, and neither the point estimates nor the standard errors are much affected by the inclusion of the additional

controls. The result in Column (4) suggests that a rise in the number of own-industry firms at county level of 1 percent decreases the price-cost margins by 0.000089, which is 0.045 percent of the average. Geographic concentration of industries strengthens the price competition relative to the externalities. Column (5)-(7) in Table 2 apply the sales-based measure of agglomeration and the estimates indicate that the results are robust to different measures of agglomeration but the negative impact is larger by using the number of own-industry firms.

Table 2: OLS Estimates of Price-cost Margins and Agglomeration

Variables	Dependent Variable: Δy						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Δx	0.20*** (0.0043)	0.13*** (0.0058)	0.28*** (0.066)	0.25*** (0.068)	0.13*** (0.0063)	0.28*** (0.066)	0.25*** (0.068)
$\Delta x \times \ln(\text{no. of own industry firms in county})$		-0.0096** (0.0047)	-0.0098** (0.0047)	-0.0089* (0.0049)			
$\Delta x \times \ln(\text{own industry sales in county, outside own firm})$					-0.0017** (0.00073)	-0.0020*** (0.00073)	-0.0019* (0.00075)
$\Delta x \times \text{size}$			0.0048 (0.0033)	0.0053 (0.0033)		0.0054 (0.0033)	0.0059* (0.0033)
$\Delta x \times \text{HHI}$			-0.089 (0.067)	-0.085 (0.068)		-0.099 (0.068)	-0.097 (0.068)
$\Delta x \times \ln(\text{population})$			-0.023*** (0.0067)	-0.022*** (0.0069)		-0.023*** (0.0068)	-0.022*** (0.0069)
$\Delta x \times \text{private}$				0.021 (0.019)			0.022 (0.019)
$\Delta x \times \text{foreign}$				0.025 (0.019)			0.026 (0.019)
$\Delta x \times \text{collective}$				0.017 (0.021)			0.018 (0.021)
$\Delta x \times \text{other}$				0.016 (0.020)			0.017 (0.020)
$\Delta x \times \text{SEZ}$				-0.025 (0.0093)			-0.024 (0.016)
Year dummies	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	27,047	27,047	27,047	27,017	27,047	27,047	27,017
Firms	15,308	15,308	15,308	15,291	15,308	15,308	15,291
Counties	1,518	1,518	1,518	1,518	1,518	1,518	1,518
R-squared	0.20	0.26	0.28	0.28	0.26	0.28	0.28

Note: Robust standard errors adjusted for clustering at the 3-digit industry level and 4-digit zip code level are in parentheses. *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. The reference group is State-owned Enterprises. The agglomeration variables, firm size, population, Herfindahl index, ownership dummies and special economic zone dummy are also included separately as additional control variables. The estimates refer to equation (16).

In addition, I also find that price-cost margins in provinces with larger population are relatively lower than that in provinces with small popula-

tion. However, firm's ownership does not have significant effect on price-cost margins.

5.2. *IV Estimation*

The results reported in Table 2 could be biased due to potential simultaneity and omitted variable problems. For instance, markups and the number of firms in the market may be jointly determined in the equilibrium. Meanwhile, geographic concentration of industries and markups of firms could be affected by some factors which are not controlled for, such as regional policy.

I use IV method to deal with these potential problems. Following Ciccone and Hall (1996), I use original sources of agglomeration, such as the transportation infrastructures and the number of firms in a region in the initial phase, as instruments for agglomeration variables. In particular, the length of highways per square kilometer and the number of enterprises at provincial level in 1965 are used as instruments for the index of geographic concentration¹². The validity of instruments rests on the hypothesis that the transportation infrastructure and initial number of firms in 1965 are important historical factors in agglomeration in 2003 and 2004 but they do not have direct impact on firm's pricing behavior (Figure 6). In terms of the first requirement, because of the self-reinforcement effects or lock-in effects, history matters for economic geography, i.e., initial conditions are essential in the forming of agglomeration (Krugman, 1991a). The second requirement is discussed in more detail below.

The construction of highways in China before 1965 was mainly driven by central planning. The highways before 1949 in China were built mainly for military use and were seriously damaged during the second Sino-Japanese War and the War of Liberation. After 1949 when People's Republic China was established, the renovation of highways was carried out gradually. The total length of highways increased dramatically during 1953-1965 except the three years of Natural Disasters. So I suppose that the length of highways in the 1960s should not be correlated with the firm's pricing behavior 40 years later. Over the period of 1960-1965, the number of enterprises in coastal areas decreased because there was a driven to relocate key industrial production from coastal areas to interior provinces for potential wars with neighboring

¹²The data are from Comprehensive Statistical Data and Materials on 50 Years of New China compiled by Department of Comprehensive Statistics of National Bureau of Statistics.

