Centre for Economic Policy Research

[redacted]

The Pinhas Sapir Center for Development
Tel Aviv University

[redacted]

TEL AVIV UNIVERSITY
The Gordon Faculty of Social Sciences
Department of Public Policy

EIGHTH CEPR/JIE SCHOOL ON APPLIED INDUSTRIAL ORGANIZATION

Hosted by

The Eitan Berglas School of Economics, Recanati School of Business, Department of Public Policy at Tel Aviv University

Supported by

The Pinhas Sapir Center for Development, Tel Aviv University
Recanati School of Business, Tel Aviv University
Department of Public Policy, Tel Aviv University
Journal of Industrial Economics (JIE)
CEPR

Tel Aviv; 24 May 2011

Markups and Agglomeration: Price Competition vs. Externalities

Liqiu Zhao (Katholieke Universiteit Leuven)

Markups and Agglomeration: Price Competition versus Externalities

Liqiu Zhao

Vives Centre for Regional Economics
LICOS Centre for Institutions and Economic Performance
Katholieke Universiteit Leuven

First draft: March, 2010
This draft: May, 2011

Abstract

In the literature, agglomeration can affect markups through two potential channels: agglomerated regions toughen competition (price competition effect); firms are more productive on average in agglomerated regions (agglomeration economies and firm selection effect). It has not reached an conclusion that which force dominates in reality. 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 find that the price competition effect dominates the effect of agglomeration externalities, i.e., firms in more agglomerated regions have 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 endogenous problems using IV approach, I find that in China an increase of own-industry firms in a region has a negative causal effect on markups of firms, which suggests that the price competition effect dominates the effect of agglomeration externalities. I also find that the markups of high-tech firms are not significantly lower in more agglomerated regions.

Key words: Markups; Price Competition; Agglomeration Externalities
JEL Classification: L11, R12

The author would like to thank Joep Konings, Gianmarco Ottaviano, Damiaan Persyn, Stijn Vanormelingen, Linke Zhu, and seminar participants at the LICOS Centre for Institutions and Economic Performance. I gratefully acknowledge the financial support from the research council (LICOS excellence center) and VIVES. (e-mail: liqiu.zhao@econ.kuleuven.be)

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 barries to entry (Bain, 1956). Some economists are also interested in the impact of trade policies (Harrison, 1994; Konings and Vandenbussche, 2005) and competition laws (Konings et al., 2001; Kee and Hoekman, 2007) 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 Kaldor, 1935; Gabszewicz and Thisse, 1986; Pinkse et al., 2002). In other words, each firm is likely to compete more intensely with its neighbours than with distant firms due to transportation costs and search costs. Meanwhile, productivity gains steming from externalities are geographically bounded as well[1]. So it would be important and interesting to introduce spatial dimension. In this context, the markups 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 Rosenthal and Strange, 2004; Melo et al., 2009, for reviews and summaries of existing findings),

[1]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 samll scale, less than fifty kilometers.

which leads to lower marginal cost and higher markups (agglomeration externalities). Therefore the impact of geographical agglomeration on markups is ambiguous, depending on the relative magnitude of agglomeration externalities (selection effect) and competition effect [2].

In this paper, I model these three channels by extending the model of Melitz and Ottaviano (2008) which incorporate endogenous markups to allow for the externalities between firms within a region. Through the simulations of the model, 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 significant negative effect of agglomeration on price-cost margins applying the approach of Roeger (1995). The empirical results confirm that the competition effect dominates agglomeration externalities in China.

The contributions of this paper are three-folded. First, to the best of my knowledge this is the first study on agglomeration and markups using Chinese data. Second, I demonstrate that, although theoretically the impact of agglomeration on markups is ambiguous, the competition effect tends to dominate the effect of agglomeration externalities in reality. Finally, I introduce a more precise measure of agglomeration which is geographical distance-based.

The research is related with a number of studies in the literature. The "new economic geography" produces models involving both monopolistic competition and increasing returns (Krugman, 1991b; Venables, 1996; Helpman, 1998), summarized in Fujita et al. (1999). However, in their model monopolistic competition à la Dixit-Stiglitz framework ignores strategic interactions between firms. Melitz and Ottaviano (2008) apply a linear model of monopolistic competition which introduces weak strategic interactions and find that bigger market induces tougher competition, leading to higher average productivity and lower average markups[3], but they do not take into account of the agglomeration externalities. Empirically, using cross section data of five different service and retail industries, Bresnahan and Reiss (1991) find that the entry of additional firms into a local oligopoly market compresses the average markups of all firms in operation. Similarly,

[2] Distinguishing agglomeration externalities from selection effect 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.

[3] Markup in their paper is defined as price minus marginal cost, but if I apply a traditional definition of markup, i.e., price over marginal cost, as shown in this paper later, the average markups will remain unchanged.

