

Sun, Churen; Yu, Zhihao; Zhang, Tao

**Conference Paper**

## Agglomeration and Exports with Heterogeneous Firms: Evidence from Chinese Firm-level and County-level Data

52nd Congress of the European Regional Science Association: "Regions in Motion - Breaking the Path", 21-25 August 2012, Bratislava, Slovakia

**Provided in Cooperation with:**

European Regional Science Association (ERSA)

*Suggested Citation:* Sun, Churen; Yu, Zhihao; Zhang, Tao (2012) : Agglomeration and Exports with Heterogeneous Firms: Evidence from Chinese Firm-level and County-level Data, 52nd Congress of the European Regional Science Association: "Regions in Motion - Breaking the Path", 21-25 August 2012, Bratislava, Slovakia, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<http://hdl.handle.net/10419/120706>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Agglomeration and Exports with Heterogeneous Firms: Evidence from Chinese Firm-level and County-level Data\*

Churen Sun<sup>†</sup> Zhihao Yu<sup>‡</sup> Tao Zhang<sup>§</sup>

Version July 1, 2012  
*Preliminary*

## Abstract

The paper proposes a model to investigate the influences of agglomeration on heterogeneous firms' exporting behaviors. Competition and thus selection effect caused by agglomeration forces less productive firms to exit the market while agglomeration externalities increase firms' productivity and decreases industrial fixed entry, fixed and variable exporting costs, and effective labor wage. The former factors decrease while the latter increase firms' exporting possibilities and sales. The model shows that the composite effect of agglomeration on firms' exports takes on a Parabola-shape pattern. Moreover, higher-productivity firms benefit more export premium from agglomeration, which explains why larger and more productive firms in larger cities are more possible to export and exports more. Empirical results based on data from Chinese Industrial Enterprises between 1998 and 2007 verify the theoretical results. The paper also investigates the influences of different agglomeration patterns on firms exports, including home market effect, urban economies and competition effect and diversification effect. It shows that the former two patterns exert a positive while the latter two have a positive influence on firms' exporting behaviors.

**Keywords:** Agglomeration, Export, Productivity heterogeneity, Parabola-shape

**JEL Subject classification:** F12, R19

---

\*The research is funded by the Key Research Base of Humanities and Social Science of Shanghai Municipal of Higher Education (Institute of International Business, Shanghai Institute of Foreign Trade). Churen Sun gratefully acknowledges financial support from General research for Humanities and Social Sciences Project, Chinese Ministry of Education (Grant No. 09YJCZH074) and Innovation Program of Shanghai Municipal Education Commission (Grant No. 10YS168). Zhihao Yu gratefully acknowledges financial support from the Social Sciences and Humanities Research Council of Canada.

<sup>†</sup>Institute of International Business, Shanghai Institute of Foreign Trade, Shanghai, 201600, China, [sunchuren@gmail.com](mailto:sunchuren@gmail.com).

<sup>‡</sup>Department of Economics, Carleton University, Canada, [zhihao.yu@gmail.ca](mailto:zhihao.yu@gmail.ca).

<sup>§</sup>International Business School, Shanghai Institute of Foreign Trade, Shanghai, China, [neotaoism@yahoo.com.cn](mailto:neotaoism@yahoo.com.cn).

## 1 Introduction

The pioneered work of the "new geography economics", [Krugman \(1991\)](#), constructs a novel model to examine the relationship between transportation cost and regional specialization pattern. It's natural to investigate the relationship between of agglomeration and trade originating from the pioneered work of [Krugman \(1991\)](#). Some scholars develop more detailed or analytically solvable models for this topic, such as [Forslid \(2002\)](#), [Cafiso \(2010\)](#), etc. Recently, some researchers even incorporate Meltiz (2003)'s heterogeneous firm model into [Krugman \(1991\)](#)'s CP model to explore the effect of firms' productivity on industrial agglomeration, such as [Okubo \(2007\)](#). On the empirical side, there are some researches on the influences of trade liberalization on industrial agglomeration for different countries, such as [Gao \(2003\)](#) and [Ge \(2006\)](#) for Chinese industrial-level data, [Puga and Venables \(1998\)](#) for French firm data, [Gatto et al. \(2008\)](#) for Italian firm-level data, [Sjoberg and Sjöholm \(2002\)](#) for Indonesian industrial-level data.

There is also some work on the reversal problem, i.e., the influences of agglomeration on trade, or more specifically in this paper, the influences of agglomeration on firms' exports. The motivation originates from the emerging findings of empirical literatures that firms benefits from agglomeration much differently, such as [Baldwin et al. \(2007\)](#) and [Strajer Madsen et al. \(2003\)](#). These results shows that firms active in international trade are different from non-traders in terms of using inputs or capital as well as in terms of performance. For example, as shown in [Bernard and Bradford \(1999\)](#), firms engaging in international trade employ more skilled workers, pay higher wages, invest more R&D, and are more productive than those only selling domestically. This implies that agglomeration has different influences on firms' performance, and thus their exports. In fact, [Harasztosi and Bekes \(2009\)](#) shows that international traders benefits twice as much as non-traders from agglomeration in productivity using Hungarian manufacturing data in 1992 and 2003. This implies that trade requires special knowledge and inputs. If agglomeration offers to firms better input sharing and knowledge spillover for trading firms, then it must have positive effects on firms exports.

The mechanism of positive effects of agglomeration on firms' exports are as follows. First, it offers an better environment for sharing of inputs which are

specialized for firms to export, where firms can acquire inputs and distribute outputs with lower trading costs in case of scale economies. Input sharing also increases firms' productivity by increasing numbers of specialized intermediate inputs for exports they can use. Second, agglomeration forms pools of skilled or specialized labor for firms' exports, which reduces their costs of searching and training and keeps them out of ineffective matches. Third, international competition requires more specialized knowledge, information, high-quality human capital, management and learning skills, absorption capacity of technology and knowledge, etc. (for production and trade). For example, exporting fixed cost consists of marketing, repackaging, finding distributional channels, and it depends on information of the foreign market. While agglomeration enhances the production of these factors and their spillover by decreasing the cost of face-to-face exchange of them and that of the mobility of human capital among firms. [Lovely et al. \(2005\)](#) finds that firms exporting to difficult-to-access countries tend to agglomerate. Agglomeration makes its knowledge spillover and thus reduces exporting fixed cost. Fourth, agglomeration of exporting firms yields export spillovers, which influences extensive or intensive margin of trade of local firms. For example, [Kneller and Pisu \(2007\)](#) and [Anwara and Nguyen \(2011\)](#) find that affiliates of foreign multinationals increase the extensive and intensive margins of trade among local firms.

Empirical results confirm the above postulation. Using Chinese Provincial-level and industrial-level data, [Ge \(2006\)](#) shows that foreign trade and foreign investment significantly affect industry agglomeration in China. It shows that industries dependent on foreign trade and FDI are more likely to locate in regions with easy access to foreign markets, and exporting industries have a higher degree of agglomeration. [Cassey and Schemeiser \(2010\)](#) finds that exporting firms agglomerate geographically with respect to their shipment's destination in addition to agglomerate around ports using Russian customs data. To explain this finding, the authors build on [Melitz \(2003\)](#) and [Chaney \(2008\)](#) to show that agglomeration affects firms' exports as it reduces costs of internationally shipping of goods. They also test their model on region- and state-level exports using Russian and U.S. data and show that it accounts for 40% more of the variation than in gravity-type models without agglomeration. In a model of industry location with endogenous transaction costs, [Duranton and Storper \(2005\)](#) constructs a theoretical model to show how and under which conditions a de-

cline in transport costs can lead to an increase in the total cost of trade. Note that decreasing transportation cost leads to less agglomeration, while increasing trade cost leads to less trade, their result implies that less agglomeration leads to less trade. More recently, [Yilmazkuday \(2011\)](#) also finds that country-side trade patterns are closely related with state-level agglomeration using data from U.S. industries based on counter-factual analysis.

In this paper, we investigate the influences of agglomeration on firms' exports theoretically and empirically. Agglomeration affects firms' exports in two ways. First, it may decrease their fixed and variable exporting costs by increasing firms' efficiency of inputs through channels discussed above. Second, it will increase firms' productivity through two channels. One is that it increases all firms' productivity while higher-productivity firms benefits more from it [Combes et al. \(2011b\)](#). Suppose the demand from the exporting market is constant. Then this channel implies that less productive firms' exporting probability and sales will decrease while those of higher-productivity ones will increase with regional agglomeration level. The other is that it selects less productive firms to exit the agglomeration area so to making regional market competition tougher and thus making concentration degree of regional productivity distribution more concentrated. This channel implies that firms are less possible to export and exports more in more agglomerative regions. They interact with each other and thus the aggregate effect is theoretically unknown.

We propose a theoretic model to investigate the influence of agglomeration on heterogeneous firms' exports in this paper. In this model, agglomeration yields technology spillover so that the efficiency of inputs (specifically, labor, in this paper) increases, while the more productive ones benefit more from this, which in return increases firms' productivity and decreases their fixed entry, fixed and variable exporting costs so that more firms enter into the industry and exporting market. However, the tougher competition due to agglomeration will force less productive firms to exit the market. The synthetic effect is that industrial productivity distribution is right-dilated and agglomeration has a Parabola-shape effect on firms' exports. Moreover, given foreign demand, this further implies that higher-productive firms export more while less-productive ones export less. Empirical results using data from Chinese manufacturing enterprises verifies these results. We also investigate the influences of different agglomeration patterns on firms exports, including home market effect, urban

economies and competition effect and diversification effect. We show that the former two patterns exert a positive while the latter two have a positive influence on firms' exporting behaviors.

The structure in the sequel of the paper is as follows. We construct a local-equilibrium trade model with heterogeneous firms in section 2 and show that agglomeration will increase firms' exporting possibilities and their exporting sales. Moreover, the larger is the region, the larger is the exporting elasticity of productivity of firms there. In section 3, we introduce the empirical models in this paper. Section 4 describes the data used in this paper and its treatments. Section 5 sums up the empirical results and relevant robust analysis. Section 6 investigates the heterogeneous export effect of agglomeration and how different agglomeration patterns affect firms' exports. The paper is concluded in Section 7.

## 2 Agglomeration and Trade: A Simple Model

Suppose there are only two countries (China, denoted by  $C$ , and the rest of the world, denoted in short by  $F$  hereafter) in the economy. In China, there are totally  $M$  regions, within each region  $j$  are there  $N_j$  workers and each worker owns one unit of labor. The land area  $A_j$  of region  $j$  is assumed to be 1 for simplification while its price is  $r_j$ . The land rents in each region is equally shared by workers. Suppose the labor wage in region  $j$  is  $w_j$ , which is common in the same region but may differ across regions. Herein the scale of each city is exogenous given. And the agglomeration of labors is represented by the worker density  $dens_j = N_j/A_j = N_j$  for each region  $j$ . We also assume that the aggregate expenditure  $Y_F$  in  $F$  is exogenously given. Moreover, we assume further that the price index in  $F$  is not influenced by China's export. While for China, the aggregate expenditure  $Y_C = \sum_{j=1}^M (w_j N_j + r_j A_j)$ . We suppose that  $M$  is sufficiently large and the variation among  $N_j$ s is not so large, so that the change of  $N_j$  does not change  $Y_C$ ,  $P$  and  $P^*$  much. Finally, goods are traded with no costs within the same country.

## 2.1 Consumer's utility maximization problem

Suppose all the workers in the two countries are homogenous and they have the following form of utility function<sup>1</sup>

$$U = \left( \sum_{j=1}^M X_j^{\frac{\sigma_C-1}{\sigma_C}} \right)^{\frac{\alpha\sigma_C}{\sigma_C-1}} \left( \int_{j \in \Omega_F} x(i)^{\frac{\sigma_F-1}{\sigma_F}} dj \right)^{\frac{(1-\alpha)\sigma_F}{\sigma_F-1}}, \quad (1)$$

where

$$X_j = \left( \int_{i \in \Omega_j} x_j(i)^{\frac{\sigma_j-1}{\sigma_j}} di \right)^{\frac{\sigma_j}{\sigma_j-1}},$$

and the elasticity of substitution between goods produced in country  $F$  is  $\sigma_k$ , that between two aggregate goods produced in two regions in country  $C$  is  $\sigma_C$ , and that between two varieties produced in region  $j$  in country  $C$  is  $\sigma_j$ .  $\Omega_F$  is the set of goods produced in country  $F$ , and  $\Omega_{Cj}$  is that produced in region  $j$  in country  $C$  available for consumption. We assume that  $\sigma_C < \sigma_j$ , that is, the substitution elasticity between varieties produced in the same region is larger than that between goods produced in two different regions.

Given (1), demand for Chinese good  $i$  produced in regional  $j$  from  $F$  and  $C$  are, respectively

$$x_j^*(i) = \frac{p_j^*(i)^{-\sigma_j} (P_j^*)^{\sigma_j - \sigma_C}}{(P^*)^{1 - \sigma_C}} \alpha Y_F, \quad x_j(i) = \frac{p_j(i)^{-\sigma_j} P_j^{\sigma_j - \sigma_C}}{P^{1 - \sigma_C}} \alpha Y_C, \quad (2)$$

where  $P$  is the aggregate price index for Chinese goods and  $P_j$  is that for Chinese goods produced in region  $j$  at home

$$P = \left( \int_{j \in \Omega_k} P_j^{1 - \sigma_C} \right)^{\frac{1}{1 - \sigma_C}}, \quad P_j = \left( \int_{j \in \Omega_j} p_j(i)^{1 - \sigma_j} di \right)^{\frac{1}{1 - \sigma_j}},$$

where  $p_j(i)$  is the price of variety  $i$  at home.  $p_j^*(i)$ ,  $P_j^*$  and  $P^*$  are accordingly the price variables in  $F$ .

---

<sup>1</sup>For the use of such kind of utility function, we refer the readers to Yilmazkuday (2011) and Cassey and Schemeiser (2010).