Figure 6: Identification Strategy

countries and regions. Thus, the number of firms in 1965 can be treated to be exogenous. In sum, it seems that the length of highways per sq. km. and the number of enterprises at provincial level in 1965 are valid instruments.

The first-stage estimates of IV regressions are presented in Appendix D. The length of highways per sq. km. and the number of firms in 1965 have significantly positive effect on agglomeration. The results of the first-stage regressions provide little evidence that the results suffer from the presence of weak instruments. The F-statistic for both first-stage regressions is well above the threshold of 10 suggested by Staiger and Stock (1997).

Table 3 reports the results of IV estimation. In column (1), I use the number of firms at county level. In column (2), the number of firms at city level (3-digit zip code) is used. The coefficients of agglomeration variables in both cases are statistically significant and the magnitude is larger compared to the OLS estimates. It suggests that 1% increase of the number of firms at county level induces a decrease of price-cost margin by 0.00056, which is 0.28% of the average. As the average number of own-industry firms at county level is 3.53, increasing the number of firms from 3 to 4 leads to a decrease in the price-cost margin by 0.018 which is remarkable. Hansen J-statistics suggest that I cannot reject the null hypothesis, i.e., the instruments are uncorrelated with the error term. Column (3) and (4) show that the results are robust to different measures of agglomeration.

5.3. Agglomeration externalities in China

Using Chinese firm-level data I find a significantly negative impact of agglomeration on price-cost margins. Does it imply that there is no or weak agglomeration externalities in China? In this section I examine the effect

of spatial agglomeration on firm-level productivity by estimating production function.

Following the empirical strategy of Henderson (2003) and Martin et al. (2011), I apply equation (17) as a regression equation,

$$\ln y_{it} = \alpha \ln K_{it} + \beta \ln L_{it} + \delta \text{Agg}_{it}^{jr} + \gamma \text{urb}_{it}^{jr} + \lambda \text{div}_t^{jr} + \eta \text{comp}_t^{jr} + \tau_t + \delta_j + \epsilon_{it} \quad (17)$$

where y_{it} is the value added of firms deflated by a provincial producer's price index of industrial products, urb_{it}^{jr} is the number of workers in other industries working in the same region, div_t^{jr} is an indicator of the diversity that firms face in region r at time t and comp_t^{jr} captures the degree of competition a firm faces¹³.

The estimates of equation (17) are shown in Table 4. The results reveal that the number of own industry firms at county level has a significantly positive effect on productivity. It suggests that 1% increase in the number of own industry firms leads to 0.055% increase of productivity. The coefficients estimated by using IV method (see column(3) to column(4)) which deals with endogeneity problems, however, are much larger. The results in Table 4 show that agglomeration externalities do exist in manufacturing industries in China.

5.4. Firm Sales and Profits in Agglomerated Regions

In agglomerated regions, firms charge lower markups despite of having lower marginal costs, indicating that firms that are geographically concentrated have lower profitability. Then why are firms geographically concentrated? The theoretical model in Section 2 predicts that firms in agglomerated regions have larger output due to the market size effect and thus may have higher profits. In this section, I test how the firm sales and profits vary with geographic concentration.

¹³The diversity index is defined as $\text{div}_t^{jr} = \ln \left(\frac{1}{\sum_{s \neq j} \left(\frac{\text{employees}_t^{sr}}{\text{employees}_t^{jr}} \right)^2} \right)$, the competition index is defined as $\text{comp}_t^{jr} = \ln \left(\frac{1}{\sum_{i \in jr} \left(\frac{\text{employees}_t^{ir}}{\text{employees}_t^{jr}} \right)^2} \right)$.

The results shown in Table 5 suggests that firms are likely to have larger sales and profits in agglomerated regions. However, the effect of agglomeration on profits is insignificant using IV method.

6. Robustness Checks

The basic results are consistent with the prediction of the model. In this section, I conduct several checks to see whether these results are robust to various sensitivity checks.