Campbell and Hopenhayn (2005) provide further evidences that firms in larger markets have lower markups but larger output basing on the data of 13 retail industries. Those two studies use data from service industries which are more geographically isolated, thus the spatial competition is probably stronger in service industries than in manufacturing industries suggested by Syverson (2007). Using the data from U.S. ready-mixed concrete plants, Syverson (2004, 2007) examines the impact of plant density on price and productivity respectively. He finds that markets with higher plant densities have higher lower-bound and average productivity but lower upper-bound and average prices. These two studies suggest that the change of markups defined as price over marginal cost is ambiguous as plant density increases. It is of policy interest to identify the impact of firm density on markups. My study is the first try to probe into it. The most relevent study is Combes and Lafourcade (2011), which applies simulations based on the parameters estimated from French data. They find that marginal costs and prices both increase from the core to the periphery[4]. In the central area, the effect of marginal cost dominates price competition, however in peripheral areas competition effect dominates the effect of marginal cost. So the markups are greater in both core and peripheral regions but lower inbetween which should produce further concentration in the core region. So the outcome in their paper is not based on the long-run equilibrium. In this paper I also shed light on these two mechanisms.

The reminder 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 generate any procompetitive effect, i.e., the markup is independent of

[4]In their paper, marginal cost increases from core to periphery because intermediate input prices are lower in more competitive regions, namely core regions.

the number of firms in the market. Ottaviano et al. (2002) introduce a linear model into economic geography, which incorporates endogenous markups and integrates weak interactions between firms. The model extends the work of Melitz and Ottaviano (2008) 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:

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

where q_0^c is the quantity of numeraire good consumed, q_i^c is the quantity of variety i. 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:

$$P_i = \alpha - \gamma q_i^c - \eta Q^c \tag{2}$$

where Q^c is the consumption level over all varieties $Q^c = \int_{i \in \Omega} q_i^c \, di$. Integrating the inverse demand equation, I have

$$\int_{i \in \Omega} P_i \, di = \alpha N - \gamma Q^c - \eta N Q^c = N\overline{P}$$

where $\overline{P}$ is the average price. I obtain the expression of Q^c as a function of N and $\overline{P}$. Plugging it into the inverse demand equation, I get the linear demand system:

$$q_i = \frac{\alpha L}{\eta N + \gamma} - \frac{L}{\gamma} P_i + \frac{\eta N}{\eta N + \gamma} \frac{L}{\gamma} \overline{P} \tag{3}$$

For any positive demand $q_i \geq 0$,

$$P_{max} = \frac{1}{\eta N + \gamma} \left(\gamma \alpha + \eta N \overline{P} \right) \equiv C_D \tag{4}$$

where P_{max} is the price at which demand is driven to 0. Given linear demand system, price elasticity of demand is given by $e_i = \left[\frac{C_D}{P_i} \quad 1\right]^{\ 1}$, varying with firm specific productivity, average price and the number of firms 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 modelling the total cost TC_i to allow for the externalities in the following way[5]:

$$TC_i = \frac{c_i}{N^{\lambda}} q_i \tag{5}$$

where c_i is firm specific productivity [6]. 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 of number of firms in a region by 1% increases the output by 0.05%. Here I assume $0 \quad \lambda < 1$. 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

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

[5]Total cost function can be derived from Cobb-Douglas production function $F(l, N) = AN^{\lambda} l$ given exogenous wage.
[6]Low c_i implies high productivity.

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

The optimal price is determined by the cut-off C_D in the market, firm specific cost c_i and the number of firms in the market N. As expected, equation (6) shows that lower cost firms charge lower prices, however, they charge higher price-cost margins[7]. 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 paremeter k, then the distribution of cost draws c is given by

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

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,

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

Then the cut-off level C_D is determined by

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

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 $\overline{P} =$

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

$\left(C_D + \frac{\bar{c}}{N^\lambda}\right)/2$ into equation (4), I get another relationship between C_D and N:

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

Figure 1 depicts equation (10) and (11) under reasonable parameters[8]. 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 equation (10) shifts downwards from short dashed line to long short dashed line. The cut-off level decreases, which suggests 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 about here.]

Figure 2 shows the markups as a function of the number of firms in two different cases given other paremeters. 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 to a dramatic decrease of marginal cost. If this effect outweights the spatial competition effect, the price-cost margin of firm i that is $1 \quad 2c_i/\left(C_D N^\lambda + c_i\right)$ 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 from roughly 0.04-0.10 (Rosenthal and Strange, 2004), depending on the sector and details of the estimation

[8] I 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 **?**, 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 curve (10) and curve (11) (see Figure 1) is the equilibrium.

procedure. Therefore, I would expect that the impact 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 about here.]

3 Estimation of Price-cost Margins

The methodology for estimating markups is based on Roeger (1995), which is an extention 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), de Loecker and Warzynski (2009), etc. I give a brief summary of the approach below, for detailed derivatinos I refer the reader to Roeger (1995).