## 2.2 Firms' profit maximization problem

### 2.2.1 Domestic market

In each region  $j$ , there is only one industry, i.e., manufacturing industry. A firm in region  $j$  must pay  $f_j$  unit fixed entry cost in terms of land before it observes its "natural" productivity  $\theta$ , which has not been influenced by agglomeration economies in the region, where  $\theta$  is drawn from the probability distribution function  $G_j(\theta)$  (with density function  $g_j(\theta)$ ). Herein  $G_j$  and thus  $g_j$  are closely related with the agglomeration economies in region  $j$ . As the firm's productivity  $\theta$  is randomly drawn from the distribution  $G_j$ , we denote the firm by its productivity  $\theta$ . After the firm enters the industry, it starts production. We assume that any firm is not free to change regions and it has to determine how much to produce to supply domestic and foreign demands. We assume that the production function in region  $j$  without agglomeration economy is  $x_j(\theta) = \theta l_j(\theta)$ , where  $x_j(\theta)$  is the output and  $l_j(\theta)$  is the quantity of workers hired for production. Involving agglomeration, we assume that there exists knowledge spillover in labor market among regions. Each worker is made more productive by interactions (such as exchanging diverse ideas) with other workers. Such interactions among workers can be found in Fujita and Ogawa (1982) and Lucas and Rossi-Hansberg (2002) among others. According to these papers, there exists spatial decay in interactions. Specifically, following Combes et al. (2011b), we assume that the effective labor supplied by one individual worker in firm  $\theta$  in region  $j$  is  $a(N_j + \delta \sum_{i \neq j} N_i) \theta^{D_j - 1}$ , where  $a(0) = 1, a' > 0, a'' < 0$ ,  $D_j = \ln d(N_j + \delta \sum_{i \neq j} N_i), d(0) = 1, d' > 0, d'' < 0$ , and  $\delta \in [0, 1]$  is the decay parameter measuring the strength of across-region relative to within-region interactions. Thus, the production function of firm  $\theta$  in region  $j$  with agglomeration economy is  $x_j(\theta) = a(N_j + \delta \sum_{i \neq j} N_i) \theta^{D_j} l_j(\theta)$ . Then the firm's profit from selling only domestically is

$$\pi_j(\theta) = p_j(\theta)x_j(\theta) - \frac{w_j x_j(\theta)}{a_j \theta^{D_j}} - r_j f_j, \quad (3)$$

where  $a_j = a(N_j + \delta \sum_{i \neq j} N_i)$ . The firm's domestically optimal pricing rule and domestic sale quantity are respectively

$$p_j(\theta) = \frac{\bar{m}_j w_j}{a_j \theta^{D_j}}, x_j(\theta) = P^{\sigma_C - 1} P_j^{\sigma_j - \sigma_C} \left( \frac{a_j \theta^{D_j}}{\bar{m}_j w_j} \right)^{\sigma_j} \alpha Y_C, \quad (4)$$



where  $\bar{m}_j = \frac{\sigma_j}{\sigma_j - 1}$ . Its domestic sale is then

$$R_j(\theta) = p_j(\theta)x_j(\theta) = P^{\sigma_C - 1}P_j^{\sigma_j - \sigma_C} \left( \frac{a_j\theta^{D_j}}{\bar{m}_j w_j} \right)^{\sigma_j - 1} \alpha Y_C. \quad (5)$$

Thus, its net profit is

$$\pi_j(\theta) = \frac{R_j(\theta)}{\sigma} - r_j f_j. \quad (6)$$

From (14), we can get the cut-off productivity condition of non-exporters as follows

$$P^{\sigma_C - 1} \alpha Y_C n_j^{\frac{\sigma_j - \sigma_C}{1 - \sigma_j}} \left( \frac{\bar{m}_j w_j}{a_j} \right)^{1 - \sigma_C} \theta_j^{D_j(\sigma_j - 1)} \tilde{\theta}_j^{\sigma_C - \sigma_j} = \sigma_j r_j f_j. \quad (7)$$

Moreover, its labor hired is thus

$$l_j = P^{\sigma_C - 1} P_j^{\sigma_j - \sigma_C} \left( \frac{a_j \theta^{D_j}}{\bar{m}_j w_j} \right)^{\sigma_j - 1} \frac{\alpha Y_C}{\bar{m}_j w_j} = \frac{R_j(\theta)}{\bar{m}_j w_j}. \quad (8)$$

As there is not fixed production cost after entry, each firm must start production and sell at home country. Suppose there are  $n_j$  firms selling domestically in region  $j$  at equilibrium. Then the price index  $P_j$  of varieties produced in region  $j$  and that at home are, respectively,

$$P_j = n_j^{\frac{1}{1 - \sigma_j}} \frac{\bar{m}_j w_j}{a_j \tilde{\theta}_j}, P = \left( \sum_{j=1}^M \left( \frac{\bar{m}_j w_j}{n_j^{\frac{1}{\sigma_j - 1}} a_j \tilde{\theta}_j} \right)^{\sigma_C - 1} \right)^{\frac{1}{\sigma_C - 1}}, \quad (9)$$

where

$$\tilde{\theta}_j \triangleq \left( \int_0^{+\infty} \theta^{D_j(\sigma_j - 1)} dG_{jD}(\theta) \right)^{\frac{1}{\sigma_j - 1}}$$

is the average productivity level in region  $j$ , where  $G_{jD}(\theta)$  is the ex post probabil-

ity distribution function of productivity of firms in region  $j$  defined as follows:

$$G_{jD}(\theta) = \begin{cases} \frac{G_j(\theta)}{1-G(\underline{\theta}_j)} & \theta \geq \underline{\theta}_j, \\ 0 & \text{else,} \end{cases} \quad (10)$$

where  $\underline{\theta}_j$  is the productivity cutoff below which a non-exporter obtains negative profits and thus exits the market.

### 2.2.2 Foreign market

We assume that a firm in region  $j$  at home has to pay another fixed cost  $f_{xj}$  units of lands to export its output abroad. We also assume that exporting any variety from each region  $j$  to abroad entails an "Samuelson" type of iceberg cost  $\tau_j > 1$ , which is common in each region, but may differ across regions. That is, to sell one unit of product abroad,  $\tau_j$  unit must be shipped out of region  $j$  in China. This cost may involve transportation costs, tariff barriers, distribution costs, marketing costs, repackaging costs, etc.

A firm with productivity  $\theta$  locating in region  $j$  has the following form of profit function:

$$\pi_j^*(\theta) = p_j^*(\theta)x_j^*(\theta) - \frac{\tau_j w_j x_j^*(\theta)}{a_j \theta^{D_j}} - r_j f_{xj}. \quad (11)$$

The firm must choose the Bertrand pricing rule to maximize its exporting profit. Similarly to the deduction in the former subsection, the firm's optimal pricing rule and its optimal export are respectively

$$p_j^*(\theta) = \frac{\bar{m}_j \tau_j w_j}{a_j \theta^{D_j}} = \tau_j p_j(\theta), x_j^*(\theta) = \frac{(P^*)^{\sigma_C - 1}}{(P_j^*)^{\sigma_C - \sigma_j}} \left( \frac{a_j \theta^{D_j}}{\tau \bar{m}_j w_j} \right)^{\sigma_C j} \alpha Y_F. \quad (12)$$

Its foreign sale is then

$$X_j(\theta) = p_j^*(\theta)x_j^*(\theta) = (P^*)^{\sigma_C - 1} (P_j^*)^{\sigma_j - \sigma_C} \left( \frac{a_j \theta^{D_j}}{\bar{m}_j \tau_j w_j} \right)^{\sigma_j - 1} \alpha Y_F. \quad (13)$$

Thus, its net profit is

$$\pi_j^*(\theta) = \frac{X_j(\theta)}{\sigma_j} - r_j f_{xj}. \quad (14)$$

From (14), we can get the cut-off productivity condition of exporters

$$(P^*)^{\sigma_C-1} \alpha Y_F n_{xj}^{\frac{\sigma_j-\sigma_C}{1-\sigma_j}} \left( \frac{\tau_j \bar{m}_j w_j}{a_j} \right)^{1-\sigma_C} \underline{\theta}_{xj}^{D_j(\sigma_j-1)} \tilde{\theta}_{xj}^{\sigma_C-\sigma_j} = \sigma_j r_j f_{xj}. \quad (15)$$

Moreover, its labor hired is thus

$$l_j^* = (P^*)^{\sigma_C-1} (P_j^*)^{\sigma_j-\sigma_C} \left( \frac{a_j \theta^{D_j}}{\bar{m}_j \tau_j w_j} \right)^{\sigma_j-1} \frac{\alpha Y_C}{\bar{m}_j w_j} = \frac{X_j(\theta)}{\bar{m}_j \tau_j w_j}. \quad (16)$$

Then the price index  $P_j^*$  and  $P^*$  are accordingly

$$P_j^* = n_{xj}^{\frac{1}{1-\sigma_j}} \frac{\bar{m}_j \tau_j w_j}{a_j \tilde{\theta}_{xj}}, P^* = \left( \sum_{j=1}^M \left( \frac{\bar{m}_j \tau_j w_j}{n_{xj}^{\frac{1}{\sigma_j-1}} a_j \tilde{\theta}_{xj}} \right)^{\sigma_C-1} \right)^{\frac{1}{\sigma_C-1}}, \quad (17)$$

where

$$\tilde{\theta}_{xj} \triangleq \left( \frac{1}{1 - G_j(\underline{\theta}_{xj})} \int_{\underline{\theta}_{xj}}^{+\infty} \theta^{D_j(\sigma_j-1)} dG_j(\theta) \right)^{\frac{1}{\sigma_j-1}} \quad (18)$$

is the average productivity level of exporting varieties from region  $j$  of China in the foreign country,  $\underline{\theta}_{xj}$  is the cutoff of productivity of exporting firms, at which a firm is indifferent from exporting or only selling domestically, and  $n_{xj}$  is the number of exporting firms in region  $j$ . Only those firms whose productivity is higher than  $\underline{\theta}_{xj}$  will export so that there holds

$$n_{xj} = \frac{1 - G_j(\underline{\theta}_{xj})}{1 - G_j(\underline{\theta}_j)} n_j, \quad (19)$$

where  $\varsigma_j \triangleq \frac{1 - G(\underline{\theta}_{xj})}{1 - G(\underline{\theta}_j)}$  is the ex ante expected exporting probability of incumbents in the domestic market.

### 2.2.3 Equilibrium

At equilibrium, both the labor market and the land market in each region clear. Thus we have

$$n_j f_j + n_{xj} f_{xj} = 1, \quad (20)$$

$$\int_0^\infty R_j(\theta) dG_{jD}(\theta) + \frac{1}{\tau_j} \varsigma_j \int_{\underline{\theta}_{xj}}^\infty X_j(\theta) dG_{jX}(\theta) = \bar{m}_j w_j N_j, \quad (21)$$

where  $G_{jX}(\theta)$  is the ex post productivity distribution of exporters in region  $j$ .

Firm's equilibrium entry condition is

$$\int_0^\infty \pi_j(\theta) dG_{jD}(\theta) + \varsigma_j \int_{\underline{\theta}_{xj}}^\infty \pi_j^*(\theta) dG_{jX}(\theta) = 0. \quad (22)$$

(15) together with (19)-(22) determines the market equilibrium given foreign expenditure  $Y_F$ .

According to (19) and (20), We have

$$n_j = \frac{A_j}{f_j + f_{xj} \varsigma_j}, n_{xj} = \frac{A_j \varsigma_j}{f_j + f_{xj} \varsigma_j}, \quad (23)$$

which implies that  $\underline{\theta}_{xj}$  is decreasing in  $n_j$ . According to (15) and (23), we have

$$\underline{\theta}_{xj} = \left\{ \left( \frac{\kappa_j}{\kappa_j - D_j(\sigma_j - 1)} \frac{A_j \varsigma_j}{f_j + f_{xj} \varsigma_j} \right)^{\frac{\sigma_j - \sigma_C}{\sigma_j - 1}} \left( \frac{\bar{m}_j \tau_j w_j}{a_j} \right)^{\sigma_C - 1} \frac{\sigma_j r_j f_{xj}}{Y_F} \right\}^{\frac{1}{D_j(\sigma_C - 1)}}. \quad (24)$$

where  $\bar{Y}_F = \frac{\alpha Y_F}{(P^*)^{1 - \sigma_C}}$  is the real incomes of the foreign country consuming goods produced in China.

Let  $\mathcal{A}_j = \int_0^\infty R_j(\theta) dG_{jD}(\theta)$ ,  $\mathcal{B}_j = \int_{\underline{\theta}_{xj}}^\infty X_j(\theta) dG_{jX}(\theta)$ . Combining (15) with (21) and (22) yields the following system of equations:

$$\mathcal{A}_j + \frac{1}{\tau_j} \varsigma_j \mathcal{B}_j = \bar{m}_j w_j N_j, \quad (25)$$

$$\mathcal{A}_j + \varsigma_j \mathcal{B}_j = \sigma_j r_j (f_j + \varsigma_j f_{xj}). \quad (26)$$

which together with (9) and (17) yields the relationship between  $\mathcal{A}_j$  and  $\mathcal{B}_j$

$$\mathcal{B}_j = \Delta^{-1} \tau_j^{1-\sigma_C} \left( \frac{n_{xj}}{n_j} \right)^{\frac{\sigma_j - \sigma_C}{1-\sigma_j}} \left( \frac{\tilde{\theta}_{xj}}{\tilde{\theta}_j} \right)^{\sigma_C - 1} \mathcal{A}_j, \quad (27)$$

where  $\Delta$  is the relative income between China and the foreign country defined as  $\Delta = \frac{\bar{Y}_C}{\bar{Y}_F}$ , wherein  $\bar{Y}_C = \frac{\alpha Y_C}{\rho^{1-\sigma_C}}$  is the real incomes of China consuming goods produced in China, respectively.

### 2.3 Parameterization

To find analytic solutions of the above equilibrium, we follow Antras and Helpman (2004), Helpman et al. (2004, 2008), Melitz and Ottaviano (2008), Eaton et al. (2011) and other many papers to assume that  $G_j(\theta)$  follows a Pareto distribution of the following form

$$G_j(\theta) = \begin{cases} 1 - \left( \frac{b_j}{\theta} \right)^{\kappa_j} & \theta \geq b_j, \\ 0 & \text{else,} \end{cases}$$

where  $\kappa_j$  is the configuration parameter (also called the concentration degree) and  $b_j$  is the lower bound of industrial productivity distribution. Under this assumption, the ex ante exporting probability of incumbents in the domestic market is

$$\varsigma_j = \frac{1 - G_j(\underline{\theta}_{xj})}{1 - G_j(\underline{\theta}_j)} = \left( \frac{\underline{\theta}_j}{\underline{\theta}_{xj}} \right)^{\kappa_j}, \quad (28)$$

and the average productivity of all the incumbents and the exporters are, respectively,

$$\tilde{\theta}_j = \left( \frac{\kappa_j}{\kappa_j - D_j(\sigma_j - 1)} \right)^{\frac{1}{\sigma_j - 1}} \underline{\theta}_j^{D_j}, \quad \tilde{\theta}_{xj} = \left( \frac{\kappa_j}{\kappa_j - D_j(\sigma_j - 1)} \right)^{\frac{1}{\sigma_j - 1}} \underline{\theta}_{xj}^{D_j}. \quad (29)$$

Dividing (7) by (15) and substituting (29) into the resultant equation, we conclude the following expression

$$\varsigma_j = \left( \frac{\underline{\theta}_j}{\underline{\theta}_{xj}} \right)^{\kappa_j} = \left( \frac{f_j}{f_{xj} \Delta \tau_j^{\sigma_C - 1}} \right)^{\frac{1}{\frac{D_j(\sigma_C - 1)}{\kappa_j} - \frac{\sigma_C - \sigma_j}{\sigma_j - 1}}}. \quad (30)$$

Thus, according to (27), we have

$$\mathcal{B}_j = \frac{f_{xj}}{f_j} \mathcal{A}_j. \quad (31)$$

Suppose now that the difference between  $\sigma_C$  and  $\sigma_j$  is not too large. Furthermore, we make the following assumption.

**Assumption 1**  $\frac{D_j(\sigma_C-1)}{\kappa_j} > \frac{\sigma_C-\sigma_j}{\sigma_j-1}$  for each  $j$ .

Obviously, Assumption 1 holds if  $\sigma_C < \sigma_j, \forall j$  holds, which is satisfied in most common situation as competition between firms within a region usually is tougher than that among regions.

Under Assumption 1, we get straightforward the following result, which coincides with our intuition and many empirical results verified in many new-new trade literatures, e.g., Melitz (2003).

**Proposition 1** *Under Assumption 1, the firms' exporting possibility  $\varsigma_j$  in region  $j$  is decreasing in the fixed exporting cost  $f_{xj}$ , the variable exporting cost (transportation cost)  $\tau_j$  and the relative real income  $\Delta$  between the two countries, and is increasing in the fixed entry cost  $f_j$ .*

From (30), we can also find the following result.

**Proposition 2** *Suppose agglomeration of region  $j$  does not affect the whole economy much (i.e., change of  $N_j$  does not affects  $\Delta$  much). If  $f_j, f_{xj}$  and  $\tau_j$  are all constant, then firms' exporting probability  $\varsigma_j$  is increasing in the agglomeration economy  $N_j$  if  $f_j < f_{xj}\Delta\tau_j^{\sigma_j-1}$  as  $D_j$  is increasing in  $N_j$ , ceteris paribus. This result is reversed if  $f_j > f_{xj}\Delta\tau_j^{\sigma_j-1}$ .*

From Proposition 2, we conclude immediately the following result.

**Corollary 1** *Suppose there are some regions  $j \in \{1, \dots, M\}$  whose fixed entry costs, fixed and variable exporting costs  $f_j, f_{xj}, \tau_j$  are such that  $f_j < f_{xj}\Delta\tau_j^{\sigma_j-1}$  while the others such that  $f_j > f_{xj}\Delta\tau_j^{\sigma_j-1}$ . Then the relationship between ex regional ante exporting probability and agglomeration level takes on a parabola-shape.*

In the sequel, we will empirically test Corollary 1 using firm-level data from Chinese manufacturing enterprises and show that it takes on a Parabola-shape trend in the acquirable Chinese data.