6.1. Randomness of the Sample

Since the sample accounts for about 5% of total number of large- and medium-sized manufacturing firms in China, one concern is that the firms in the sample may not be randomly drawn from the population. Using the total number of state-owned and non-state-owned enterprises above designated size (SIC 2-digit) at provincial level from the China Industry Economy Statistical Yearbook, I am able to check whether the results are robust to alternative data sources¹⁴.

The correlation between the number of firms at provincial level based on the sample and that from yearbook is 0.88, suggesting an evidence for the representative sample. The estimates using census-based aggregate variables from the yearbook are shown in Table 6. The coefficients of the number of firms at provincial level are significantly negative in all columns, and the IV estimate is higher than the OLS estimates.

6.2. Selection Bias

The strengthened competition in agglomerated regions drives out relatively inefficient firms with low markups, which is likely to cause a downward bias in the estimation. In this section, I use the sub-sample of existing manufacturing firms across sample period to partially deal with the selection bias. The results are reported in Table 7, indicating that the coefficients of interest do not change much except the first column. The results provide an evidence that the selection bias may not be serious. And I will get a higher estimate of the coefficient of interest if the potential selection problem can be controlled for.

¹⁴This category includes all state owned enterprises (SOEs) and those not-state-owned firms with annual revenue above 5 million RMB (equivalent to around 700 thousand USD).

6.3. *Geographical distance-based measure of Agglomeration*

The measure of agglomeration above is based on the standard approach which define geography based on political boundaries, such as cities and counties. It is calculated with measurement error when firms locate near to the boundary of an administrative region. In this section, I explore a distance-based measure of agglomeration. By mapping zip codes with geographic coordinates (latitude and longitude), I am able to calculate the straight-line distances between each 6-digit zip code and construct circles of radius r around the geographic centroid of each zip code (There are 6,602 zip codes in my sample). The number of own-industry firms within a given circle is calculated as a measure of agglomeration. I consider two cases of radius, i.e., 10km and 20km. The correlation between the number of own-industry firms at county level and the number of firms in a region with a radius of 10 km is 0.83. Table 8 presents the results.

The first two columns in Table 8 report the estimates using a radius of 10km. And the last two columns present the results using a radius of 20km. The results reveal that the coefficients on agglomeration variables are similar as those using the measures based on administrative regions. Doubling the number of own-industry firms within 10km ring results in 0.055 decrease of price-cost margins, while doubling the number of own-industry firms within 20km ring results in 0.032 decrease of price-cost margins.

7. Conclusion

The impact of geographic concentration on markups of firms has been far less researched. In this paper, I investigate the effect of geographical agglomeration on the markups of manufacturing firms. Agglomeration can affect markups through the price competition effect and the agglomeration externalities. By modeling these channels, I demonstrate that the competition effect tends to dominate the effect of agglomeration externalities under reasonable parameter values, i.e., firms charge lower markups in more agglomerated regions. Using a unique firm level data from China, I find a significantly negative effect of agglomeration on price-cost margins and a significantly positive effect on firm-level productivity. The empirical results confirm that the price competition effect dominates the effect of agglomeration externalities in China. In addition, firms are likely to sell more products and have higher profit in agglomerated regions.

This study can be extended in several ways for future work. First, as the industrial agglomeration in China has increased consistently since the 1990s, it will be interesting to investigate the relationship between agglomeration and the evolution of markups over time when longitudinal firm-level data is available. Second, future work should be directed at having closer look at the impact of geographic concentration on markups for narrowly defined sectors with different degrees of geographic isolation, which can provide some explanations for the sectoral differences in the spatial distribution patterns.

Appendix A. Data description and cleaning process

The dataset used are from the Oriana database of Bureau Van Dijk (BvD). This is a commercial database of company accounts and includes information of the balance sheets and income statements of medium and large companies in the Asia-Pacific regions. For the purpose of this study, I retrieved detailed information on 23,613 plants for the People’s Republic of China. Firm level data often contains outlier observations that may bias the estimated coefficients. Hence, I carefully clean the original dataset to handle the missing observations and outliers. Several cleaning procedures are applied to the sample:

- I work with unconsolidated accounts only.
- I eliminated the observations that were based on irregular reports or unreasonable data values in the level of variables (such as negative material costs).
- I restricted the data to the manufacturing sectors based on US SIC 1987 classification (SIC 20-39).
- To make results less driven by the extreme values, I exclude observations with extreme values of Δy and Δx (<1st percentile and >99th percentile).

Appendix B. Variable description

- PQ = turnover in thousand dollar.
- $P_L L$ = cost of employees in thousand dollar.