Consider a linear homogeneous production function which also allows for the externalities between firms $Q_{it} = A_{it} N_{it}^{jr^{\lambda}} F(K_{it}, L_{it}, M_{it})$, where i is the firm index, t is 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}^{jr} is the number of own-industry firms in region r. Under imperfect competition, Hall (1988) shows that the Solow residual can be decomposed into an imperfect competition term and a productivity term[9].

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

with $\alpha_{Jit} = P_{Jt}J_{it}/P_{it}Q_{it}(J = K, L, M)$ is the cost share of inputs in turnover. The coefficient $B_{it} \quad (P_{it} \quad MC_{it})/P_{it}$ is the price-cost margin, which is directly related to the markups via $\mu_{it} = 1/(1 \quad B_{it})$. Δq_{it}, Δk_{it}, Δl_t and Δm_{it} are growth rates of output, capital, employment and material input, respectively. Δn_{it}^{jr} is the growth rate of the number of firms in region r. To estimate

[9]As the externalities are allowed for, the Solow residual also depends on the growth of the number of firms in the market

B_{it}, instrumental variables are required because of the potential correlation between unobserved productivity shocks and input factors. 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. Using the properties of the dual Solow residual under imperfect competition, Roeger derives a similar expression for the price based, or dual, Solow residual:

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

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 substracting (13) from (12), I get

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

Equation (14) can easily be estimated with firm level data since all bracketed terms can be observed from the data. Equation (14) also shows that in order to obtain an estimate of the price-cost margin, I need the information on growth rate of turnover, growth in cost of employees and growth in material costs, which are computable in my dataset. I also need the information on fixed tangible assets and the price of capital. The rental price for capital is constructed based on Hall and Jorgenson (1967) and Hsieh (2002), $P_{Kit} = P_I(r - \pi + \delta_i)$, 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 level.

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

$$\Delta y_{it} = B_{it}\Delta x_{it} + u_{it} \tag{15}$$

To investigate the effect of agglomeration on firms' 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 have

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

I investigate the impact of spatial agglomeration on price-cost margins by estimating equation (16), controlling for individual terms as well.

4 Data and Variables

4.1 Data

The data used in this paper are drawn from the Oriana CD-ROM (2007 January version) 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) USD[10]. 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 about here.]

[10]Because of the limitation of the dataset, I can only examine the effect of the number of large firms on markups of large firms.

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 about here.]

Figure 3 shows the maps of the distribution of firms in two industries in China. Firms in both industries 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 about here.]

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, [illegible]

as:

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

There are totally 1,542 counties in the sample. The agglomeration variable capturing the total sale is:

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

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

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

$$HHI = \sum_{i2j} \left(\frac{\text{sales}_{ijt}}{\text{sales}_{jt}} \right)^2$$

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. Firm size is measured by log of sales, $\text{size}_{it} = ln(\text{sales}_{it})$.

5 Basic Empirical Results

5.1 OLS

The basic results applying equation (16) are shown 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 would decrease the

price-cost margins by 0.000089, which is 0.045 percent of the average. Geographical concentration of industries strengthens the competition relative to the externalities, however, the magnitude is small. Column (5)-(7) in Table 2 apply the sales-based measure of agglomeration and the results show that the negative impact is larger by using the number of own-industry firms. The results are robust to different measures of agglomeration.

[Table 2 about here.]

In addition, I also find price-cost margins in provinces with larger population are relatively lower than that in provinces with small population. However, firm's ownership does not have significant effect on price-cost margins.

The magnitude of the effects of price competition and agglomeration economies depends on the product differentiation and the external accumulated local knowledge which varies across industries. Thus, I examine whether the markups of high-tech manufacturing firms are less affected by agglomeration, because of the weaker price competition or the stronger agglomeration economies. For the definition of high-tech manufacturing industries, I follow Henderson (2003) and define them as computers (357), communications (366), electronic components (367), aircraft (372), missiles and space vehicles (386), search and navigation equipment (381), measuring devices (382), and medical instruments (384). The results shown in Table 3 indicate that high-tech manufacturing firms charge higher markups than the other firms and the agglomeration does not have significantly negative effect on the markups for high-tech firms.

[Table 3 about here.]

5.2 IV Estimation

The results reported in Table 2 could be biased due to the simultaneity problem and the omitted variable problem. For instance, markups and the number of firms in the market may be jointly determined in the equilibrium. Due to the self-selection problem, firms in the agglomerated market have relatively higher unobserved productivity and thus have higher markups, which leads to a downward bias in the estimate of β.

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 this section, the length of highways and the number of enterprises at provincial level in 1965 are used as instruments for the index of geographical concentration[12]. 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 today but they do not have direct impact on firm's pricing behavior (Figure 5). 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.

[Figure 5 about here.]

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 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 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 and the number of enterprises at provincial level in 1965 are valid instruments.

[Table 4 about here.]

The first-stage estimates of IV regressions are presented in Table 4. The length of highways

[12]The 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.

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 5 about here.]