From (30), we know that Combining (7), (25), (26) and (31), we can find that

$$\mathcal{A}_j = \frac{\sigma_j r_j f_j \kappa_j}{\kappa_j - D_j(\sigma_j - 1)}, \mathcal{B}_j = \frac{\sigma_j r_j f_{xj} \kappa_j}{\kappa_j - D_j(\sigma_j - 1)}. \quad (32)$$

which together with (25) yields

$$\frac{w_j N_j}{r_j \mathcal{A}_j} = (\sigma_j - 1) [f_j + s_j f_{xj} / \tau_j]. \quad (33)$$

Plugging (33) into (30) and combining the resulted expression with (7) and (15) yields  $M$  equations with  $2M$  unknown variables  $w_j$  and  $r_j$  for  $j = 1, \dots, M$ . Substituting (30) and (23) into (32), we get another  $M$  equations with unknown variables  $w_j$  and  $r_j$  for  $j = 1, \dots, M$ . Solving the total  $2M$  equations yields all  $w_j$  and  $r_j$ .

Substituting (15), (23) and (30) into (20) and (21), we have<sup>2</sup>

$$X_j(\theta) = \underbrace{\bar{Y}_F^{\frac{\sigma_j - 1}{\sigma_C - 1}} \left( \bar{m}_j \tau_j \frac{w_j}{a_j} \right)^{1 - \sigma_j} \left( \frac{f_j}{f_{xj}} \frac{s_j}{f_j + f_{xj} s_j} \mathcal{B}_j \right)^{\frac{\sigma_C - \sigma_j}{\sigma_C - 1}} A_j^{\frac{\sigma_j - \sigma_C}{\sigma_C - 1}}}_{\text{common export effect caused by agglomeration}} \underbrace{\theta^{D_j(\sigma_j - 1)}}_{\text{heterogeneous export effect}} \quad (34)$$

In the above expression, we divide the effect of agglomeration on firms' export into two forms. The first is common for all the firms in each region, called the **common export effect of agglomeration**. The second is idiosyncratic across firms as their productivity is heterogeneous, which is called **heterogeneous export effect**. As  $D_j = \ln d(N_j + \delta \sum_{i \neq j} N_i)$  is increasing in  $N_j$ . Define the exporting elasticity of productivity in region  $j$  as  $\varepsilon_j = D_j(\sigma_j - 1)$ . Then we conclude immediately the following proposition from (34).

**Proposition 3** *The exporting elasticity of productivity is increasing in the scale of all the regions.*

If  $\sigma_j = \sigma_C$ , then it's easy to conclude the following result.

**Proposition 4** *Suppose  $\sigma_j = \sigma_C$ . Then the common export effect of agglomeration in region  $j$  increases with its agglomeration level.*

<sup>2</sup>(34) implies further that firms' exporting sales are



Proposition 4 satisfies only if  $\sigma_j = \sigma_C$ . It implies that agglomeration has a positive influence on each firm's exporting sale with the increase of integration of domestic economy. The decrease of market segregation will eventually lead a country to engage more deeply into the international division of labor. However, this is not the case for most regions. In the general situation, agglomeration affects firms' exports in a complicate way. Note that  $\bar{Y}_F$  is the real income consuming Chinese goods in the foreign country,  $\frac{w_j}{a_j}$  is the effective wage, and  $B_j$  is the firm-level average export in region  $j$ . Suppose  $\bar{Y}_F$  is not influenced by agglomeration, and the effective wage in region  $j$  decreases with agglomeration. As  $r_j$  increases with  $N_j$ , the firm-level average export increases with agglomeration. Moreover,  $\frac{s_j}{f_j + f_{xj}s_j}$  is increasing in  $s_j$ . We conclude that  $\left(\bar{m}_j \tau_j \frac{w_j}{a_j}\right)^{1-\sigma_j}$  (which can be considered as one unit of variable cost of foreign sale as  $\frac{\bar{m}_j \tau_j w_j}{a_j}$  is each worker's effective wage for producing goods selling oversea) increases, while  $\left(\frac{f_j}{f_{xj}} \frac{s_j}{f_j + f_{xj}s_j} B_j\right)^{\frac{\sigma_C - \sigma_j}{\sigma_C - 1}}$  (which can be considered as each firm' average exporting sale per unit of fixed exporting cost) decreases with agglomeration in region  $j$ . Thus its difficult to just the synthetic common export effect of agglomeration in the region. In practice, agglomeration in one region will influences both the (relative) real incomes consuming Chinese goods between the two countries. It may increase the effective labor wage in the region. It may also affects fixed entry cost, fixed and variable exporting costs. Thus, it's very difficult to judge how agglomeration affects firms' export behaviors in the general case. We left it to empirical analysis to be stated in the sequel.

However, under a certain condition, we have the following scenario.

**Hypothesis 1** *Suppose Assumption 1 holds and  $\sigma_j \neq \sigma_C$ . Assuming that agglomeration in region  $j$  does not change the real income consuming Chinese goods of the foreign country, decrease labors' effective wage and increase the region's firm-level average export. If agglomeration induces more exporting competition between firms so that the decrease of each firm's average exporting sale per unit of fixed exporting cost dominates the increase of the save of each unit of variable exporting cost in the first agglomerating stage while it reverses in the second agglomerating stage, then the influence of regional agglomeration on firms' exporting sales takes on a Parabola-shape pattern.*

In the sequel , we will use firm-level data from Chinese manufacturing enterprises to test and show that Proposition 3 and Hypothesis 1 hold for China's

case.

### 3 Econometric model

#### 3.1 Employment density and trade

From Proposition 2, we know that an incumbent's exporting probability is increasing in the regional agglomeration level (measured in labor density in the region) under some conditions. Carefully investigating (34), we set the econometrical model of estimating firms' exporting choices as follows

$$\begin{aligned} \mathbb{X}_{rjit} = & \phi \ln dens_{rj} + \varphi \ln dens_{rj}^2 + \psi \ln A_r + \vartheta \ln \theta_{rjit} + \lambda \ln dens_{rj} \ln \theta_{rjit} + Z_{rjit} \zeta \\ & + \alpha_r + \beta_j + \gamma_i + \eta_t + cons + \varepsilon_{rjit}, \end{aligned} \quad (35)$$

where  $r, j, i, t$  are indices of region, industry, firm and time, respectively,  $\mathbb{X}_{rjit}$  is firm  $i$ 's exporting decision variable in time  $t$ , which is 0 if it does not export, and is the logarithm of the foreign sale  $\ln X_{rjit}$  if its exporting sale is positive,  $\alpha_r, \beta_j, \gamma_i$  and  $\eta_t$  are respectively the regional, the industrial, the firm and the time fixed effect,  $dens_{rj} = \frac{N_{rj}}{A_r}$  is the density of industrial activity in industry  $j$  in regional  $r$ , which is measured by the number of employment (industrial workers) per square kilometer,  $N_{rj}$  is the sum of the employment of firms in industry  $j$  in region  $r$ ,  $A_r$  is the area of the county in square kilometer, and  $Z_{rjit}$  is the vector of control variables for firm  $i$  in industry  $j$  in region  $r$  in time  $t$ . In our empirical analysis,  $Z_{rjit}$  includes fiscal expenditure, human capital, institute level in region  $j$ , ownership, belonging, operating status, size, industry, and region dummy in time  $t$ . Here we involve a quadratic term of  $\ln dens_{rj}$  to investigate whether agglomeration has an inverted-U effect on firms' exporting choices, and we use the cross term  $\ln dens_{rj} \ln \theta_{rjit}$  to investigate whether the exporting elasticity of firm productivity in region  $r$  is increasing in the agglomeration level  $dens_{rj}$ . Note that in (35) the term  $\Delta = \frac{P^{\sigma_C - 1} Y_C}{(P^*)^{\sigma_C - 1} Y_F}$  is incorporated into the three fixed effect terms.

We divide (35) into two stages. The first is to examine the influence of agglomeration on firm exporting choice. In this stage, according to Proposition 3, there should be  $\vartheta > 0, \lambda < 0$ . That is, the influence of productivity on firm exporting possibility is positive and it's decreasing in the agglomeration level. The

second stage is to investigate the influence of agglomeration on firm exporting sale. In this stage, according to Proposition 3, there should be  $\vartheta\lambda > 0$ , i.e.,  $\vartheta$  and  $\lambda$  have the same sign. That is, higher-productivity firms' exporting sales are affected more by agglomeration. Moreover, according to Proposition 4, there shall be  $\phi > 0, \psi > 0$  if the conditions in the proposition hold.

### 3.2 Local market effect, urban economy and trade

As agglomeration generate two kinds of effects, the localization effect ( within industry external economies) and the urbanization effect (between industry external economies), it's necessary to separate agglomeration economies into these two effects and investigate how they affect firms' exports.

According to Martin et al. (2008), the measure of localization economy at firm level is

$$Loc_{rjit} = \ln[Emp_{rjt} - Emp_{rjit} + 1],$$

where  $Emp_{rjit}$  is the number of workers in firm  $i$  in manufacturing industry  $j$  of region in year  $t$ ,  $Emp_{rjt}$  is the total employment of industry  $j$  in region  $r$  in year  $t$ , and the urbanization economy is measured by

$$Urb_{rjt} = \ln[Emp_{rt} - Emp_{rjt} + 1],$$

where  $Emp_{rt}$  is the number of workers in region  $r$  in year  $t$ . Following Martin et al. (2008), we can estimate the impact of these two agglomeration economies on firms' exports. We also add an industrial index  $Div_{rjt}$  and the degree of competition a firm faces in the region it locates control other effects from local and external competition, where

$$Div_{rjt} = -\ln HSD_{rjt},$$

and

$$Comp_{rjt} = -\ln Herf_{rjt},$$

where  $HSD$  is the index of Herfindal sectoral diversity in industry  $j$  of region  $r$

in year  $t$  defined as

$$HSD_{rjt} = \sum_{j \neq j'} \left( \frac{Emp_{rj't}}{Emp_{rt} - Emp_{rjt}} \right)^2,$$

and  $Herf_{rjt}$  is defined as

$$Herf_{rjt} = \sum_{k \in \mathbb{S}_{rjt}} \left( \frac{Emp_{rjkt}}{Emp_{rjt}} \right)^2$$

is the Herfindahl index of employment concentration and  $\mathbb{S}_{rjt}$  is the set of firms belonging to industry  $j$  in region  $r$  at time  $t$ . Then our econometrical model is as follows

$$\begin{aligned} \mathbb{X}_{rjit} = & \phi \ln dens_{rjt} + \varphi \ln dens_{rjt}^2 + \psi \ln A_{rt} + \vartheta \ln \theta_{rjit} + \lambda \ln \theta_{rjit} \ln dens_{rjt} + \xi Loc_{rjit} \\ & + \rho Urb_{rjt} + \varrho Div_{rjt} + \chi Comp_{rjt} + Z_{rjit} \zeta + \alpha_r + \beta_j + \gamma_i + \eta_t + cons + \varepsilon_{rjit} \end{aligned} \quad (36)$$

## 4 Data description and treatments

### 4.1 Data Source

According to the nature of this research, our empirical dataset is composed and organized in two folds: Chinese city-level and firm-level data. The formal city-level data is collected from three major sources: *China Urban Statistic Yearbook*, *China Statistic Yearbook* and *China Regional Statistic Yearbook (1990-2009)*. We take the city statistic variables of longer period than the firm level data with following concerns: the aggregate urban social political and demographic indexes as the external environment has time-lag effect as well as the sorting-effect on firm location and labor migration, and can be utilized as instrumental movement estimation in micro-macro econometric application. The later firm-level data comes from the Annual Survey of Industrial Firms (**ASIF**) cross-sectional data collected by China National Bureau of Statistics between 1998 and 2007.

The data description section here mainly discusses the usage and treatments of **ASIF** dataset. The whole **ASIF** contains all detailed information for all state-owned and non-state firms above designated scale (above 5 million Yuan) with all operational, financial and managerial items (e.g. there are 135 variables de-

scribing firm information in year 2004) in (1) mining, (2) manufacturing and (3) production and distribution of electricity, gas and water sector with all 40 industries, and 90 4-digits and 600 6-digits sub-industries (see the appendix of industrial categories). The number of firms covered by this dataset is 165,118 in 1998 and 336,768 in 2007, respectively. The industry section of *China Statistic Yearbook* and *China Industrial Statistic Yearbook* and reports in *China Markets Yearbook* are compiled and based on this dataset, which covers 95% of the industry gross output in these yearbooks ((Lin et al. 2009); (Lu and Tao 2009); (Brandt et al. 2011)). The only difference between **ASIF** and *China Industrial Statistic Yearbook* is that the later is reported in aggregated industries and sectors while the former is individual firms and plants. The duration of this dataset includes the WTO entry year 2001, and new industrial information calculation in year 2004, which is sensitive to the impact and fluctuations of structural change.

The **ASIF** dataset provide us unique opportunity to observe China enterprises performance with large and comprehensive sample and the time duration also enables us to avoid some radical economic policy changes in the early and middle 1990s (structural change, SOE reform, etc.). China undertook a series of economic policy reform since 1978, and such structural adjustments stabilized in the later years. Especially in the late 1990s, more and more domestic firms and plants are emerging and competing with their foreign counterparts for the unconditional governmental fiscal loans, abolishing industrial licensing, equal foreign direct investment opportunities, cutting import duties, deregulating capital markets and reducing tax rates. Therefore, the time period of this dataset with relatively stable price indices and deflators for all variables is suitable to indicate the firm performance with specific effects.

## 4.2 Variables Definition and Consistency

To facilitate this research, we take variables closed related to the firm-idiosyncratic variables with all commonly used operational, managerial and financial statistics. Notably, since year 2003, the China National Bureau of Statistics started to implement new National General Specifications of Industry (GBT/4757), therefore there were two systems registered and coded in this dataset: before 2003 and afterwards. In this research, we adjust all industry-related codes by new GBT/4757 system. Apart from these direct changes, some variables in differ-

ent years are noted differently or partially, which requires close examination on their explicit definitions - some attributive descriptions varied for different situations. For example, words like gross value, net value, total value, sum amount are missing in some case or years, 'fixed capital' refers to net value of fixed capital, 'asset' refers to the gross value of asset, 'employment' refers to the average employment numbers, etc. And, since year 2004, the total value of sales was no longer surveyed; instead, this term was replaced by sum income of major revenue. And, the geographic codes were largely extended to 12 digits since 2004 (6 digits for province-city-county, plus other 6 digits for village-district-street accordingly). By adjusting above major statistics changes, we try to make sure the whole dataset is consistent and tractable both in terms of cross-section and longitudinal manner.

### 4.3 Missing variables, missing values and treatments

To most empirical studies, incomplete sample information or missing values have critical influence on the validity of econometric estimation. There are missing variables and values in differently years of **ASIF**.

Missing values can be found in year 2001 to 2007. Some are due to statistical index changes or accounting system adjustment, e.g. 'total production value constant value' and 'product sale value' (2004-2007, both variables are no longer accounted in the national survey). Some are due to miss accounting, e.g. 'firm status' (2001), 'firm size' (2002, 2004), 'total value of fixed capital' (2003), 'total export value' (2004), 'current value of gross output' (2004) and 'current value of sales' (2005). Some of these variables can be computed by accounting principles, such as the 'total value-add' (2001, 2004) equalizes 'sum of current value of output' - 'middle inputs plus' + 'value add tax'. Missing variables like 'total export value' can only be measured by matching whole **ASIF** among same firms of different years by linear interpolation method.

Like all large scale survey data, some variables contain missing values. We treat missing values with two principles: 1) leave all accounting variables (operational, financial and managerial variables) unchanged; 2) match and refill all possible firm-idiosyncratic variables not related to accounting variables, e.g. firm id, location and postal address, operation status, founding year, registration type, belonging, stock share. In which we attempts to maximize the cover-

age and usage of the whole **ASIF** dataset.