- $P_M M$ = cost of materials in thousand dollar (total costs of the goods sold minus the cost of employees).
- K = tangible fixed assets.

Appendix C. Zipcodes and Regions

Regions are defined based on zip codes. Zip codes in the People's Republic of China are a numeric 6-digit system for the whole country. The first two digits show the province, province-equivalent municipality, or autonomous region. The third digit is the postal zone, the fourth digit refers to the prefectures or prefecture-level city, and the last two digits denote the delivery post office. Regions defined at 4-digit zip code level is equivalent to county level.

Appendix D. First Stage Results

References

- Bain, J. S., 1956. Barriers to New Competition. Cambridge, MA: Harvard University Press.
- Bresnahan, T., Reiss, P., 1991. Entry and competition in concentrated markets. *The Journal of Political Economy* 99 (5), 977–1009.
- Campbell, J., Hopenhayn, H., 2005. Market size matters. *The Journal of Industrial Economics* 53 (1), 1–25.
- Ciccone, A., Hall, R., 1996. Productivity and the density of economic activity. *The American Economic Review* 86 (1), 54–70.
- Collins, N., Preston, L., 1969. Price-cost margins and industry structure. *The Review of Economics and Statistics* 51 (3), 271–286.
- Combes, P., Duranton, G., Gobillon, L., Puga, D., Roux, S., 2009. The productivity advantages of large cities: Distinguishing agglomeration and firm selection, cEPR Discussion paper No. DP7191.
- Combes, P., Lafourcade, M., 2011. Competition, market access and economic geography: Structural estimation and predictions for France. *Regional Science and Urban Economics* forthcoming.

- Corcos, G., Gatto, M. D., Mion, G., Ottaviano, G., 2009. Productivity and firm selection: Quantifying the “new” gains from trade, sERC DP 28.
- Duranton, G., Overman, H., 2005. Testing for localization using micro-geographic data. *Review of Economic Studies* 72 (4), 1077–1106.
- Fujita, M., Krugman, P., Venables, A., 1999. *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press, Cambridge, MA.
- Gabszewicz, J., Thisse, J., Fujita, M., Schweizer, U., 1986. *Location Theory*. Harwood Academic Publisher GmbH, Ch. Spatial Competition and the Location of Firms, pp. 1–71.
- Hall, R., 1988. The relation between price and marginal cost in u.s. industry. *Journal of Political Economy* 96 (5), 921–947.
- Hall, R., 1990. Growth, Productivity, Unemployment: Essays to celebrate Bob Solow’s Birthday. Cambridge, MA: MIT Press, Ch. Invariance Properties of Solow’s Productivity Residual, p. 71112.
- Hall, R., Jorgenson, D., 1967. Tax policy and investment behavior. *American Economic Review* 57 (3), 638–655.
- Harrison, A. E., 1994. Productivity, imperfect competition and trade reform. *Journal of International Economics* 36 (1-2), 53–73.
- Helpman, E., 1998. *Topics in Public Economics*. Cambridge University Press, Cambridge, Ch. The Size of Regions, p. 3354.
- Henderson, J., 2003. Marshall’s scale economies. *Journal of Urban Economics* 53 (1), 1–28.
- Kaldor, N., 1935. Market imperfection and excess capacity. *Economica* 2 (5), 35–50.
- Kee, H., Hoekman, B., 2007. Imports, entry and competition law as market disciplines. *European Economic Review* 51 (4), 831–858.
- Konings, J., Cayseele, P. V., Warzynski, F., 2001. The dynamics of industrial mark-ups in two small open economies: Does national competition policy matter? *International Journal of Industrial Organization* 19 (5), 841–859.