Table 5 reports the results of IV estimation. In column (1), I use the number of firms at county level. In column (2), I use the number of firms at city level (3-digit zip code). The coefficients of agglomeration variables in both cases are statistically significant and the magnitude is larger comparing 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 implies 33% increase and thus leads to the decrease of price-cost margin by 0.018 which is remarkable. Hansen J-statistics suggest that I can not reject the null hypothesis, i.e., the instruments are uncorrelated with the error term. The results in column (3) and (4) show that the conclusion still holds by using an alternative agglomeration variable which is sales-based.

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 specifications.

6.1 Randomness of the Sample

The sample accounts for about 5% of total number of large- and medium-sized manufacturing firms in China. The firms in the sample may not be randomly drawn from 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 Year Book, I am able to check

whether the results are robust to alternative data sources [13].

The correlation between the number of firms at provincial level based on the sample and that from year book is 0.878, 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.

[Table 6 about here.]

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 of the results. In this section, I use the information of existing manufacturing firms across sample period to partially deal with the selection problem. The results are reported in Table 7, suggesting that the coefficients of interest do not change much except the first column.

[Table 7 about here.]

Though I cannot solve the selection problem fully, I show that I may underestimate the impact of agglomeration on markups if there is a selection problem.

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 boundries, such as cities and counties. It is calculated with measurement error when firms locate near to the boundary of an administrative region. And the areas of counties widely differ across regions which also leads to the measurement error. 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

[13] This 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).

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 0.828. Table 8 presents the results.

[Table 8 about here.]

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 would result in 0.055 decrease of price-cost margins, while doubling the number of own-industry firms within 20km ring would result in 0.032 decrease of price-cost margins.

7 Conclusion

In this paper, I investigate the effect of geographical agglomeration on the markups of manufacturing firms. Agglomeration can affect markups through the competition effect and the agglomeration externalities. By modelling 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. The empirical results confirm that the price competition effect dominates the effect of agglomeration externalities in China. In addition, I find that the markups of high-tech manufacturing firms are not significantly affected by agglomeration which may be due to the higher product differenciation or higher degree of externalities.

The results have several policy implications in a context in which cluster policy are popular for local governments and the economy is emerging. This paper provides some evidences that

agglomeration on one hand promotes productivity and on the other hand strengthens competition though the effects are moderate, which both increase the allocative efficiency of resources. It gives support for the establishments of industrial parks to encourage firms to concentrate in a few locations in China. And it also suggests that other measures, like removing interregional trade barries and impoving competition rules, are important in China.

A Data description and cleaning process

The dataset used are from the Oriana database of Bureau Van Dijk (BvD). This is a commerial database of company accounts and includes information of the balcance sheests 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).

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.

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.

References

Joe S. Bain. *Barries to New Competition*. Cambridge, MA: Harvard University Press, 1956.

T.F. Bresnahan and P.C. Reiss. Entry and compeition in concentrated markets. *The Journal of Political Economy*, 99(5):977–1009, Oct. 1991.

J.R. Campbell and H.A. Hopenhayn. Market size matters. *The Journal of Industrial Economics*, 53 (1):1–25, Mar. 2005.

A. Ciccone and R.E. Hall. Productivity and the density of economic activity. *The American Economic Review*, 86(1):54–70, March 1996.

N.R. Collins and L.E. Preston. Price-cost margins and industry structure. *The Review of Economics and Statistics*, 51(3):271–286, Aug. 1969.

P. Combes and M. Lafourcade. Competition, market access and economic geography: Structural estimation and predictions for france. *Regional Science and Urban Economics*, forthcoming, 2011.

P.P. Combes, G. Duranton, L. Gobillon, D. Puga, and S. Roux. The productivity advantages of large cities: Distinguishing agglomeration and firm selection. CEPR Discussion paper No. DP7191, March 2009.

J. de Loecker and F. Warzynski. Markups and firm-level export status. NBER Working Paper No. 15198, July 2009.

G. Duranton and H.G. Overman. Testing for localization using micro-geographic data. *Review of Economic Studies*, 72(4):1077–1106, 2005.

M. Fujita, P. Krugman, and A.J. Venables. *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press, Cambridge, MA, 1999.

J.J. Gabszewicz and J.F. Thisse. *Location Theory*, chapter Spatial Competition and the Location of Firms, pages 1–71. Harwood Academic, Chur, 1986.

R.E. Hall. The relation between price and marginal cost in u.s. industry. *Journal of Political Economy*, 96(5):921–947, Dec. 1988.

R.E. Hall. *Growth, Productivity, Unemployment: Essays to celebrate Bob Solow's Birthday*, chapter Invariance Properties of Solow's Productivity Residual, page 71112. Cambridge, MA: MIT Press, 1990.

R.E. Hall and D. Jorgenson. Tax policy and investment behavior. *American Economic Review*, 57(3): 638–655, 1967.

Ann E. Harrison. Productivity, imperfect competition and trade reform. *Journal of International Economics*, 36(1-2):53–73, February 1994.

E. Helpman. *Topics in Public Economics*, chapter The Size of Regions, page 3354. Cambridge University Press, Cambridge, 1998.