First of all, we checked the dataset's firm id yearly. Not as simple as previous research ((Lin et al. 2009; Lu and Tao 2009; Brandt et al. 2011)) claimed that the **ASIF** contains unique firm id for each firm throughout 10 year period. Instead, there are repetitive firm ids in each year (463 firms and one wrongly registered firm). Comparing with the over two million firms dataset, those firms seems minute and no harm - we are arguing that many of those firms are large SOEs, such as provincial petro line companies, regional tobacco companies and provincial electricity groups who are sharing the same registered names and controlled by the central government. They are influential both in terms of local gross domestic product, revenue, taxation income, industrial forward-backward linkages as well as local employment. Some of them are even 80 times than some 'ordinary' surveyed firms. We believe deleting those repetitive firms could cause the very problem of selection bias. To deal with such problem, we match these firms' locations, owners' name (legalistic representative name), major product items accordingly, identify, assign and add these firms with new firm ID in the **ASIF** dataset respectively.

Secondly, we exam our newly defined firm-idiosyncratic variables accordingly. To save the miss value treatment discussion, here we list two examples. Take the operational status in year 1999 as the first case. There are 3,904 firms are either missing or noted as 0, capital A or letter a. We take these lots as the sub-database to match with year 1998, 2000 to 2007 firm survey data accordingly. Our rational is that if these firms exited after year 1998, they would not be accounted or noted in the 1999 survey, vice versa, if they appear in the 2000 or later years' survey, it means their operational status are active instead of frozen or closed. In fact, there are 3,276 firms being noted in 1998 survey, and the other 628 firms have 471 firms not exist in the following years' survey since 2000. We can safely conclude that these 471 firms only exist in year 1999, and by the end of this year, their operational status shall be registered as "canceled" (not belonged to status as 'establishing', 'operating', 'frozen', or 'other'). Another example is treating the missing values of location. Take the location variable in year 2000 as the case. There are 67 firms' location codes are missing, noted as 0 or wrongly registered (some letters replaced the literal 6 digits location codes). Same mechanism as previous case, we match and sort these firms with pre and after years' survey data. There are 57 firms being confirmed with 1998-1999



survey data, 6 of 10 firms are found in the 2001-2007 survey data. We match the left 4 firms with their exclusive information - phone number, mail address, firm name, major products, owner names, and assign their location codes accordingly.

Based on above techniques, we match missing values in variables like location, operation status, founding year, registration type, belonging and stock shares respectively. Likewise, we generate these variables' dummy variables for our later empirical estimation. These treatments enables us to utilize the **ASIF** dataset with multiple purposes, for instance, we can generate all location codes into regional, provincial and urban dummies, so as take the registration type into variables as 2-digit, 4-digit and 6-digit industry specifications.

#### **4.4 Further discussion of data treatment**

Some noteworthy drawbacks in the **ASIF** dataset need further discussions. We believe these characteristics are partial reasons causing the estimates' standard errors comparatively large and less converging in our later empirical tests. The first is that the manufacturing firms covered in the sample period increased dramatically since the year 2004. Apart from more and more firms having annual sales reached the official statistical category, the year 2004 was an industry census year, there was more comprehensive survey coverage in that year, which may explain the jump from 2003 to 2004 in the number of firms ([Lu and Tao 2009](#)). The second is the ASIF does not cover small non-state-owned firms with annual sales less than five million Yuan, which could cause the sample estimation upward biased. The third and most challenging problem is that **ASIF** does not provide organization relation information among multi-plant firms. We could only recognize all dataset as individual plant and ignore the situation that enterprises having more than one plants in different regions. The disaggregate composition of plant TFP could not review some multi-plant firms real performance.

Comparing with researches applied with the same data source, this research designed not delete firms with zero monetary inputs or outputs (gross assets, net sum of fixed capital value, sales, gross output) or employment less than 10 persons ([Jefferson et al. 2008](#); [Cai and Liu 2009](#)). The endogeneity issue of firm behavior is our major concern. We are arguing that if researchers need to ob-

serve firms endogenous behavior, henceforth estimate their self-adjustments in capital and labor investment and yearly middle inputs from year to year, the zero monetary accounting is useful and sensitive in indicating their entry and exit dynamics. Since we assume firms are aware of their productivity changes, so as the profitability, there is less solid ground to assume they have static decision making for each year's productivity shock. [Levinsohn and Petrin \(2003\)](#) proposed method on firm-level productivity estimation only requiring middle input information as the state variable of capital stock. Here we apply the L-P methods to observe firm dynamics of market entry-exit in each year. To provide the analytical benchmark, we also apply the pooled OLS and Fixed effect panel data methods. The purpose of applying these estimation methods are solely for seeking consistent and robust results for accounting firm level capital stock and productivity estimation.

For the computation of total factor productivity, gross production value, net sales of the plants, investment, middle inputs and all other monetary variables were deflated using price deflators (1978 as the benchmark year).

#### **4.5 Firm Productivity Estimation**

There are different methods in measuring productivity. In this paper, we apply econometric approach to observe the productivity change and its sources. The plant-level estimates of TFP are computed using the LP method that was first proposed in [Levinsohn and Petrin \(2003\)](#). See the appendix for details on this method. The LP method has several advantages over the other methods, such as the pooled OLS method, the fixed-effect method, etc. (see [Sun et al. \(2011\)](#) for detailed comparisons about these methods as well as the OP method proposed in [Olley and Pakes \(1996\)](#)). Primarily, the assumption of constant returns to scale is not required here. And, perfectly competitive market is not realistic to developing countries, and factors are not compulsorily be paid their marginal product.

### **5 Econometric results**

We regress Equation (35) in two steps. In the first step, we apply the Probit model to investigate how agglomeration affects firms' exporting choices by con-

troling other factors. In the second step, we examine the influence of agglomeration on exporters' foreign sales.

In our empirical analysis, a region unit is a prefecture-level city. A region's area  $A$  is the area of its central district. The density of the employment  $dens$  in a city is measured by the density of workers employed in manufacturing industries in it. To control some other factors that may affect firms' exports, we involve firms' ownership structure (classified by *Stock dummy*), belonging relationship (classified by *Belonging dummy*), size levels (classified by *Size dummy*) and their operating status (classified by *Status dummy*) for firm-level characteristics (which are defined by China Statistical Bureau in the original dataset), and human capital (denoted by *humancapital*) and fiscal expenditure (denoted by *fiscalexpend*) for city-level characteristics, and institute level (denoted by *institute*) for province-level one. We also consider the influence of industry and region that firms locate on their exports, which is controlled by industry- and region- dummy (classified by *Industry dummy* and *Region dummy*, respectively). In our baseline empirical analysis, firms' total factor productivity is estimated applying LP (Levinsohn-Petrin method), which we denote by *lpbyind*. In the sequel, we will use productivity estimated using OLS method (denoted by *olsbyind*) to do robust analysis. In the former case, the cross term between productivity and agglomeration is denoted by *lpdens*, while it is *olsdens* in the latter one.

## 5.1 Agglomeration and exporting choice

Table 5 shows the six estimation results in the first step by controlling the city-level, province-level, firm-level, and industry- and region-level characteristics. It shows that the coefficients of  $\ln dens$  and  $\ln dens^2$  are both positive and significant at 1% level. This implies that agglomeration has a significant influence on firms' exporting choices. In the former regression results, the influence of agglomeration on firms' exporting choices decreases first and then increases (i.e., a Parabola-shape effect). This implies that agglomeration has positive influences on firms' exporting choices for some Chinese cities while negative ones for some other cities, by noting that the minimum and the maximum of the variable  $\ln dens$  are  $-3.9376$  and  $-0.2012$ , respectively, while those of *lpbyind* are  $-2.8707$  and  $5.5220$ , respectively. This verifies Proposition 2. While it becomes

positive when industry- and region-level characteristics is controlled. This may be due to that competition and thus selection effect overtakes positive agglomeration externality in the first agglomerating stage and then the latter exceeds the former ones in the second agglomerating stage.

We can also see that this result is robust when controlling different qualitative variables, such as industry, region, ownership dummies, etc. However, the coefficient of the cross term of  $lndens$  and  $lpbyind$  is not significant. This result may be caused by the correlation between the two variables as agglomeration results in positive or negative externality on firms' productivity.

One other interesting result deserve noting here. We see from Table 5 that city size has also a significant influences on firms' exporting choices. This can be derived by revising our assumption that each region's area is 1. This result implies that city size may interact with firms' characteristics (such as productivity), which results finally in their more possible exporting behaviors.

## 5.2 Agglomeration and exporting sale

Table 6 shows the estimation result in the second step. We can see directly that agglomeration has a similar Parabola-shape influence on firms' foreign sale as it does to their exporting choices in all regression equations (1)-(6). This implies that agglomeration has a positive influence on firms' exporting sales for all Chinese cities by recalling the scope of  $lndens$  and  $lpbyind$ . However, we can also see from this table that the coefficient of the cross term of  $lndens$  and  $lpbyind$  is not significant either, whose rationale is the same to the above. Interestingly, the influence of city size has also a positive and significant effect on firms' exporting sales.

## 5.3 Robust analysis

Table 7 and Table 8 show the regression results on the influence of agglomeration on firms' exporting choices and sales with their productivity estimated using OLS method. The results shown in the two table are much similar to those given in Table 5 and Table 6, which takes on a stable Parabola-shape relationship between firms' exporting behaviors and agglomeration.

Table 9 and Table 10 show the empirical results estimated using Tobit ap-

proach, which confirm further the Parabola-shape influencing pattern of agglomeration on firms' exporting sales. However, similar to those in Table 5, 6, 7 and 8, the heterogeneous export effect of agglomeration is not significant even at 10% significance level. The rationale is due to the intrinsic influence of agglomeration on firms' productivity. In the next section, we will investigate whether this effect is positive and significant by considering this intrinsic influence and by controlling different agglomeration patterns.

## 6 Export premium of agglomeration under different agglomeration patterns

In the former one-step estimation, we did not deal with the endogeneity between agglomeration and firms' productivity. However, firms' productivity estimated using the TFP estimation methods (Pooled OLS, Fixed Effects, and Levinshon and Petrin) is not the one ex ante given prior to the spatial agglomeration. Instead, it is the one influenced ex post by agglomeration patterns. This endogeneity issue could jeopardize the validity of estimation of the effect of agglomeration on firms' exports. To identify the real effect of agglomeration on trading activities, we following the two-step regression methods in two stage regressions.

### 6.1 First stage regression on industrial level total factor productivity and net industrial amenity effect

The effect of agglomeration on firms' total factor productivity can be estimated by the following equation:

$$TFP_{rjit} = \phi \ln dens_{rjt} + \psi \ln A_{rt} + \delta_1 \ln fiscalexpend_{rt} + \delta_2 institute_{rt} + \varphi_1 stock_{rjit} + \varphi_2 belong_{rjit} + \varphi_3 status_{rjit} + \varphi_4 size_{rjit} + \alpha_r + \beta_j + \gamma_i + \eta_t + cons + \epsilon_{rjit}$$

by controlling the regional-, industry-, firm-, and time- fixed effects  $\alpha_r, \beta_j, \gamma_i$  and  $\eta_t$ , where  $fiscalexpend_{rt}$ ,  $institute_{rt}$ ,  $stock_{rjit}$ ,  $belong_{rjit}$ ,  $status_{rjit}$  and  $size_{rjit}$  are, respectively, fiscal expenditure and institution index in region  $r$  in year  $t$ , and ownership (state-owned, collective owned or private owned, etc.), be-

longing relationship (referring to whether a firm belongs to the central government, a local one, etc.), operating status (newly established, incumbent, exiting/ closed ), and size dummies of firm  $i$  in industry  $j$  in region  $r$  in year  $t$ . we estimate the above model and then generate the predicted value of  $\widehat{TFP}_{rjit}$ . The above estimation assumes that a firm's productivity is determined by all the region-, industry-, firm- and time- combination effects. Ruling out other factors, such as regional agglomeration patterns and intra- and inter industries activities, we suppose that a firm's inborn productivity is determined only by its inherent characteristics (ownership structure, operating status, size, belonging relationship, etc.) as well as local **amenities** . Therefore, the error term  $\epsilon_{rjit}$  was designed to explained the ex post exogenous externalities of local amenities to individual firms, agglomeration patterns in particular.

As were predicted in Proposition 1 and 2 and asserted in Corollary 1 and Hypothesis 1, agglomeration significantly affects firm's productivity and thus diverts their exporting decisions both in terms of exporting possibilities and exporting sales. Hence, the net effects of local amenities such as agglomeration on firms' productivity can be identified from the above first stage estimation. We generate the residual from the TFP equation and then the cross-term between it and local employment density, as follows

$$Amenity_{rjit} \triangleq TFP_{rjit} - \widehat{TFP}_{rjit}, Interact_{rjit} = Amenity_{rjit} \times \ln dens_{rjt}$$

where  $Amenity_{rjit}$  and  $Interact_{rjit}$  correspond to  $\vartheta \ln \theta_{rjit}$  and  $\lambda \ln dens_{rjt} \ln \theta_{rjit}$  in (35), accordingly. Hence, the net effect of agglomeration on firms' productivity can be identified. The effects of spatial agglomeration on their exporting behaviors are identified in the following second stage.

## 6.2 Agglomeration patterns, firms' productivity and their exporting behaviors

From Table 11 (estimation in the first and the second step) , it is straight forward that urban employment density has a positive effect on firms' productivity. So does the region-specific effect, such as local government fiscal expenditure, institution quality, city area, etc. If we take local government expenditure and institution quality as the proxy to represent local infrastructure, then the estima-

tion result shows that the higher quality of local infrastructure brings forward positively higher spillovers on firms' productivity.

Controlling industry dummies as well as firm-specific effect, stock, belonging relationship, operating status and size dummies, the first-stage estimation provides us an opportunity to inspect the pre-assumption that local employment density and other regional effects have positive influences on firms' productivity. Furthermore, we can rule out those "unobserved" effects influencing industrial productivity by the Solow residual approach. This implies that coefficient of local employment density for instance in the first stage estimation could be over-estimated without considering the dynamics of "firm" choices. For example, larger and more productive firms could self-select into larger cities and thus receive higher knowledge spillovers. How firms interact with others is thus crucial for their productivity growth. That is to say, suppose that the first stage estimation's right hand size variables are exogeneously given, agglomeration patterns matters!

To start our second-stage estimation, it's of sense for us to get some intuition on how agglomeration patters as well as local amenities on firms' productivity and exports.

A summary of the relationship between agglomeration, firms' exports and their productivity is given in Table 1, 2 and 3. It shows that agglomeration significantly contributes to firms' the urban export rate with different agglomeration indicators. Conceptually, we have

$$\mathbb{X}_{rjit} = \beta_0 + \beta_1 G(A),$$

where  $G(A)$  includes the urban employment density (10 thousands workers per square of kilometer), localization economy, urbanization economy, diversification economy and competition economy accordingly. Without considering the productivity-agglomeration interaction terms, the primary investigation of agglomeration effects on trade in cities are tested.

By introducing various agglomeration patterns, Table 1, 2 and 3 indicated that agglomeration has positive effects on average urban export volume as well as the numbers of exporting firm among all active firms, where *loc*, *urb*, *div* and *com* measure, respectively, localization economy, urbanization economy, diversification economy and competition economy, whose measurement is shown



in Section 3.2. Though the first-step analysis of the effect of agglomeration on the rate of exporters in city does not consider the quadratic effect of agglomeration and other control variables for firm- and city- specific effects, the preliminary results still indicates that high agglomeration has significant effect on firms' exports.

In order to provide a mechanism for the above analysis, we consider the correlations among agglomeration, productivity and export.

We calculate firms' average export in each city value by dividing total exporting firms' export by the total firm number in it, as follows

$$\text{Average export value}_{rt} = \frac{\text{Total export value}_{rt}}{\text{Total firm number}_{rt}}$$

where Average export value<sub>rt</sub>, Total export value<sub>rt</sub> and Total firm number<sub>rt</sub> are average, total exporting values, and total number of firms in region  $r$  in year  $t$ , respectively.

The fitted line between average employment density and average export value across cities takes on a U- configuration as the employment density increases, which indicates that the rising of labor agglomeration is highly correlated with local export, as shown in Figure 1. It shows that firms benefit from externalities of increasing labor pool. The intriguing fact is that firms' average export is not increasing with urban employment density, though it increases with the specialization of labor market.