- Konings, J., Cayseele, P. V., Warzynski, F., 2005. The effects of privatisation and competitive pressure on firm's price-cost margins: Micro evidence from emerging economies. *The Review of Economics and Statistics* 87 (1), 124–134.
- Konings, J., Vandenbussche, H., 2005. Antidumping protection and markups of domestic firms: Evidence from firm level data. *Journal of International Economics* 65 (1), 151–165.
- Krugman, P., 1991a. History versus expectations. *The Quarterly Journal of Economics* 106 (2), 651–667.
- Krugman, P., 1991b. Increasing returns and economic geography. *Journal of Political Economy* 99 (3), 483–499.
- Krugman, P., 2011. The new economic geography, now middle-aged. *Regional Studies*.
- Levinsohn, J., 1993. Testing the imports-as-market-discipline hypothesis. *Journal of International Economics* 35 (1-2), 1–22.
- Marshall, A., 1920. *Principles of Economics*. London: MacMillan.
- Martin, P., Mayer, T., Mayneris, F., 2011. Spatial concentration and plant-level productivity in france. *Journal of Urban Economics* 69 (2), 182–195.
- Melitz, M. J., Ottaviano, G. I., 2008. Market size, trade, and productivity. *Review of Economics Studies* 75 (1), 295–316.
- Melo, P., Graham, D., Noland, R., 2009. A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics* 39 (3), 332–342.
- Ottaviano, G. I., Tabuchi, T., Thisse, J.-F., 2002. Agglomeration and trade revisited. *International Economic Review* 43 (2), 409–436.
- Pinkse, J., Slade, M. E., Brett, C., 2002. Spatial price competition: A semi-parametric approach. *Econometrica* 70 (3), 1111–1153.
- Roeger, W., 1995. Can imperfect competition explain the difference between primal and dual productivity measures? estimates for u.s. manufacturing. *Journal of Political Economy* 103 (2), 316–330.

- Rosenthal, S. S., Strange, W. C., 2003. Geography, industrial organization, and agglomeration. *Review of Economics and Statistics* 85 (2), 377–393.
- Rosenthal, S. S., Strange, W. C., 2004. *Handbook of Regional and Urban Economics*. Amsterdam: North-Holland, Ch. Evidence on the nature and sources of agglomeration economies, pp. 2119–2171.
- Staiger, D., Stock, J., 1997. Instrumental variables regression with weak instruments. *Econometrica* 65 (3), 557–586.
- Syverson, C., 2004. Market structure and productivity: A concrete example. *Journal of Political Economy* 112 (6), 1181–1222.
- Syverson, C., 2007. Prices, spatial competition, and heterogeneous producers: An empirical test. *The Journal of Industrial Economics* 55 (2), 197–222.
- Venables, A. J., 1996. Equilibrium locations of vertically linked industries. *International Economic Review* 37 (2), 341–359.

Table 3: IV Estimates of Price-cost Margins and Agglomeration

	Dependent Variable: Δy			
	(1)	(2)	(3)	(4)
Δx	0.31*** (0.088)	0.34*** (0.087)	0.31*** (0.090)	0.38*** (0.094)
$\Delta x \times \ln(\text{no. of own industry firms in county})$	-0.056** (0.026)			
$\Delta x \times \ln(\text{no. of own industry firms in city})$		-0.039*** (0.013)		
$\Delta x \times \ln(\text{own industry sales in county, outside own firm})$			-0.011** (0.0043)	
$\Delta x \times \ln(\text{own industry sales in city, outside own firm})$				-0.011*** (0.0038)
$\Delta x \times \text{size}$	0.0071* (0.0042)	0.0071* (0.0039)	0.010** (0.0044)	0.010** (0.0042)
$\Delta x \times \text{HHI}$	-0.24** (0.11)	-0.27*** (0.11)	-0.28** (0.12)	-0.35*** (0.13)
$\Delta x \times \ln(\text{population})$	-0.028*** (0.0090)	-0.031*** (0.0089)	-0.030*** (0.0093)	-0.034*** (0.0094)
$\Delta x \times \text{private}$	0.041* (0.023)	0.041* (0.022)	0.045* (0.024)	0.044* (0.023)
$\Delta x \times \text{foreign}$	0.052** (0.025)	0.057** (0.024)	0.057** (0.025)	0.062** (0.025)
$\Delta x \times \text{collective}$	0.033 (0.025)	0.037 (0.024)	0.040 (0.026)	0.044* (0.026)
$\Delta x \times \text{other}$	0.030 (0.023)	0.032 (0.023)	0.034 (0.024)	0.037 (0.024)
$\Delta x \times \text{SEZ}$	0.012 (0.028)	-0.010 (0.019)	0.0084 (0.024)	-0.0068 (0.019)
Year Dummy	Yes	Yes	Yes	Yes
Observations	21,735	21,735	21,735	21,735
Hansen J statistic/p-value	0.13	0.42	0.23	0.45
Centered R-squared	0.26	0.26	0.26	0.26

Note: Robust standard errors adjusted for clustering at the 3-digit industry level and 4-digit zip code level are in parentheses. *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. The reference group is State-owned Enterprises. The agglomeration variables, firm size, population, Herfindahl index, ownership dummies and special economic zone dummy are also included separately as additional control variables. The estimates refer to equation (16).