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

C.T. Hsieh. What explains the industrial revolution in east asia? evidence from the factor markets. *American Economic Review*, 2002.

N. Kaldor. Market imperfection and excess capacity. *Economica*, 2(5):35–50, Feb. 1935.

H.L. Kee and B. Hoekman. Imports, entry and competition law as market disciplines. *European Economic Review*, 51(4):831–858, May 2007.

J. Konings and H. Vandenbussche. Antidumping protection and markups of domestic firms: Evidence from firm level data. *Journal of International Economics*, 65(1):151–165, January 2005.

J. Konings, Van Cayseele P., and Warzynski F. 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, April 2001.

J. Konings, P. Van Cayseele, and F. Warzynski. 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, 2005.

P. Krugman. History versus expectations. *The Quarterly Journal of Economics*, 106(2):651–667, May 1991a.

P. Krugman. Increasing returns and economic geography. *Journal of Political Economy*, 99(3):483–499, Jun. 1991b.

J. Levinsohn. Testing the imports-as-market-discipline hypothesis. *Journal of International Economics*, 35(1-2):1–22, August 1993.

A. Marshall. *Principles of Economics*. London: MacMillan, 1920.

M.J. Melitz and G.I.P. Ottaviano. Market size, trade, and productivity. *Review of Economics Studies*, 75(1):295–316, 2008.

P.C. Melo, D.J. Graham, and R.B. Noland. A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics*, 39(3):332–342, 2009.

G.I.P. Ottaviano, T. Tabuchi, and J.-F. Thisse. Agglomeration and trade revisited. *International Economic Review*, 43(2):409–436, May 2002.

J. Pinkse, Margaret E. Slade, and Craig Brett. Spatial price competition: A semiparametric approach. *Econometrica*, 70(3):1111–1153, May 2002.

W. Roeger. 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, Apr. 1995.

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

S.S. Rosenthal and W.C. Strange. *Handbook of Regional and Urban Rconomics*, chapter Evidence on the nature and sources of agglomeration economies, pages 2119–2171. Amsterdam: North-Holland, 2004.

D. Staiger and J.H. Stock. Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586, May 1997.

C. Syverson. Market structure and productivity: A concrete example. *Journal of Political Economy*, 112(6):1181–1222, Dec. 2004.

C. Syverson. Prices, spatial competition, and heterogeneous producers: An empirical test. *The Journal of Industrial Economics*, 55(2):197–222, Jun. 2007.

A. J. Venables. Equilibrium locations of vertically linked industries. *International Economic Review*, 37(2):341–359, May 1996.

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.

Table 2: OLS Estimates of Price-cost Margins and Agglomeration: 2002-2004

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

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.

Table 3: Are High-tech industries different?

	High-tech			Non high-tech		
	(1)	(2)	(3)	(4)	(5)	(6)
Δx	0.22	0.67	0.60	0.20	0.25	0.22
	(0.016)	(0.20)	(0.22)	(0.0045)	(0.070)	(0.072)
Δx ln(no. of own industry firms in county)		-0.013	-0.010		-0.013	-0.012
		(0.011)	(0.013)		(0.0056)	(0.0057)
Δx size		-0.012	-0.012		0.0060	0.0066
		(0.010)	(0.011)		(0.0035)	(0.0035)
Δx HHI		-0.82	-0.80		-0.069	-0.065
		(0.70)	(0.74)		(0.068)	(0.069)
Δx ln(population)		-0.043	-0.044		-0.021	-0.020
		(0.021)	(0.022)		(0.0070)	(0.0071)
Δx private			0.15			0.020
			(0.073)			(0.020)
Δx foreign			0.064			0.023
			(0.072)			(0.020)
Δx collective			0.0057			0.019
			(0.084)			(0.022)
Δx other			0.051			0.016
			(0.078)			(0.021)
Δx SEZ			-0.020			-0.026
			(0.034)			(0.019)
Year Dummy	No	Yes	Yes	No	Yes	Yes
Observations	2,230	2,230	2,219	24,817	24,817	24,798
R-squared	0.21	0.29	0.30	0.20	0.28	0.28
Firms	1,252	1,252	1,246	14,056	14,056	14,045

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.