Likewise, other agglomeration patterns identifying different sources of agglomeration all are positively correlated with average firm export. We are particularly interested in three aspects: (1) the shape of the fitted line between agglomeration economy and average firm export, (2) the different curvature between average firm export and different sources of agglomeration economy, and (3) the fitness of the above mentioned correlations.

Without losing generosity, Figure 2 shows that agglomeration economies, including localization, urbanization, diversification and competition economy as several specific patterns, take on a U- shape effect on average firm export in cities.

If we examine closely the influencing configuration of each agglomeration pattern on average firm export in cities, it is apparent that the diversification economy is less steep than others. That may be because exporting firms are

less likely to enhance its' horizontal linkages with other industries, which would directly lead to increasing competition both in the factor and product market. and decreasing markup. Therefore, exporting firms could enjoy longer period of externality from localization and urbanization economy while the diversification and competition of industries have higher selection thresholds for high-productivity firms.

As diversification pattern of agglomeration is designed to indicate the relative strength linking with other industries, the fitness of urban diversification economy and average firm export is less converged or fitted. The increasing variance of diversification level would lead average firm export to fluctuate dramatically according to the firm-, industry-, and region- specific effects. Therefore, further analysis controlling the influencing variables are immediately necessary.

Another descriptive summary of agglomeration and trade can be illustrated by the correlation between the ratio of urban exporting firm numbers and various source of agglomeration economies. We calculate the ratio of urban exporting firm numbers which is divided by the total firm numbers located in the same city, and graph the correlation with agglomeration economies, shown as Figure 3 and Figure 4. They all have the same implication as the above tables, and thus we ignore their detailed illustration.

### 6.3 Second stage regression on agglomeration patterns and firm export preferences

Combining the estimated net amenity effects and industry-, region- and cross-effect, we estimate the influence of agglomeration patterns on individual firms' trading activities with the following model, which was previously given as the exogenous factors

$$\begin{aligned} \mathbb{X}_{rjit} = & \phi \ln dens_{rjt} + \varphi \ln dens_{rjt}^2 + \psi \ln A_{rt} + \vartheta Amenit_{rjt} + \lambda Interact_{rjt} \\ & + \xi Loc_{rjit} + \rho Urb_{rjt} + \varrho Div_{rjt} + \chi Comp_{rjt} + \beta_{jt} + \gamma_1 Stock_{rjit} + \gamma_2 Belong_{rjit} \\ & + \gamma_3 Status_{rjit} + \gamma_4 Size_{rjit} + \varepsilon_{rjit}. \end{aligned}$$

According to Equation (34), the effect of agglomeration on firms' export can be divided into two terms. The first is common for all the firms in the same

region, which is called the "**common export effect of agglomeration**". The second is idiosyncratic across firms as their productivity is heterogeneous, which is called the "**heterogeneous export effect of agglomeration**". That is to say, we need to specify and testify Proposition 3, 4 and Hypothesis 1. That is, the exporting elasticity of productivity is increasing in the scale of all the regions, and agglomeration in region  $j$  increases each incumbent's export without considering its initial productivity level.

Table 11 (column 2-8) and 12 (column 2-8) shows that the agglomeration has a Parabola-shape effect on firm exporting choices and exporting sales with interesting and different interpretations, as are predicted in Corollary 1 and Hypothesis 1. We divide the differences into two categories.

### **Common export effect of agglomeration**

Dilation effect of agglomeration on firms' exporting choices can be found in various estimations combining different controlling variables. Agglomeration, taken employment density, urbanization, localization, specialization and competition indexes as proxies, exerts uniformly a contributing source for firms' productivity, and then enhances firms' exports, which thus verifies Proposition 4. Such dilation effect directly increases firms' exporting probability and sales.

In particular, Table 11 (model (1) and (3) in the second stage estimation) and Figure 5 show that employment density has a quadratic effect on firms' exporting probability, which takes on a consistent U- shape like previous estimation results predicted in Corollary 1. so is implied in Table 12. It is decisive that the common effect of agglomeration benefits all firms as the home market effects implies. Also, such common effect is the whole "package" of agglomeration without considering the inner-city agglomeration patterns, which the estimated coefficients are over-estimated.

### **Heterogeneous export effect caused by Agglomeration patterns and interactions**

Controlling the agglomeration patterns within the direct effect of agglomeration, the second stage estimation models (model 4-7 in Table 11) and model 1-7 in Table 12) indicate that as the direct effect of agglomeration decreases with the

rising agglomeration patterns contributing to the exporting choices and sales simultaneously.

As shown in Figure 6, it is interesting that diversification and competition effects have negative effects on firms' exporting choices, which is different from previous literatures in urban economics. The more diversified is the local market, the less backward and forward linkages are provided for local firms' exporting choices. As firms could not find substantial firms dedicating on the up- and down-stream production and services, the exporting conditions are comparatively poor. Also, local industrial competition pushing low productivity firms out of the market. Only firms with high productivity could survive in the increasing market competition, which is consistent with the results asserted in the Melitz model (Melitz, 2003). The negative effects of diversification and competition to exporting dynamics are counter-factual to classic predictions in urban economic theories but realistic in emerging nations with market segmentations. On the other hand, the localization and urbanization economies are all promising factors to most firms' exporting choices. Different with diversification and competition indices that measure industry- and region- correlation, localization and urbanization indices measure the local "firm-industry" distributional differences. Firms benefit from local home market effects instead of from other effects. The more fragmented or distorted is the local market, the more protection or less competition firms are faced with. Firms are conscious about the labor pool effects and market size (employment density and region size as the proxies). However, how they rank themselves within the particular market is more important than how strong is competition they are faced with. Consistent with Proposition 3, the empirical estimation implies some crucial findings that firms enjoy more from the home-market effects which were "localized" and fragmented by heterogeneous locations if the national/regional income and population are given constant. Such findings are explicitly revealing the trade-geography mechanism caused by spatial competition. The more protection and isolated is a region, the more benefits firms receive in stimulating their production and exporting propensities in it.

Another supporting finding is that the estimation results of the models in the second stage (model 2,5,7) all indicate that interactions ( $Interact_{rjt}$ ) between firms' inborn characteristics (such as inherent productivity) and local agglomeration have positive effects on export propensities. This finding is interesting

as it's not revealed in the former four propositions, Corollary 1 and Hypothesis 1. Moreover, Table 12 shows that the parameters  $\vartheta$  and  $\lambda$  in Equation 36 satisfies  $\vartheta/\lambda > 0$ . This implies that firms with higher productivity are influenced more by agglomeration, as is predicted in Proposition 3. The intuition embedded in the interaction term implies that firms productivity growth and agglomeration status together have the "chemical effect" in production growth, there are two-way causalities among productivity and agglomeration effects.

## 7 Conclusion

Instead of classic literature on trade induced agglomeration (TIA) studies, our paper investigates the influences of agglomeration on heterogeneous firms' exporting behaviors (AIT).

The paper proposes a model to investigate the influences of agglomeration on heterogeneous firms' exporting behaviors. Competition and thus selection effect caused by agglomeration forces less productive firms to exit the market while agglomeration externalities increase firms' productivity and decreases industrial fixed entry, fixed and variable exporting costs, and effective labor wage. The former factors decrease while the latter increase firms' exporting possibilities and sales. The model shows that the composite effect of agglomeration on firms' exports takes on a Parabola-shape pattern. Moreover, higher-productivity firms benefit more export premium from agglomeration, which explains why larger and more productive firms in larger cities are more possible to export and exports more. Empirical results based on data from Chinese Industrial Enterprises between 1998 and 2007 verify the theoretical results. The paper also investigates the influences of different agglomeration patterns on firms exports, including home market effect, urban economies and competition effect and diversification effect. It shows that the former two patterns exert a positive while the latter two have a positive influence on firms' exporting behaviors.

The timing for firms making exporting decisions are of two steps. Firstly, firms decide which location to start up their production facing sunk costs whose productivity levels are heterogeneous. This setting is sufficient to explain the spatial heterogeneity on firms exporting decisions. As pre-assumed, firms are consciously aware of externalities in potential locations. Such externalities can

be explained as local labor pool effects, home market effect, and regional amenities in terms of fiscal expenditure, etc. These "X factors" are common to all firms, which provide positive influences on aggregate industrial productivity.

Secondly, after firms enter the local market, the story of technological spillover is no longer new to each new-entry or incumbent firm. The previous "unobserved" local market magnification effects enlarge the demand and supply premier for firms, while there are new decisions for firms to make. After large and productive firms paid their sunk costs to enter into large cities that have higher level of productivity spillovers and labor pools, they make their decisions on how they interact with other firms and industries. How firms compete and self-organize themselves become the "neo-X factor" for productivity growth and exporting dynamics. That is to say, agglomeration patterns matters exogenously given local home market effect for "New Firms in Town!"

In all, firm-level exporting dynamics is threefold. Firm-level productivity determines its initial preference of export, local market effect increases firms' productivity and export potential while the self-organizing of firms - agglomeration patterns- enhances the selection and sorting mechanism for individuals firms. The first and third dynamics are firm-made, while the second is more or less exogenous to firms.

The policy suggestion is nothing fancy but straight forward. Providing supportive local amenities and services to attracting firms is important for a local government. How firms interact with each other is as private as their heterogeneous nature could be less intervened by government policies.

## References

- Antras, Pol and Elhanan Helpman (2004). Global Sourcing, *Journal of Political Economy*, 112 (3), 552C580.
- Anwara, Sajid and Lan Phi Nguyen (2011). "Foreign Direct Investment and Export Spillovers: Evidence from Vietnam," *International Business Review*, 20(2): 177-193.
- Alfaro, Laura and Maggie Xiaoyang Chen (2010). "The Global Agglomeration of Multinational Firms," mimeo.
- Amiti, Mary (2005). "Location of Vertically Linked Industries: Agglomeration versus Comparative Advantage," *European Economic Review*, 49(4), 809-832.
- Baldwin, Richard E. and Toshihiro Okubo (2005). "Agglomeration and the Heterogeneous Firms Trade Model," mimeo.
- Baldwin, Richard E. and Toshihiro Okubo (2006). "Heterogeneous Firms, Agglomeration and Economic Geography: Spatial Selection and Sorting," *Journal of Economic Geography*, 6(3): 323-346.
- Baldwin, J. R., D. Beckstead, W. M. Brown and D. Right (2007). "Urban Economies and Productivity, Economic Analysis (EA) Research Paper Series 2007045e, Statistics Canada, Analytical Studies Branch.
- Bernard, A. and J. Bradford (1999). "Exceptional Exporter Performance: Cause, Effect, or Both," *Journal of International Economics*, 47(1): 1-25.
- Brandt, L., J. Van Biesebroeck, and Y. Zhang (2011). Creative Accounting or Creative Destruction Firm-level Productivity Growth in Chinese Manufacturing, forthcoming in *Journal of Development Economics*.
- Cafiso, Gianluca (2010). "Trade Costs and the Agglomeration of Production," mimeo.
- Cai, Hongbin and Qiao Liu (2009). Competition and Corporate Tax Avoidance: Evidence from Chinese Industrial Firms, forthcoming in *Economic Journal*.
- Cassey, Andrew J. and Katherine N. L. Schmeiser (2010). "The Agglomeration of Exporters by Destination," mimeo.



- Chaney, Thomas (2008). "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 98(4): 1707-1721.
- Chen, Maggie Xiaoyang and Michael O. Moore (2010). "Location Decision of Heterogeneous Multinational Firms," *Journal of International Economics*, 80, 188-199.
- Chyi, Y.-L., L. Yi, E.-S. Lin and S. Y. Wu (2010). "Do Local Industrial Agglomeration and FDI to China Enhance the Productivity of Taiwanese Firms," mimeo.
- Combes, P., Duranton, G., Gobillon, L. (2008). "Spatial Wage Disparities: Sorting Matters!," *Journal of Urban Economics*, 63: 723-742.
- Combes, P., Duranton, G., Gobillon, L., Roux, S. (2010). "Estimating Agglomeration Economies with History, Geology, and Worker Effects," in Glaeser E.L. (eds.), *Agglomeration Economics*, NBER, Cambridge, MA.
- Combes, P., Duranton, G., Gobillon, L. (2011a). "The Identification of Agglomeration Economies," *Journal of Economic Geography*, 11(2):253-265.
- Combes, P., Laurent Gobillon, Diego Puga, and Sebastien Roux (2011b). "The productivity advantages of large cities: : Distinguishing agglomeration from firm selection," *Econometrica*, forthcoming.
- Duranton, Gilles and Michael Storper (2005). "Rising Trade Costs? Agglomeration and Trade with Endogenous Transaction Costs," CEP Discussion Paper No 683.
- Eaton, Jonathan, Samuel Kortum, and Francis Kramarz (2011). "An Anatomy of International Trade: Evidence from French Firms," *Econometrica*, 2011, 79(5), 1453-1498.
- Forslid, Rikard and Gianmarco I.P. Ottaviano (2002). "Trade and Agglomeration: An Analytically Solvable Case," mimeo.
- Fujita, M. and Hideaki Ogawa (1982). "Multiple Equilibria and Structural Transition of Non-monocentric Urban Configurations," *Regional Science and Urban Economics*, 12(2), 161-196.

- Gao, T (2003). "The Impact of Foreign Trade and Investment Reform on Industry Location: the Case of China," *Journal of International Trade and Economic Development*, 11 (4): 367-86.
- Gatto, Massimo Del, Gianmarco I. P. Ottaviano and Marcello Pagnini (2008). "Openness to Trade and Industry Cost Dispersion: Evidence from a Panel of Italian Firms," *Journal of Regional Science*, 48(1), 97-129.
- Gatto, M.D., A.D. Liberto, and Petraglia, C. (2008). "Measuring Productivity," *CRENoS*, working paper 2008/19.
- Ge, Ying (2006). "Globalization and Industry Agglomeration in China." *World Development*, 37(3), 550-559.
- Griliches, Z. and Mairesse, J. (1995). "Production Functions: the Search for Identification," *NBER Working Paper*, No. 9617.
- Harasztosi, Peter and Gabor Bekes (2009). "Agglomeration Economics and Trading Activity of Firms," mimeo.
- Helpman, Elhanan, Mark J. Melitz, and Stephen R. Yeaple (2004). Export versus FDI with Heterogeneous Firms, *American Economic Review*, 2004, 94(1), 300C316.
- Helpman, Elhanan, Mark J. Melitz, and Yona Rubinstein (2008). Estimating Trade Flows: Trading Partners and Trading Volumes, *Quarterly Journal of Economics*, 2008.
- Jefferson, Gary, Thomas G. Rawski and Yifan Zhang (2008). Productivity Growth and Convergence Across China's Industrial Economy, *Journal of Chinese Economic and Business Studies*, 6(2), 121-140.
- Kang, Youngho (2011). "Is Agglomeration a Free Lunch for New Exporters? Evidence from Chile." mimeo, University of Colorado.
- Kneller, R. and M. Pisu (2007). "Industrial Linkages and Export Spillovers from FDI," *World Economy*, 30(1): 105-134.
- Koenig, Pamina (2009). "Agglomeration and the Export Decision of French Firms." *Journal of Urban Economics*, 66, 186-195.

- Krugman, Paul (1991). "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99(3), 483-499.
- Levinsohn, J. and A. Petrin (2003). Estimating production function using inputs to control for unobservables, *Review of Economic Studies*, 70, 317-342.
- Lin, P., Z. Liu, and Y. Zhang (2009). Do Chinese domestic firms benefit from FDI inflow-star, open- Evidence of horizontal and vertical spillovers-back, *China Economic Review*, 20(4), 677-691.
- Long, Cheryl and Xiaobo Zhang (2011). " Cluster-Based Industrialization in China: Financing and Performance," *Journal of International Economics*, 84(1), 112-123.
- Lovely, M. E., S. S. Rosenthal and S. Sharma (2005). "Information, Agglomeration, and the Headquarters of U.S. Exporters," *Regional Science and Urban Economics*, 35(2): 167-191.
- Lu, J. and Z. Tao (2009). Trends and determinants of China's industrial agglomeration, *Journal of Urban Economics*, 65, 167-180.
- Lucas, Robert E. JR. and Esteban Rossi-Hansberg (2002). On the Internal Structure of Cities, *Econometrica*, 70(4), 1445-1476.
- Martin, Philippe, Thierry Mayer and Florian Mayneris (2008). "Spatial Concentration and Plant-level Productivity in France." *Journal of Urban Economics*, 69, 182-195.
- Maticchioni, A. and P. Naticchioni (2011). "Wage Distribution and the Spatial Sorting of Workers," *Journal of Economic Geography*, volume, issue, pp.
- Melitz, Marc J., 2003, "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71(6), 1695-1725.
- Melitz, Marc J. and G. Ottaviano, Market Size, Trade, and Productivity, *Review of Economic Studies*, 2008, 75, 295C316.
- Mion, G., Naticchioni, P. (2009). "The Spatial Sorting and Matching of Skills and Firms," *Canadian Journal of Economics*, Canadian Economic Association, 42(1):28-55.