Table 4: Estimates of Agglomeration Externalities

	Dependent Variable: value added			
	(1)OLS	(2)OLS	(3)IV	(4)IV
ln(no. of own industry firms in country)	0.055*** (0.012)		0.67*** (0.15)	
ln(no. of own industry firms in city)		0.063*** (0.0085)		0.33*** (0.057)
ln(employees)	0.39*** (0.011)	0.39*** (0.011)	0.36*** (0.015)	0.39*** (0.012)
ln(capital)	0.49*** (0.0072)	0.49*** (0.0072)	0.51*** (0.011)	0.49*** (0.0087)
Urbanization	-0.065*** (0.010)	-0.055*** (0.010)	-0.13*** (0.022)	-0.042*** (0.013)
Diversity	-0.019 (0.020)	-0.019 (0.019)	-0.0080 (0.042)	-0.024 (0.026)
Competition	-0.28*** (0.025)	-0.25*** (0.025)	-0.19*** (0.056)	-0.12*** (0.044)
Year fixed effect	yes	yes	yes	yes
Industry fixed effect	yes	yes	yes	yes
Observations	21,198	21,198	18,126	18,126
R-squared	0.66	0.66	0.54	0.62

Note: Robust standard errors adjusted for clustering at the 3-digit industry level and 4-digit zip code level are in parentheses. *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. The reference group is State-owned Enterprises. Centered R^2 is reported for IV estimation.

Table 5: Output and Profitability of Firms

	(1) ln(sales)	(2) profit	(3)IV ln(sales)	(4)IV profit
ln(no. of own industry firms in country)	0.079*** (0.016)	0.10** (0.045)	1.79* (0.94)	0.18 (1.88)
HHI	0.024 (0.57)	12.88* (6.87)	1.43 (1.09)	12.42 (13.58)
ln(population)	0.92*** (0.33)	1.85 (2.80)	0.83* (0.48)	1.87 (5.32)
private	-0.31*** (0.041)	-0.96 (0.82)	-0.34*** (0.073)	-1.59 (1.01)
foreign	-0.00028 (0.041)	0.49 (0.81)	-0.029 (0.064)	-0.16 (1.01)
collective	-0.31*** (0.051)	-1.49 (1.00)	-0.31*** (0.079)	-2.10 (1.29)
other	-0.27*** (0.041)	-0.66 (0.79)	-0.22*** (0.064)	-1.19 (1.12)
SEZ	-0.81*** (0.13)	0.37 (0.54)	-2.40** (0.96)	0.40 (3.52)
Year fixed effect	yes	yes	yes	yes
County fixed effects	yes	yes	yes	yes
Industry fixed effects	yes	yes	yes	yes
Observations	27,017	27,017	21,735	21,735
R-squared	0.29	0.04	.	0.05

Note: Robust standard errors adjusted for clustering at the 3-digit industry level and 4-digit zip code level are in parentheses. *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Profit is expressed in millions of U.S. dollar. The reference group is State-owned Enterprises.

Table 6: Robustness Checks I: based on yearbook data

	Dependent Variable: Δy			
	(1)	(2)	(3)	(4)IV
Δx	0.22*** (0.027)	0.29*** (0.079)	0.27*** (0.078)	0.31*** (0.094)
$\Delta x \times \ln(\text{no. of own industry firms at provincial level})$	-0.013*** (0.0039)	-0.010** (0.0049)	-0.012** (0.0049)	-0.033*** (0.010)
$\Delta x \times \text{size}$		-0.00075 (0.0037)	-0.00063 (0.0037)	-0.000073 (0.0042)
$\Delta x \times \text{HHI}$		-0.030 (0.13)	-0.037 (0.13)	-0.086 (0.16)
$\Delta x \times \ln(\text{population})$		-0.010 (0.010)	-0.0081 (0.010)	0.00074 (0.015)
$\Delta x \times \text{private}$			0.025 (0.019)	0.048** (0.023)
$\Delta x \times \text{foreign}$			0.037* (0.019)	0.062*** (0.023)
$\Delta x \times \text{collective}$			0.021 (0.022)	0.044* (0.026)
$\Delta x \times \text{other}$			0.021 (0.019)	0.038* (0.022)
$\Delta x \times \text{SEZ}$			-0.025* (0.014)	-0.015 (0.013)
Year dummy	Yes	Yes	Yes	Yes
Observations	22,218	22,218	22,188	17,567
R-squared	0.29	0.30	0.30	0.29

Note: Robust standard errors adjusted for clustering at the 3-digit industry level and 4-digit zip code level are in parentheses. *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. The reference group is State-owned Enterprises. The agglomeration variables, firm size, population, Herfindahl index, ownership dummies and special economic zone dummy are also included separately as additional control variables. The estimates refer to equation (16).