Table 4: First-stage Estimates of IV Regression

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

Table 5: IV Estimates of Price-cost Margins and Agglomeration

	(1)	(2)	(3)	(4)
Δx	0.31	0.34	0.31	0.38
	(0.088)	(0.087)	(0.090)	(0.094)
Δx ln(no. of own industry firms in county)	-0.056			
	(0.026)			
Δx ln(no. of own industry firms in city)		-0.039		
		(0.013)		
Δx ln(own industry sales in county)			-0.011	
			(0.0043)	
Δx ln(own industry sales in city)				-0.011
				(0.0038)
Δx size	0.0071	0.0071	0.010	0.010
	(0.0042)	(0.0039)	(0.0044)	(0.0042)
Δx HHI	-0.24	-0.27	-0.28	-0.35
	(0.11)	(0.11)	(0.12)	(0.13)
Δx ln(population)	-0.028	-0.031	-0.030	-0.034
	(0.0090)	(0.0089)	(0.0093)	(0.0094)
Δx private	0.041	0.041	0.045	0.044
	(0.023)	(0.022)	(0.024)	(0.023)
Δx foreign	0.052	0.057	0.057	0.062
	(0.025)	(0.024)	(0.025)	(0.025)
Δx collective	0.033	0.037	0.040	0.044
	(0.025)	(0.024)	(0.026)	(0.026)
Δx other	0.030	0.032	0.034	0.037
	(0.023)	(0.023)	(0.024)	(0.024)
Δx SEZ	0.012	-0.010	0.0084	-0.0068
	(0.028)	(0.019)	(0.024)	(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.

Table 6: Robustness Checks: based on yearbook data

	(1)	(2)	(3)	(4)IV
Δx	0.22	0.29	0.27	0.31
	(0.027)	(0.079)	(0.078)	(0.094)
Δx ln(no. of own industry firms at provincial level)	-0.013	-0.010	-0.012	-0.033
	(0.0039)	(0.0049)	(0.0049)	(0.010)
Δx size		-0.00075	-0.00063	-0.000073
		(0.0037)	(0.0037)	(0.0042)
Δx HHI		-0.030	-0.037	-0.086
		(0.13)	(0.13)	(0.16)
Δx ln(population)		-0.010	-0.0081	0.00074
		(0.010)	(0.010)	(0.015)
Δx private			0.025	0.048
			(0.019)	(0.023)
Δx foreign			0.037	0.062
			(0.019)	(0.023)
Δx collective			0.021	0.044
			(0.022)	(0.026)
Δx other			0.021	0.038
			(0.019)	(0.022)
Δx SEZ			-0.025	-0.015
			(0.014)	(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.

Table 7: Robustness Checks: Existing Firms across Sample Period

	(1)	(2)IV	(3)	(4)IV
Δx	0.27	0.48	0.27	0.48
	(0.078)	(0.12)	(0.078)	(0.12)
Δx ln(no. of own industry firms in county)	-0.0081	-0.064		
	(0.0052)	(0.027)		
Δx ln(own industry sales in county, outside own firm)			-0.0016	-0.011
			(0.00079)	(0.0047)
Δx size	0.0044	0.0077	0.0048	0.011
	(0.0035)	(0.0047)	(0.0035)	(0.0049)
Δx HHI	-0.086	-0.24	-0.093	-0.27
	(0.066)	(0.098)	(0.066)	(0.11)
Δx ln(population)	-0.023	-0.049	-0.023	-0.049
	(0.0077)	(0.013)	(0.0077)	(0.013)
Δx private	0.017	0.048	0.018	0.052
	(0.020)	(0.027)	(0.021)	(0.029)
Δx foreign	0.010	0.049	0.011	0.051
	(0.020)	(0.029)	(0.020)	(0.030)
Δx collective	0.0057	0.028	0.0064	0.035
	(0.022)	(0.028)	(0.022)	(0.030)
Δx other	0.0065	0.031	0.0071	0.034
	(0.022)	(0.027)	(0.022)	(0.028)
Δx SEZ	-0.0018	0.043	-0.0023	0.035
	(0.017)	(0.030)	(0.017)	(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.

Table 8: Distance-based measure of agglomeration

		(1)	(2)IV	(3)	(4)IV
Δx		0.25	0.34	0.25	0.33
		(0.070)	(0.093)	(0.070)	(0.088)
Δx	lnnofirm10km	-0.0090	-0.054		
		(0.0047)	(0.021)		
Δx	lnnofirm20km			-0.012	-0.033
				(0.0041)	(0.012)
Δx	size	0.0050	0.0069	0.0051	0.0070
		(0.0034)	(0.0041)	(0.0034)	(0.0040)
Δx	HHI	-0.071	-0.21	-0.097	-0.19
		(0.069)	(0.10)	(0.070)	(0.090)
Δx	ln(population)	-0.022	-0.032	-0.022	-0.031
		(0.0070)	(0.0095)	(0.0070)	(0.0091)
Δx	private	0.015	0.033	0.017	0.033
		(0.020)	(0.023)	(0.020)	(0.023)
Δx	foreign	0.022	0.048	0.027	0.048
		(0.019)	(0.024)	(0.019)	(0.023)
Δx	collective	0.012	0.022	0.014	0.025
		(0.022)	(0.025)	(0.022)	(0.025)
Δx	other	0.011	0.023	0.013	0.024
		(0.021)	(0.023)	(0.021)	(0.023)
Δx	SEZ	-0.027	-0.000082	-0.023	-0.0043
		(0.016)	(0.023)	(0.017)	(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.