- Moller, J., Haas, A. (2003). "The Agglomeration Differential Reconsidered: an Investigation with German Micro Data 1984-1997," in Broecker, J., D. Dohse, R. Soltwedel (eds.), *Innovation Clusters and Interregional Competition*. Berlin: Springer.
- Olley, G.S., A. Pakes (1996). "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64(6): 1263-1297.
- Ottaviano, G., J. Thisse and T. Tabuchi (2002). "Agglomeration and Trade Revisited," *International Economic Review*, 43, 409-436.
- Okubo, Toshihiro (2007). "Trade Liberalisation and Agglomeration with Firm Heterogeneity—Forward and Backward Linkages," mimeo.
- Puga, Diego and Anthony J. Venables (1998). "Agglomeration and Economic Development: Import Substitution vs. Trade Liberalisation," mimeo.
- Saito, Hisamitsu, Munisamy Gopinath, and JunJie Wu (2011). "Heterogeneous Firms, Trade Liberalization and Agglomeration." *Canadian Journal of Economics*, 44(2), 541-560.
- Sidorov, Alexander V. (2011). "The Impact of Exogenous Asymmetry on Trade and Agglomeration in Core-Periphery Model," mimeo.
- Sjoberg, Orjan and Fredrik Sjöholm (2002). "Trade Liberalization and the Geography of Production: Agglomeration, Concentration and Dispersal in Indonesia's Manufacturing Industry," Stockholm School of Economics Working Paper Series Vol. 2002-21.
- Strajer Madsen, E., V. Smith and M. Dilling-Hansen (2003). "Industrial Clusters, Firm Location and Productivity: Some Empirical Evidence for Danish Firms," Working Papers 03-26, University of Aarhus, Aarhus School of Business, Department of Economics.
- Sun, Churen, Guoqiang Tian and Tao Zhang (2011). When Melitz Meets Pareto. MPRA Working Paper No.35597, <http://mpra.ub.unimuenchen.de/35597/>.
- Van Biesebroeck, J. (2005). "Exporting raises productivity in sub-Saharan African manufacturing firms," *Journal of International Economics*, 67(2):373-391.

- Van Biesebroeck, J. (2007). "Robustness of productivity estimates," *The Journal of Industrial Economics*, 55(3):529-569.
- Van Biesebroeck, J. (2008). "Aggregate and decomposing Productivity," *Review of Business and Economics*, LIII(2):122-146.
- Venerables, Anthony (1996). "Equilibrium Locations of Vertically Linked Industries," *International Economic Review*, 37, 341-359.
- Wheeler, C.H. (2004). "Wage Inequality and Urban Density," *Journal of Economic Geography*, 4: 421-437.
- Wheeler, C.H. (2007). "Industry Localization and Earnings Inequality: Evidence from US Manufacturing," *Papers in Regional Science*, 86(1):77-100.
- Yilmazkuday, Hakan (2011). "Agglomeration and Trade: State-level Evidence from U.S. Industries," *Journal of Regional Science*, 51, 139-166.

## Appendix

### The Levinsohn-Petrin Approach

Since the firm's asymmetry knowledge of its productivity is unavailable to the econometrician, the problem of simultaneity will affect firm's endogenous decision on hiring and investment factor inputs. This will lead the pooled OLS estimation of a production function to estimates of the coefficients of exogenous inputs that are biased upwards.

The LP method proposed an alternative for firm-level data estimation which requires no further information about input values, nor subtracting them from the gross-output number to get value added. Since the investment proxy is only valid for plants reporting nonzero investment, firms with "zero investment" are likely to be dropped in previous approach. Instead, L-P method uses intermediate input proxies avoids truncating all the zero investment firms. In many empirical studies (so as in our **ASIF** dataset), firms always report positive use of intermediate inputs like electricity or materials.

Start with the Cobb-Douglas production technology

$$y_t = \beta_0 + \beta_l l_t + \beta_k k_t + \beta_m m_t + \omega_t + \eta_t,$$

where  $y_t$  is the logarithm of the firm's output, such as value added;  $l_t$  and  $m_t$  are the logarithm of the freely variable inputs labor and the intermediate input; and  $k_t$  is the logarithm of the state variable capital.

The error has two components: the transmitted productivity component given as  $\omega_t$  and  $\eta_t$ , an error term that is uncorrelated with input choices. They key difference between  $\omega_t$  and  $\eta_t$  is that the former is a state variable and impacts the firm's decision rules. It is not observed by the econometrician, and it can impact the choices of inputs, leading to the simultaneity problem in production function estimation.

Demand for the intermediate input is assumed to depend on the firm's state variables  $k_t$  and  $\omega_t$ :

$$m_t = m_t(k_t, \omega_t).$$

In the [Levinsohn and Petrin \(2003\)](#) assumption, demand function is mono-

tonically increasing in  $\omega_t$ . This allows inversion of the intermediate demand function,  $\omega_t$  can be written as a function of  $k_t$  and  $m_t$ :

$$\omega_t = \omega_t(k_t, m_t).$$

The unobservable productivity term is now expressed solely as a function of two observed inputs.

A final identification restriction follows [Olley and Pakes \(1996\)](#), which productivity is governed by a first-order Markov process:

$$\omega_t = E[\omega_t | \omega_{t-1}] + \xi_t,$$

where  $\xi_t$  is an innovation to productivity that is uncorrelated with  $k_t$ , but not necessarily with  $l_t$ .

For the value-added production function, it can be written as

$$v_t = \beta_0 + \beta_l l_t + \beta_k k_t + \omega_t + \eta_t = \beta_l l_t + \phi_t(k_t, m_t) + \eta_t,$$

where

$$\phi_t(k_t, m_t) = \beta_0 + \beta_k k_t + \omega_t(k_t, m_t).$$

Substituting a third order polynomial approximation in  $k_t$  and  $m_t$  in place of  $\phi_t(k_t, m_t)$ , makes it possible to consistently estimate parameters of the value-added equation using OLS as

$$v_t = \delta_0 + \beta_l l_t + \sum_{i=0}^3 \sum_{j=0}^{3-i} \delta_{ij} k_t^i m_t^j + \eta_t,$$

where  $\beta_0$  is not separately identified from the intercept of  $\phi_t(k_t, m_t)$ . As the first stage of estimation routine from [Levinsohn and Petrin \(2003\)](#), estimates of  $\beta_l$  and  $\phi_t$  are available.

The second stage of the routine identifies the coefficient  $\beta_k$ . It begins by computing the estimated value for  $\phi_t$  using

$$\hat{\phi}_t = \hat{v}_t - \hat{\beta}_l l_t = \hat{\delta}_0 + \sum_{i=0}^3 \sum_{j=0}^{3-i} \hat{\delta}_{ij} k_t^i m_t^j - \hat{\beta}_l l_t.$$



For any candidate value  $\beta_k^*$ , a prediction for  $\omega_t$  of all periods  $t$  can be computed by

$$\hat{\omega}_t = \hat{\phi}_t - \beta_k^* k_t.$$

Using these values, a consistent (nonparametric) approximation to  $E[\omega_t | \omega_{t-1}]$  is given by the predicted values from the regression:

$$\hat{\omega}_t = \gamma_0 + \gamma_1 \omega_{t-1} + \gamma_2 \omega_{t-1}^2 + \gamma_3 \omega_{t-1}^3 + \epsilon_t,$$

which in [Levinsohn and Petrin \(2003\)](#) is given as  $E[\omega_t | \hat{\omega}_{t-1}]$ .

Given  $\hat{\beta}_0$ ,  $\beta_k^*$  and  $E[\omega_t | \hat{\omega}_{t-1}]$ , [Levinsohn and Petrin \(2003\)](#) writes the sample residual of the production function as

$$\widehat{\eta}_t + \widehat{\xi}_t = v_t - \hat{\beta}_l l_t - \beta_k^* k_t - E[\omega_t | \hat{\omega}_{t-1}].$$

The estimation  $\hat{\beta}_k$  of  $\beta_k$  is defined as the solution to:

$$\min_{\beta_k^*} \sum_t \left( v_t - \hat{\beta}_l l_t - \beta_k^* k_t - E[\omega_t | \hat{\omega}_{t-1}] \right)^2.$$

Table 1: Summary Results of first-step analysis on export firms rate and agglomeration economies

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
exportnumrat							
ln dens	0.00000925 (1.20)						0.0000186* (2.39)
loc		0.000204*** (27.55)				0.0000471*** (4.43)	0.0000729*** (17.67)
urb			0.000206*** (27.58)			0.000285*** (25.23)	0.0000657*** (14.80)
div				- 0.00924*** (-35.68)		- 0.0137*** (-49.04)	-0.00432*** (-37.28)
com					0.0000590*** (43.38)	0.0000574*** (42.07)	- 0.00000184*** (-3.66)
_cons	0.259*** (25955.62)	0.246*** (3455.89)	0.245*** (2585.11)	0.249*** (9388.94)	0.248*** (54644.02)	0.245*** (2566.60)	0.258*** (6723.29)
N	2096116	2226426	2226426	2226426	2226266	2226266	2096020

Note:  $t$  statistics in parentheses, \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 2: Summary Results of first-step analysis on export value and agglomeration economies

averexportvalu(€)	(2)	(3)	(4)	(5)	(6)	(7)	
ln dens	2.105*** (9.05)					2.816*** (11.93)	
loc	4.720*** (51.25)				1.434*** (10.82)	1.512*** (12.10)	
urb		5.412*** (58.29)			5.010*** (35.58)	1.757*** (13.08)	
div			0.102 (0.03)		- 74.55*** (-21.40)	- 46.02*** (-13.10)	
com				0.359*** (21.18)	0.276*** (16.23)	0.0662*** (4.35)	
_cons	3701.1*** (12244.40)	3471.9*** (3914.48)	3448.5*** (2916.85)	3517.3*** (10638.31)	3517.3*** (62149.57)	3447.4*** (2893.17)	3669.4*** (3156.48)
N	2096116	2226426	2226426	2226426	2226266	2226266	2096020

Note: *t* statistics in parentheses, \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 3: Summary Results of first-step analysis on export value standard deviation and agglomeration economies

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
sdexport							
ln dens	121.1*** (25.91)						138.6*** (29.24)
loc		48.16*** (27.96)				11.36*** (4.58)	11.39*** (4.54)
urb			57.12*** (32.88)			55.37*** (21.02)	63.94*** (23.69)
div				-19.14 (-0.32)		-827.5*** (-12.70)	- 2108.9*** (-29.90)
com					5.208*** (16.44)	4.360*** (13.71)	3.826*** (12.53)
_cons	45381.7*** (7474.27)	42423.2*** (2557.35)	42159.6*** (1905.66)	42888.0*** (6941.24)	42885.4*** (40548.91)	42156.0*** (1890.23)	44681.1*** (1913.97)
N	2096116	2226412	2226412	2226412	2226253	2226253	2096020

Note:  $t$  statistics in parentheses, \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 4: The General Information of ASIF dataset

Year	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	Total
<b>Original</b>	165,118	162,033	162,885	169,031	181,557	196,222	279,092	271,835	301,961	336,768	2,226,502
<b>ASIF ids Corrected</b>	165,118	162,032	162,880	169,027	181,556	196,222	279,032	271,833	301,959	336,767	2,226,426
<b>ASIF ids Variables</b>	98	96	97	96	86	74	135	126	128	123	

Table 5: The influence of agglomeration on firms' exporting choices (baseline regression)

	(1)	(2)	(3)	(4)	(5)	(6)
	exportdum	exportdum	exportdum	exportdum	exportdum	exportdum
ln dens	1.191*** (52.31)	1.369*** (59.65)	0.436*** (18.37)	0.190*** (37.53)	0.226*** (44.05)	0.0923*** (14.57)
ln dens <sup>2</sup>	0.244*** (35.59)	0.283*** (41.07)	0.0835*** (12.11)	0.0294*** (20.31)	0.0351*** (23.77)	0.00908*** (5.12)
ln A	0.156*** (16.22)	0.139*** (14.34)	0.116*** (11.23)	0.0737*** (30.46)	0.0772*** (31.43)	0.0603*** (20.17)
lpbyind	0.214*** (8.36)	-0.0231 (-0.89)	-0.389*** (-12.92)	0.0270*** (4.78)	-0.0528*** (-9.15)	-0.149*** (-18.11)
lpdens	0.00352 (0.19)	-0.0272 (-1.48)	-0.181*** (-8.98)	0.00726 (1.91)	0.00249 (0.64)	-0.0345*** (-6.43)
ln fiscalex- pend	0.0324*** (5.02)	0.0795*** (12.17)	0.155*** (22.33)	-0.00343* (-2.27)	0.0129*** (8.36)	0.0291*** (14.90)
ln humcapital	-0.144***	-0.208***	-0.286***	-0.0331***	-0.0539***	-0.0716***

	(-16.09)	(-23.13)	(-30.49)	(-14.78)	(-23.67)	(-26.17)
institute	0.223***	0.228***	0.0662***	0.0451***	0.0424***	0.0205***
	(87.94)	(88.61)	(21.68)	(72.65)	(67.12)	(24.68)
Stock dummy	Yes	Yes	Yes	Yes	Yes	Yes
Belonging	Yes	Yes	Yes	Yes	Yes	Yes
dummy	Yes	Yes	Yes	Yes	Yes	Yes
Status dummy	No	Yes	Yes	No	Yes	Yes
Size dummy	No	Yes	Yes	No	Yes	Yes
Industry	No	No	Yes	No	No	Yes
dummy	No	No	Yes	No	No	Yes
Region	No	No	Yes	No	No	Yes
dummy	No	No	Yes	No	No	Yes
_cons	-4.677***	-3.450***	-5.166***	-1.174***	-1.019***	-1.805***
	(-81.23)	(-57.27)	(-58.11)	(-79.58)	(-66.44)	(-64.74)
lnsig2u						
_cons	2.670***	2.535***	2.157***			
	(575.96)	(520.53)	(405.19)			
N	1930027	1930027	1930027	1930027	1930027	1930027

Note: *t* statistics in parentheses, \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 6: The influence of agglomeration on firms' exporting sales (baseline regression)

	(1)	(2)	(3)	(4)	(5)	(6)
	lnexport	lnexport	lnexport	lnexport	lnexport	lnexport
ln dens	0.101*** (6.77)	0.369*** (23.97)	0.279*** (17.78)	0.251*** (16.07)	0.251*** (16.07)	0.244*** (15.59)
ln dens <sup>2</sup>	-0.0429*** (-12.88)	0.0372*** (10.36)	0.0321*** (8.87)	0.0310*** (8.63)	0.0310*** (8.63)	0.0313*** (8.71)
ln A	0.432*** (59.01)	0.250*** (27.69)	0.181*** (19.29)	0.148*** (15.85)	0.148*** (15.85)	0.143*** (15.24)
lpbyind	0.447*** (24.25)	0.103*** (5.46)	0.0705*** (3.71)	0.0366 (1.92)	0.0366 (1.92)	0.107*** (5.09)
lpdens	0.00376 (0.30)	-0.0420*** (-3.34)	-0.00670 (-0.53)	0.00301 (0.24)	0.00301 (0.24)	0.0172 (1.35)
ln fiscalex- pend		0.187*** (29.61)	0.141*** (21.74)	0.140*** (21.69)	0.140*** (21.69)	0.137*** (21.15)
ln humcapital		-0.237***	-0.165***	-0.181***	-0.181***	-0.182***