Table 7: Robustness Checks II: Existing Firms across Sample Period

	Dependent Variable: Δy			
	(1)	(2)IV	(3)	(4)IV
Δx	0.27*** (0.078)	0.48*** (0.12)	0.27*** (0.078)	0.48*** (0.12)
$\Delta x \times \ln(\text{no. of own industry firms in county})$	-0.0081 (0.0052)	-0.064** (0.027)		
$\Delta x \times \ln(\text{own industry sales in county, outside own firm})$			-0.0016** (0.00079)	-0.011** (0.0047)
$\Delta x \times \text{size}$	0.0044 (0.0035)	0.0077 (0.0047)	0.0048 (0.0035)	0.011** (0.0049)
$\Delta x \times \text{HHI}$	-0.086 (0.066)	-0.24** (0.098)	-0.093 (0.066)	-0.27** (0.11)
$\Delta x \times \ln(\text{population})$	-0.023*** (0.0077)	-0.049*** (0.013)	-0.023*** (0.0077)	-0.049*** (0.013)
$\Delta x \times \text{private}$	0.017 (0.020)	0.048* (0.027)	0.018 (0.021)	0.052* (0.029)
$\Delta x \times \text{foreign}$	0.010 (0.020)	0.049* (0.029)	0.011 (0.020)	0.051* (0.030)
$\Delta x \times \text{collective}$	0.0057 (0.022)	0.028 (0.028)	0.0064 (0.022)	0.035 (0.030)
$\Delta x \times \text{other}$	0.0065 (0.022)	0.031 (0.027)	0.0071 (0.022)	0.034 (0.028)
$\Delta x \times \text{SEZ}$	-0.0018 (0.017)	0.043 (0.030)	-0.0023 (0.017)	0.035 (0.027)
Year Dummy	Yes	Yes	Yes	Yes
Observations	23,452	17,066	23,452	17,066
Hansen J statistic /p-value	-	0.90	-	0.88
R-squared / Centered R-squared	0.30	0.28	0.30	0.27

Note: Robust standard errors adjusted for clustering at the 3-digit industry level and 4-digit zip code level are in parentheses. *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. The reference group is State-owned Enterprises. The agglomeration variables, firm size, population, Herfindahl index, ownership dummies and special economic zone dummy are also included separately as additional control variables. The estimates refer to equation (16).

Table 8: Robustness Checks III: Distance-based measure of agglomeration

	Dependent Variable: Δy			
	(1)	(2)IV	(3)	(4)IV
Δx	0.25*** (0.070)	0.34*** (0.093)	0.25*** (0.070)	0.33*** (0.088)
$\Delta x \times \ln(\text{no. of own industry firms within 10km ring})$	-0.0090* (0.0047)	-0.054** (0.021)		
$\Delta x \times \ln(\text{no. of own industry firms within 20km ring})$			-0.012*** (0.0041)	-0.033*** (0.012)
$\Delta x \times \text{size}$	0.0050 (0.0034)	0.0069* (0.0041)	0.0051 (0.0034)	0.0070* (0.0040)
$\Delta x \times \text{HHI}$	-0.071 (0.069)	-0.21** (0.10)	-0.097 (0.070)	-0.19** (0.090)
$\Delta x \times \ln(\text{population})$	-0.022*** (0.0070)	-0.032*** (0.0095)	-0.022*** (0.0070)	-0.031*** (0.0091)
$\Delta x \times \text{private}$	0.015 (0.020)	0.033 (0.023)	0.017 (0.020)	0.033 (0.023)
$\Delta x \times \text{foreign}$	0.022 (0.019)	0.048** (0.024)	0.027 (0.019)	0.048** (0.023)
$\Delta x \times \text{collective}$	0.012 (0.022)	0.022 (0.025)	0.014 (0.022)	0.025 (0.025)
$\Delta x \times \text{other}$	0.011 (0.021)	0.023 (0.023)	0.013 (0.021)	0.024 (0.023)
$\Delta x \times \text{SEZ}$	-0.027 (0.016)	-0.000082 (0.023)	-0.023 (0.017)	-0.0043 (0.020)
Year Dummy	Yes	Yes	Yes	Yes
Observations	26,276	21,156	26,276	21,156
Hansen J-statistic / p-value	-	0.20	-	0.31
R-squared / Centered R-squared	0.28	0.26	0.28	0.27