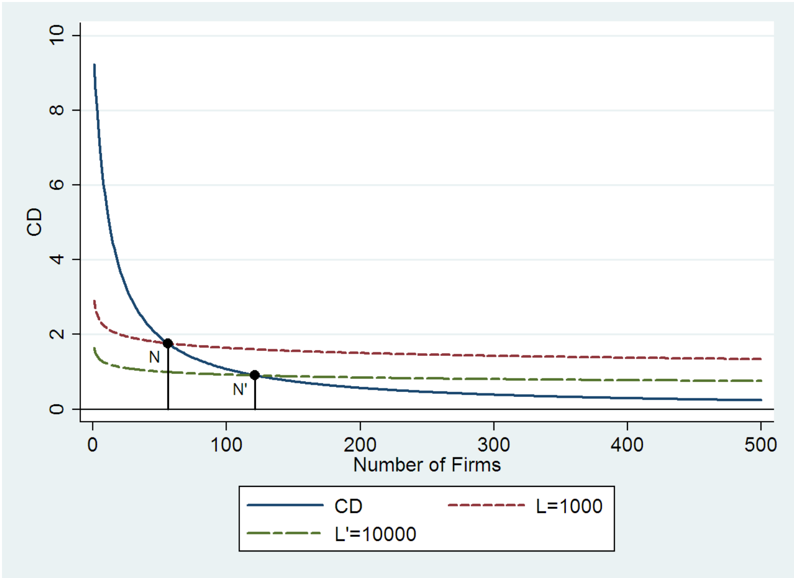

Figure 1: Market Size, Cut-off, and Equilibrium Number of Firms

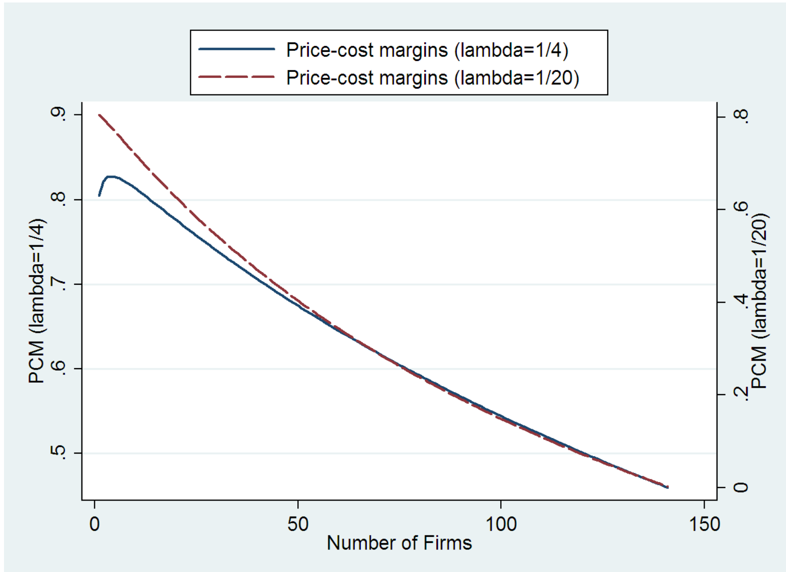

Figure 2: Price-cost Margins and Number of Firms: $\lambda = 1/4$ and $\lambda = 1/20$