			(-33.94)	(-22.40)	(-24.75)	(-24.75)	(-24.76)
institute			0.0664***	0.0594***	0.0539***	0.0539***	0.0513***
			(29.21)	(25.51)	(23.19)	(23.19)	(21.87)
Stock dummy	No	No	Yes	Yes	Yes	Yes	yes
Belong	No	No	Yes	Yes	Yes	Yes	Yes
dummy	No	No	No	Yes	Yes	Yes	Yes
Status dummy	No	No	No	Yes	Yes	Yes	Yes
Size dummy	No	No	No	Yes	Yes	Yes	Yes
Industry	No	No	No	No	No	No	Yes
dummy	No	No	No	No	No	No	Yes
Region	No	No	No	No	No	No	Yes
dummy	No	No	No	No	No	No	Yes
_cons	6.000***	5.366***	5.770***	6.387***	6.387***	6.387***	6.726***
	(220.24)	(86.52)	(80.13)	(88.35)	(88.35)	(88.35)	(9.13)
N	525105	515238	514885	514885	514885	514885	514885

Note: *t* statistics in parentheses, \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 7: The influence of agglomeration on firms' exporting choices (TFP estimated with OLS method)

	(1)	(2)	(3)	(4)	(5)	(6)
	exportdum	exportdum	exportdum	exportdum	exportdum	exportdum
ln dens	1.139*** (24.32)	1.221*** (25.83)	0.404*** (8.45)	0.203*** (18.45)	0.228*** (21.06)	0.0818*** (5.85)
ln dens <sup>2</sup>	0.288*** (41.53)	0.310*** (44.44)	0.0973*** (13.99)	0.0435*** (26.82)	0.0458*** (28.73)	0.0161*** (8.11)
ln A	0.141*** (14.64)	0.140*** (14.37)	0.131*** (12.74)	0.0697*** (26.65)	0.0751*** (29.04)	0.0556*** (17.26)
olsbyind	0.464*** (60.01)	0.391*** (49.60)	0.472*** (60.32)	0.124*** (66.61)	0.103*** (56.00)	0.149*** (63.50)
olstdens	0.0309*** (5.71)	0.0360*** (6.54)	0.00688 (1.26)	0.00599*** (4.71)	0.00543*** (4.35)	0.00265 (1.64)
ln fiscalex- pend	-0.00731 (-1.13)	0.0358*** (5.45)	0.0808*** (11.70)	-0.0164*** (-9.93)	-0.00306 (-1.88)	0.00922*** (4.35)
ln humcapital	-0.120***	-0.177***	-0.244***	-0.0268***	-0.0427***	-0.0596***

	(-13.38)	(-19.52)	(-26.08)	(-10.90)	(-17.65)	(-20.04)
institute	0.223***	0.223***	0.0327***	0.0467***	0.0423***	0.00875***
	(87.16)	(85.94)	(10.77)	(69.13)	(63.31)	(9.64)
Stock dummy	Yes	Yes	Yes	Yes	Yes	Yes
Belong	Yes	Yes	Yes	Yes	Yes	Yes
dummy	Yes	Yes	Yes	Yes	Yes	Yes
Status dummy	No	Yes	Yes	No	Yes	Yes
Size dummy	No	Yes	Yes	No	Yes	Yes
Industry	No	No	Yes	No	No	Yes
dummy	No	No	Yes	No	No	Yes
Region	No	No	Yes	No	No	Yes
dummy	No	No	Yes	No	No	Yes
_cons	-7.353***	-6.209***	-8.620***	-1.917***	-1.672***	-2.840***
	(-89.30)	(-71.66)	(-79.56)	(-92.70)	(-79.24)	(-82.50)
lnsig2u						
_cons	2.526***	2.508***	2.082***			
	(522.37)	(511.87)	(389.89)			
<i>N</i>	1930027	1930027	1930027	1930027	1930027	1930027

Note: *t* statistics in parentheses, \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 8: The influence of agglomeration on firms' exporting sales (TFP estimated with OLS method)

	(1)	(2)	(3)	(4)	(5)	(6)
	lnexport	lnexport	lnexport	lnexport	lnexport	lnexport
ln dens	0.829*** (32.10)	0.843*** (32.06)	0.343*** (12.76)	0.283*** (10.49)	0.283*** (10.49)	0.274*** (10.16)
ln dens <sup>2</sup>	0.0518*** (16.96)	0.0625*** (18.95)	0.0458*** (13.97)	0.0451*** (13.79)	0.0451*** (13.79)	0.0451*** (13.77)
ln A	0.107*** (16.32)	0.0961*** (11.70)	0.0574*** (6.80)	0.0421*** (4.98)	0.0421*** (4.98)	0.0357*** (4.21)
olsbyind	0.575*** (138.37)	0.567*** (134.64)	0.754*** (166.51)	0.762*** (166.32)	0.762*** (166.32)	0.764*** (166.45)
olsdens	-0.0805*** (-29.79)	-0.0770*** (-27.79)	-0.0219*** (-7.72)	-0.0157*** (-5.53)	-0.0157*** (-5.53)	-0.0152*** (-5.33)
ln fiscalex- pend		0.0178** (3.10)	0.0104 (1.79)	0.00369 (0.63)	0.00369 (0.63)	0.00327 (0.56)
ln humcapital		-0.0544***	-0.0465***	-0.0405***	-0.0405***	-0.0427***

			(-8.61)	(-7.03)	(-6.12)	(-6.12)	(-6.44)
institute		0.0119***	-0.000376	-0.000996	-0.000996	-0.000996	-0.00334
		(5.72)	(-0.18)	(-0.47)	(-0.47)	(-0.47)	(-1.58)
Stock dummy	No	No	Yes	Yes	Yes	Yes	Yes
Belong	No	No	Yes	Yes	Yes	Yes	Yes
dummy	No	No	Yes	Yes	Yes	Yes	Yes
Status dummy	No	No	No	Yes	Yes	Yes	Yes
Size dummy	No	No	No	Yes	Yes	Yes	Yes
Industry	No	No	No	No	No	No	Yes
dummy	No	No	No	No	No	No	Yes
Region	No	No	No	No	No	No	Yes
dummy	No	No	No	No	No	No	Yes
_cons	3.305***	3.404***	2.093***	2.096***	2.096***	2.096***	2.964***
	(89.24)	(53.89)	(28.89)	(28.50)	(28.50)	(28.50)	(4.45)
<i>N</i>	525105	515238	514885	514885	514885	514885	514885

Note: *t* statistics in parentheses, \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 9: The influence of agglomeration on firms' exporting sales estimated with Tobit model

	(1)	(2)	(3)
	lexport	lexport	lexport
ln dens	2.306*** (48.08)	2.556*** (53.77)	0.948*** (19.07)
ln dens <sup>2</sup>	0.397*** (30.31)	0.441*** (34.03)	0.120*** (9.00)
ln A	0.666*** (29.29)	0.651*** (28.95)	0.480*** (19.93)
lpbyind	0.333*** (5.57)	-0.304*** (-5.13)	-0.983*** (-14.40)
lpdens	-0.0446 (-1.08)	-0.0949* (-2.31)	-0.343*** (-7.79)
ln fiscalexpend	0.162*** (10.09)	0.283*** (17.74)	0.413*** (24.58)
ln humcapital	-0.490*** (-22.72)	-0.661*** (-30.95)	-0.765*** (-34.27)
institute	0.451*** (72.89)	0.429*** (69.67)	0.159*** (23.02)
Stock dummy	Yes	Yes	yes
Belong dummy	Yes	Yes	Yes
Status dummy	No	Yes	Yes
Size dummy	No	Yes	Yes
Industry dummy	No	No	Yes
Region dummy	No	No	Yes
_cons	-14.93***	-12.49***	-18.18***

	(-100.55)	(-83.47)	(-74.74)
sigma_u			
_cons	11.63***	11.15***	9.516***
	(578.18)	(567.33)	(516.01)
sigma_e			
_cons	3.354***	3.314***	3.359***
	(824.81)	(822.16)	(824.09)
<i>N</i>	1930024	1930024	1930024

---



---

Note: *t* statistics in parentheses, \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$



Table 10: The influence of agglomeration on firms' exporting sales estimated with Tobit model (OLS productivity)

	(1)	(2)	(3)
	lexport	lexport	lexport
ln dens	2.325*** (24.18)	2.513*** (26.32)	1.051*** (10.68)
ln dens <sup>2</sup>	0.467*** (35.91)	0.497*** (38.53)	0.157*** (11.89)
ln A	0.613*** (27.31)	0.629*** (28.27)	0.462*** (19.55)
olsbyind	1.475*** (91.57)	1.312*** (81.34)	1.547*** (93.47)
olsdens	0.0268* (2.47)	0.0262* (2.43)	-0.0160 (-1.44)
ln fiscalexpend	-0.0171 (-1.07)	0.0796*** (5.04)	0.152*** (9.17)
ln humcapital	-0.353*** (-16.53)	-0.479*** (-22.62)	-0.550*** (-25.00)
institute	0.385*** (62.92)	0.364*** (59.69)	0.0347*** (5.15)
Stock dummy	Yes	Yes	Yes
Belong dummy	Yes	Yes	Yes
Status dummy	No	Yes	Yes
Size dummy	No	Yes	yes
Industry dummy	No	No	Yes
Region dummy	No	No	Yes
_cons	-22.66***	-20.61***	-28.44***

	(-118.03)	(-105.37)	(-104.43)
sigma_u			
_cons	10.93***	10.76***	8.948***
	(563.47)	(561.41)	(510.06)
sigma_e			
_cons	3.300***	3.267***	3.312***
	(822.64)	(821.88)	(827.81)
<i>N</i>	1930024	1930024	1930024

---



---

Note: *t* statistics in parentheses, \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 11: Agglomeration and firms' exporting choices considering their interactions

	First stage		Second stage					
	(1)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	lpbyind	exportdum	exportdum	exportdum	exportdum	exportdum	exportdum	exportdum
ln dens	0.0390*** (60.31)	0.941*** (42.11)	0.951*** (42.34)		0.339*** (14.38)	0.344*** (14.55)		0.511*** (21.85)
ln dens <sup>2</sup>		0.180*** (26.49)	0.183*** (26.80)		0.112*** (16.32)	0.114*** (16.49)		0.157*** (23.50)
elpbyind		-0.343*** (-25.11)	-0.176*** (-4.73)	-0.0519*** (-3.68)	-0.0539*** (-3.81)	0.0401 (1.09)	-0.0316* (-2.22)	-0.00731 (-0.20)
Interactionioin			0.126*** (4.80)			0.0714** (2.75)		0.0198 (0.76)
loc				0.171*** (54.37)	0.171*** (54.33)	0.171*** (54.34)	0.160*** (43.16)	0.157*** (42.64)
urb				0.475*** (68.73)	0.465*** (58.91)	0.465*** (58.84)	0.306*** (41.16)	0.280*** (33.88)
div				-1.330*** (-17.83)	-1.290*** (-17.28)	-1.294*** (-17.32)	-0.619*** (-8.01)	-0.573*** (-7.43)
com				-0.0168***	-0.0163***	-0.0162***	- 0.0164***	- 0.0157***

				(-27.24)	(-26.38)	(-26.24)	(-26.46)	(-25.32)
Infispend	0.0370***	-0.0641***	-0.0625***	-0.0621***	-0.0555***	-0.0546***	0.0328***	0.0333***
	(102.31)	(-14.58)	(-14.29)	(-11.08)	(-9.17)	(-9.00)	(5.73)	(5.34)
Incityarea	0.00830***	0.172***	0.169***	-0.230***	-0.240***	-0.242***	-0.299***	-0.293***
	(11.77)	(20.76)	(20.45)	(-28.33)	(-24.88)	(-24.98)	(-36.89)	(-30.10)
institute	0.0262***	0.185***	0.185***	0.119***	0.128***	0.128***	0.141***	0.157***
	(164.57)	(79.76)	(79.54)	(58.61)	(47.98)	(47.86)	(60.53)	(54.85)
Industry dummy	Yes	No	No	No	No	No	Yes	Yes
Stock dummy	Yes	No	No	No	No	No	Yes	Yes
Belong dummy	Yes	No	No	No	No	No	Yes	Yes
Status dummy	Yes	No	No	No	No	No	Yes	Yes
Size dummy	Yes	No	No	No	No	No	Yes	Yes
Constant	-0.403***	-4.197***	-4.195***	-14.09***	-9.820***	-9.815***	-9.084***	-8.521***
	(-51.09)	(-91.25)	(-91.21)	(-149.43)	(-114.60)	(-114.50)	(-94.25)	(-77.17)
Constant		2.929***	2.928***	2.765***	2.751***	2.751***	2.132***	2.124***
		(686.62)	(686.36)	(625.44)	(619.27)	(619.23)	(389.64)	(387.23)
R-squared	0.278							
N	1972732	1972732	1972732	1972732	1972732	1972732	1972732	1972732

Note: *t* statistics in parentheses, \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 12: Agglomeration and firms' exports considering their interactions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	lnexport	lnexport	lnexport	lnexport	lnexport	lnexport	lnexport
ln dens	0.0612*** (108.04)	0.0607*** (107.15)		0.0375*** (62.77)	0.0371*** (62.05)		0.0415*** (69.02)
ln dens <sup>2</sup>	0.00342*** (23.18)	0.00320*** (21.60)		0.000689*** (4.66)	0.000467** (3.14)		0.000941*** (6.31)
elpyind	- 0.00680*** (-17.95)	-0.0195*** (-22.42)	- 0.00240*** (-5.71)	- 0.00145*** (-3.46)	-0.0139*** (-15.98)	- 0.00330*** (-7.83)	-0.0162*** (-18.53)
lnfisexp	0.0141*** (88.68)	0.0141*** (88.30)	0.0150*** (94.11)	0.0140*** (87.98)	0.0140*** (87.73)	0.0170*** (104.74)	0.0163*** (100.27)
lncityarea	0.0580*** (184.88)	0.0583*** (185.57)	0.0132*** (61.07)	0.0434*** (125.19)	0.0438*** (126.01)	0.0124*** (57.07)	0.0452*** (129.97)
institute	- 0.00759*** (-109.75)	- 0.00760*** (-109.88)	-0.0119*** (-163.06)	-0.0103*** (-137.58)	-0.0103*** (-137.46)	-0.0111*** (-147.56)	- 0.00936*** (-120.79)
crosselpden		- 0.00961*** (-16.22)			- 0.00961*** (-16.32)		-0.0104*** (-17.57)
loc			0.00193*** (20.45)	0.00153*** (16.27)	0.00153*** (16.21)	0.00225*** (21.76)	0.00187*** (18.21)

urb			0.0469***	0.0329***	0.0328***	0.0508***	0.0362***
			(175.14)	(110.69)	(110.61)	(183.45)	(119.68)
div			-0.318***	-0.308***	-0.307***	-0.327***	-0.318***
			(-115.89)	(-114.55)	(-114.41)	(-118.85)	(-115.91)
com			0.0000612***	-	-	0.000161***	-0.0000236
			(4.51)	(-9.52)	(-10.64)	(11.55)	(-1.69)
Industry dummy	No	No	No	No	No	Yes	Yes
Stock dummy	No	No	No	No	No	Yes	Yes
Belong dummy	No	No	No	No	No	Yes	Yes
Status dummy	No	No	No	No	No	Yes	Yes
Size dummy	No	No	No	No	No	Yes	Yes
Constant	16.98***	16.98***	16.55***	16.64***	16.64***	16.46***	16.54***
	(9672.84)	(9665.15)	(4916.47)	(4784.05)	(4782.96)	(3536.57)	(3520.59)
R-squared	-0.346	-0.346	-0.336	-0.324	-0.324	-0.331	-0.318
N	1972325	1972325	1972325	1972325	1972325	1972325	1972325

Note: *t* statistics in parentheses, \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$