Note: Robust standard errors adjusted for clustering at the 3-digit industry level and 4-digit zip code level are in parentheses. *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. The reference group is State-owned Enterprises. The agglomeration variables, firm size, population, Herfindahl index, ownership dummies and special economic zone dummy are also included separately as additional control variables. The estimates refer to equation (16).

Table D.9: First-stage Estimates of IV Regression

	(1) ln(no. of own industry firms in county)	(2) $\Delta x \times$ ln(no. of own industry firms in county)	(3) ln(no. of own industry firms in city)	(4) $\Delta x \times$ ln(no. of own industry firms in city)
Instrumental variables				
highway per sq.km. 1965	-0.000019 (0.000043)	0.000020* (0.000012)	0.00037*** (0.000041)	.000031** (.000016)
nofirm 1965	0.000066*** (7.39e-06)	-1.64e-06 (2.06e-06)	0.00011*** (7.16e-06)	-2.80e-06 (2.52e-06)
Δx highway per sq.km. 1965	0.000041** (0.000019)	-0.000047 (0.000059)	0.000042* (0.000024)	0.00028*** (0.000069)
Δx nofirm 1965	-1.07e-06 (3.04e-06)	0.000074*** (0.00001)	-3.17e-06 (3.69e-06)	0.00012*** (0.000011)
Control variables				
size	0.022** (0.011)	-0.0024 (0.0047)	-0.0042 (0.011)	-0.0023 (0.0050)
HHI	-3.40*** (0.32)	-0.15 (0.21)	-5.78*** (0.42)	-0.34 (0.34)
ln(population)	-0.19*** (0.031)	-0.0021 (0.0088)	-0.28*** (0.031)	0.024** (0.011)
private	0.097** (0.038)	-0.021 (0.016)	0.21*** (0.040)	-0.019 (0.019)
foreign	0.21*** (0.034)	-0.024 (0.015)	0.44*** (0.036)	-0.013 (0.019)
collective	0.10** (0.041)	-.0074 (0.018)	0.24*** (0.047)	0.0049 (0.023)
other	0.047 (0.031)	-0.0065 (0.014)	0.19*** (0.035)	-0.0049 (0.018)
SEZ	0.62*** (0.12)	0.026 (0.034)	0.19 (0.12)	0.0097 (0.0331)
Δx	-0.070 (0.14)	1.15*** (0.40)	-0.26 (0.16)	2.60*** (0.45)
$\Delta x \times$ size	-0.0040 (0.0069)	0.022 (0.018)	0.00095 (0.0078)	0.0037 (0.019)
$\Delta x \times$ HHI	-0.17 (0.21)	-3.07*** (0.80)	-0.39 (0.34)	-4.92*** (1.29)
$\Delta x \times$ ln(population)	0.0091 (0.014)	-0.15*** (0.041)	0.031* (0.017)	-0.32*** (0.048)
$\Delta x \times$ private	-0.0065 (0.032)	0.16*** (0.062)	-0.025 (0.038)	0.25*** (0.074)
$\Delta x \times$ foreign	0.015 (0.031)	0.34*** (0.063)	0.021 (0.038)	0.56*** (0.076)
$\Delta x \times$ collective	0.035 (0.037)	0.20** (0.092)	0.049 (0.046)	0.33*** (0.11)
$\Delta x \times$ other	0.025 (0.031)	0.11** (0.058)	0.0063 (0.038)	0.23*** (0.072)
$\Delta x \times$ SEZ	0.029 (0.037)	0.55*** (0.14)	0.017 (0.039)	0.17 (0.14)
Year dummies	Yes	Yes	Yes	Yes
Observation	21,735	21,735	21,735	21,735
R-squared	0.18	0.53	0.28	0.70
F value	17.88	40.10	61.98	127.07

Note: Robust standard errors adjusted for clustering at the 3-digit industry level and 4-digit zip code level are in parentheses. *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Column (1) and (2) show the results of the first-stage regressions of IV estimation in column (1) in Table 3. Column (3) and (4) are the first-stage results of IV estimation in column (2) Table 3.