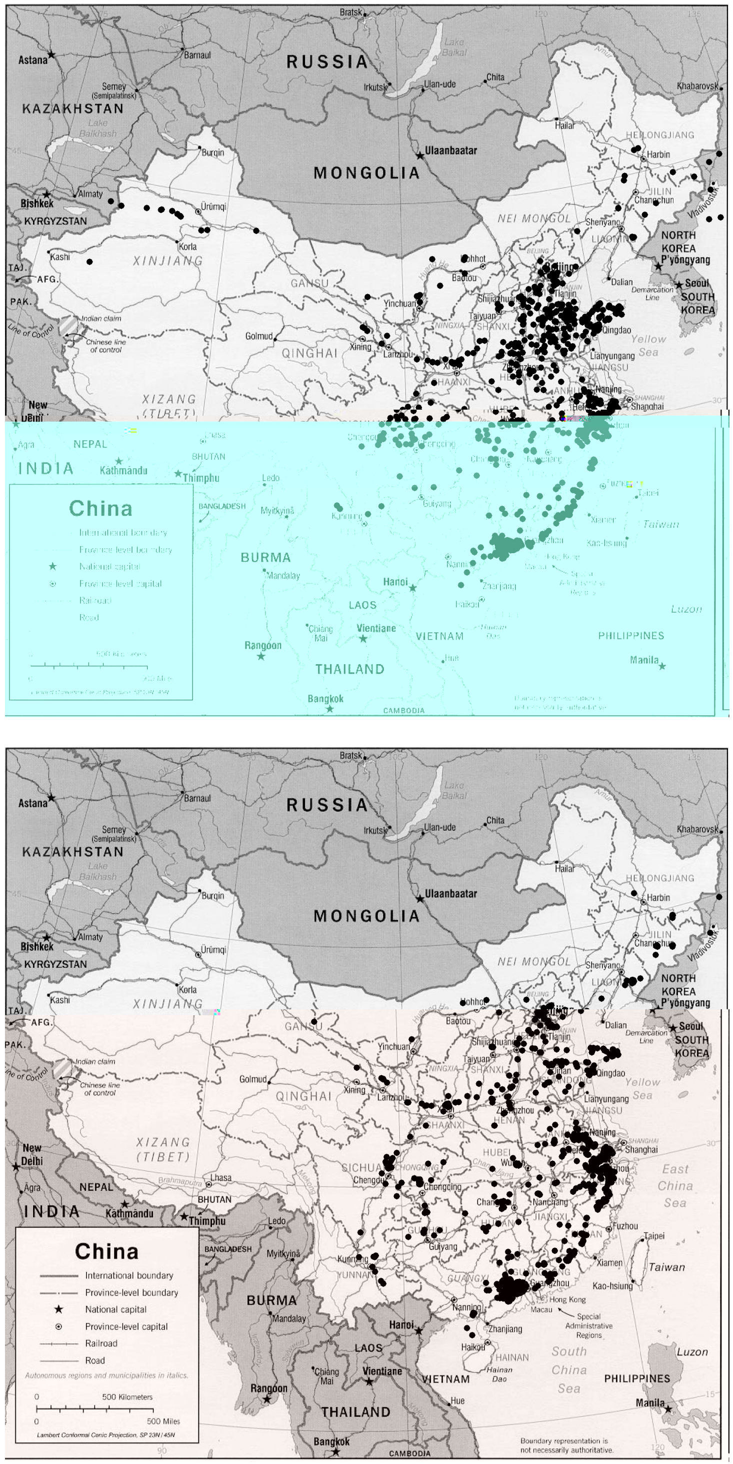

Figure 3: Maps of the distribution of plants in two industries in China (2004): (a)Textile Industry and (b)Electrical and Electronic Equipment Industry.

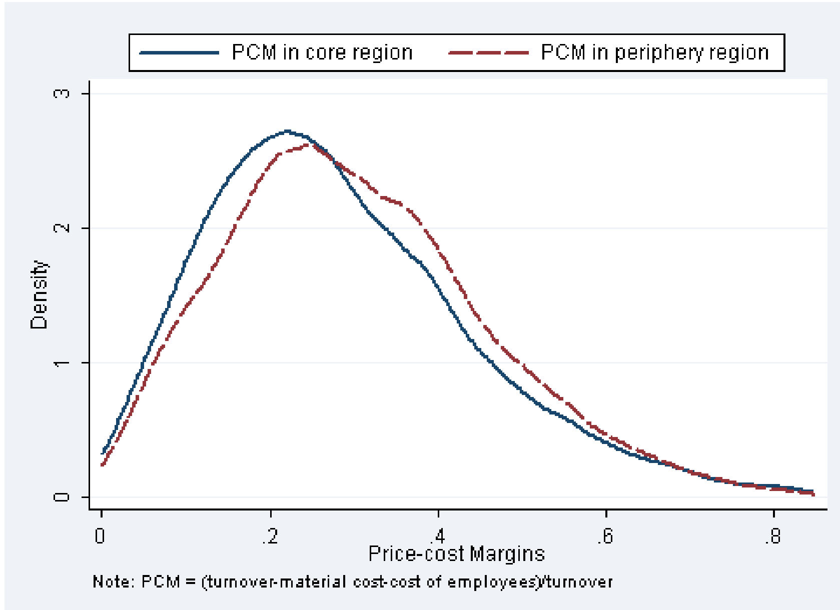

Figure 4: PCM kernal estimates(2004): firms in core and periphery region

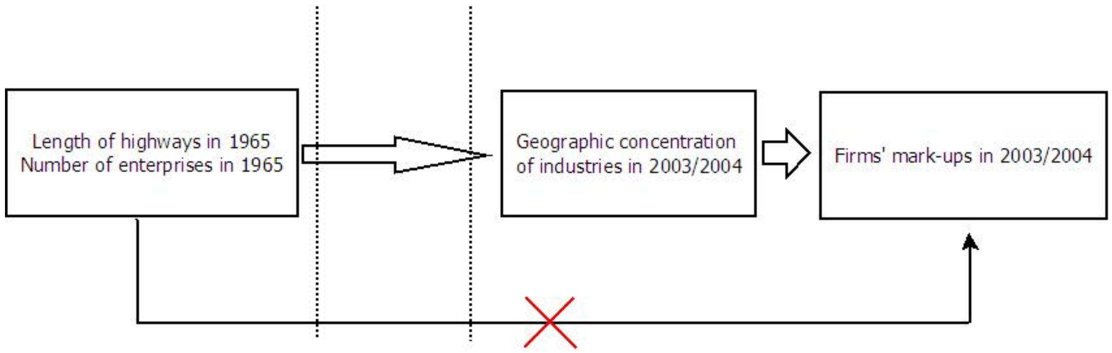

Figure 5: Identification Strategy