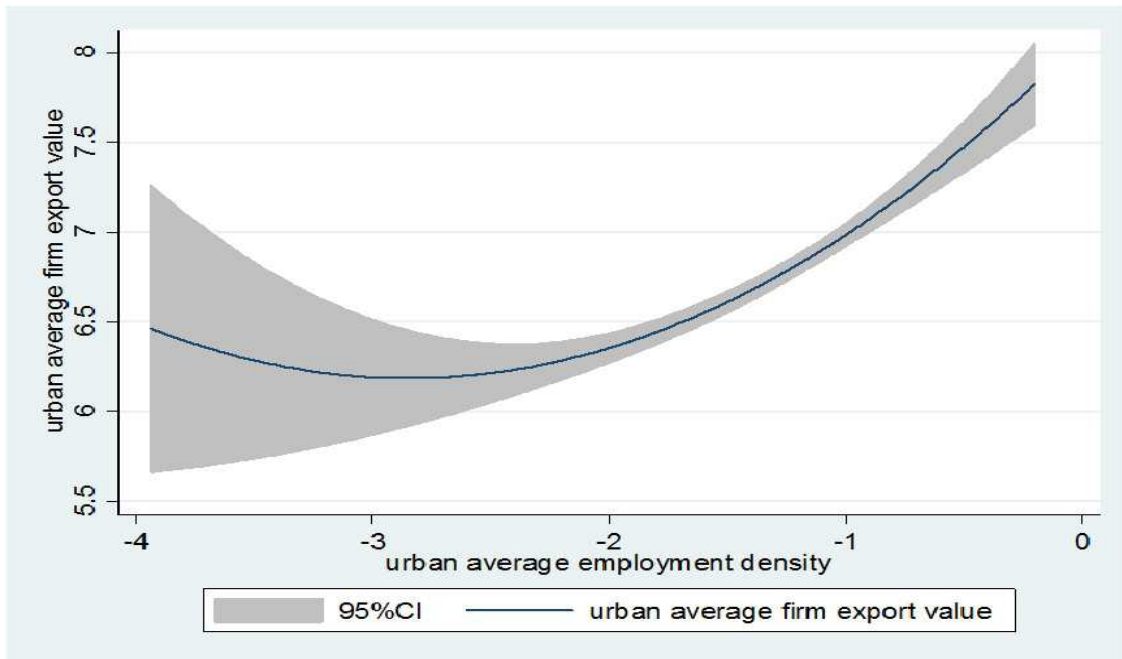


Figure 1: The correlation between urban average export value and urban labor density

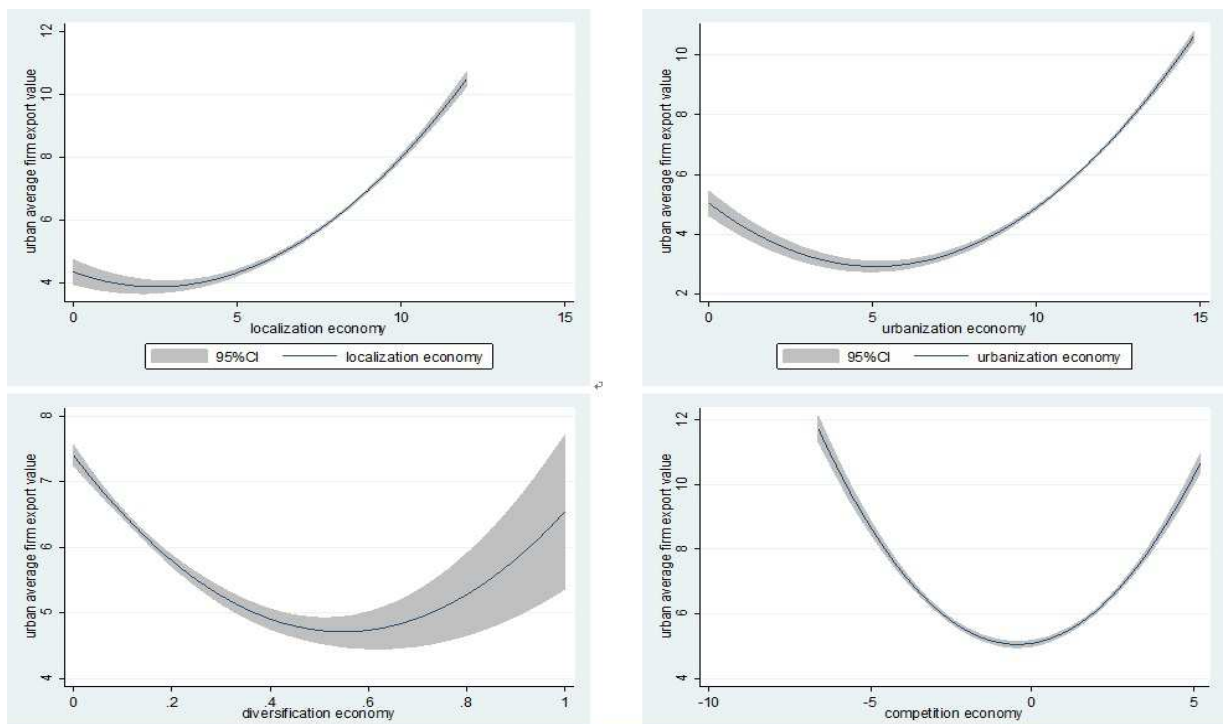


Figure 2: The correlation among urban average export value and agglomeration economy

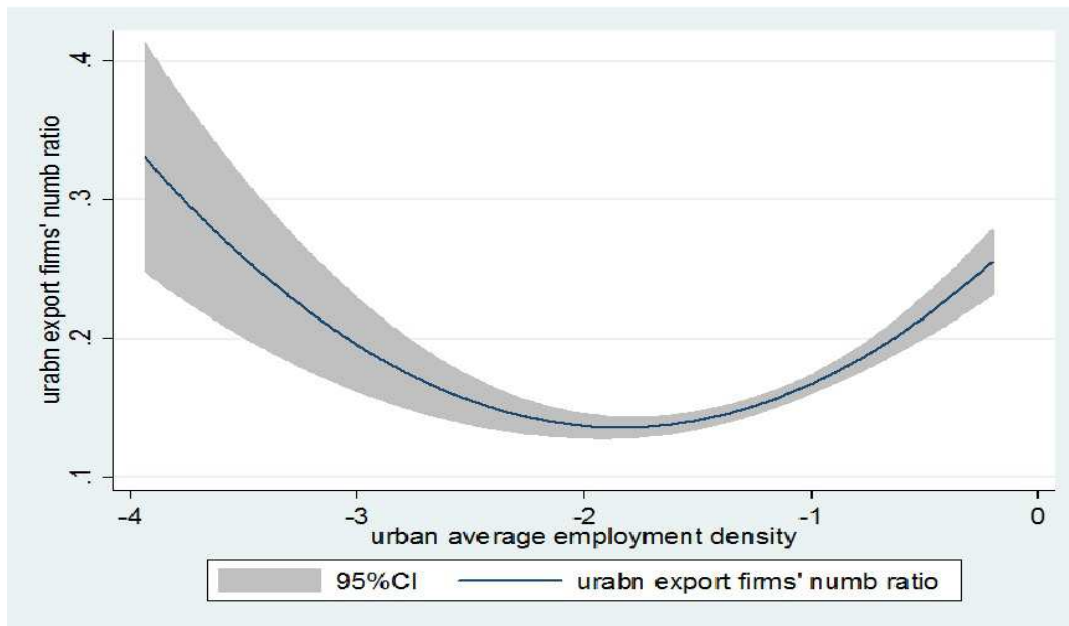


Figure 3: The correlation between urban export firm rate and labor density

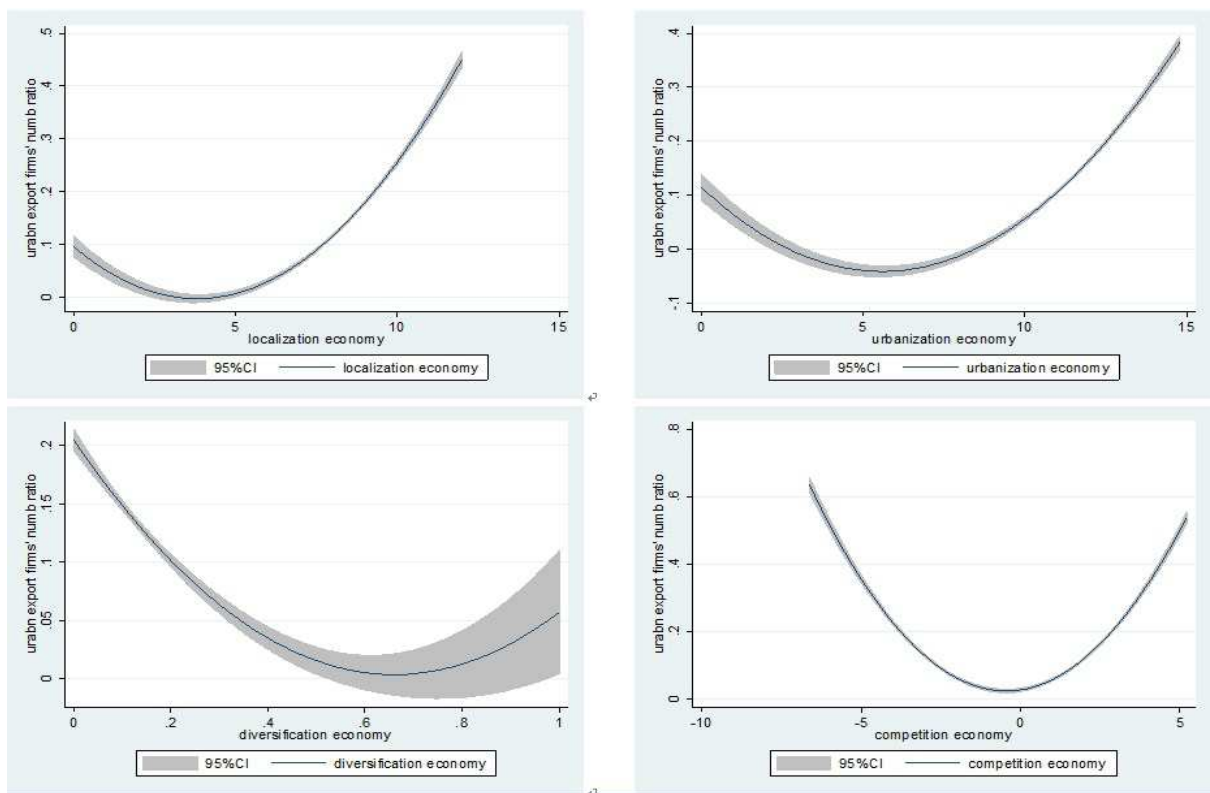


Figure 4: The correlation among urban average export value and agglomeration economy

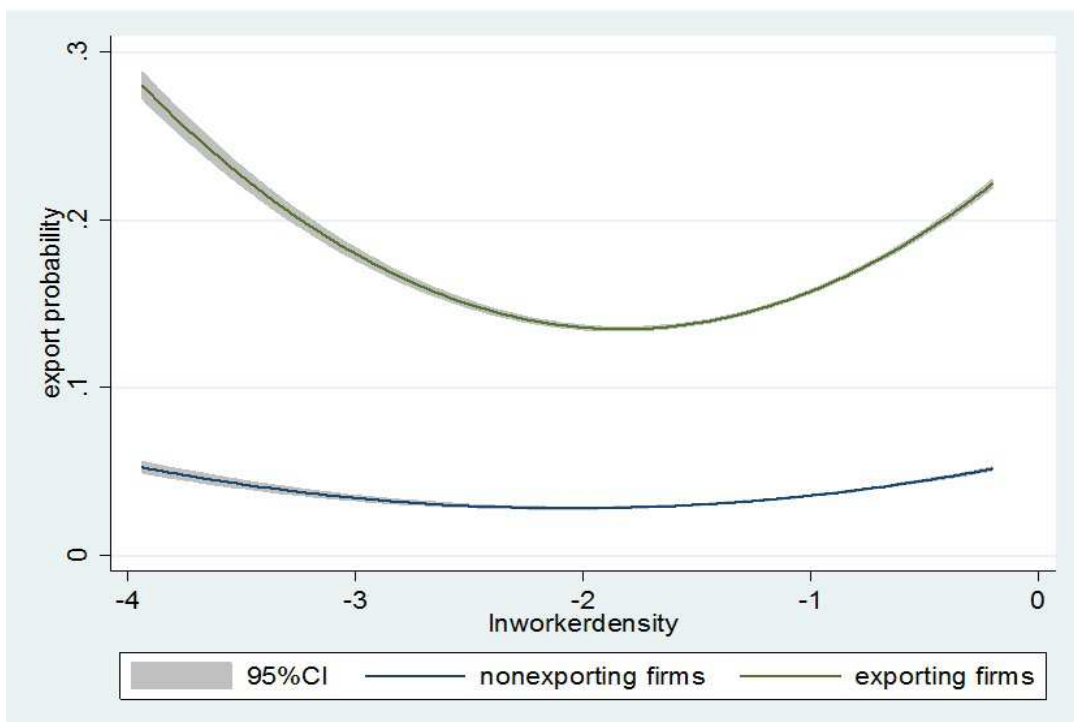


Figure 5: The correlation among firm exporting choices and direct export effect of agglomeration

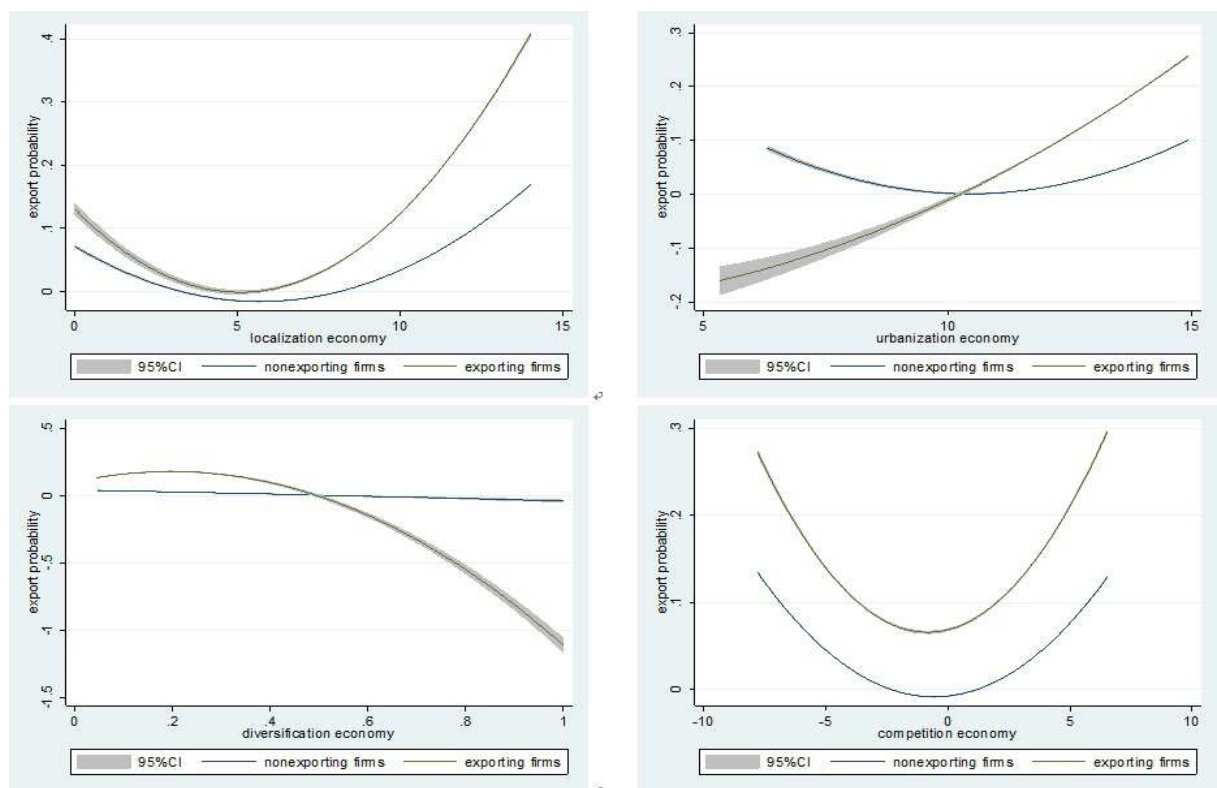


Figure 6: The correlation among firm exporting choices and agglomeration patterns