INSTITUTE OF DEVELOPING ECONOMIES

IDE Discussion Papers are preliminary materials circulated
to stimulate discussions and critical comments

IDE DISCUSSION PAPER No. 339

Assessing agglomeration economies in the Yangtze River Delta, China: A Bayesian spatial Durbin model approach

Yoshihiro HASHIGUCHI* and Kuang-hui CHEN

Revised October 22, 2012

Abstract

This paper estimates the elasticity of labor productivity with respect to employment density, a widely used measure of the agglomeration effect, in the Yangtze River Delta, China. A spatial Durbin model is presented that makes explicit the influences of spatial dependence and endogeneity bias in a very simple way. Results of Bayesian estimation using the data of the year 2009 indicate that the productivity is influenced by factors correlated with density rather than density itself and that spatial spillovers of these factors of agglomeration play a significant role. They are consistent with the findings of Ke (2010) and Artis, et al. (2011) that suggest the importance of taking into account spatial dependence and hitherto omitted variables.

Keywords: agglomeration economies, endogeneity, omitted variables, Bayesian, spatial Durbin model

JEL classification: C21, C51, R10, R15

---

*Corresponding author. Development Studies Center, IDE-JETRO.

The Institute of Developing Economies (IDE) is a semigovernmental, nonpartisan, nonprofit research institute, founded in 1958. The Institute merged with the Japan External Trade Organization (JETRO) on July 1, 1998. The Institute conducts basic and comprehensive studies on economic and related affairs in all developing countries and regions, including Asia, the Middle East, Africa, Latin America, Oceania, and Eastern Europe.

INSTITUTE OF DEVELOPING ECONOMIES (IDE), JETRO
3-2-2, WAKABA, MIHAMA-KU, CHIBA-SHI
CHIBA 261-8545, JAPAN

Assessing agglomeration economies in the Yangtze River Delta, China: A Bayesian spatial Durbin model approach

Yoshihiro Hashiguchi*
Institute of Developing Economies, Japan External Trade Organization

Kuang-hui Chen
Graduate School of International Cooperation Studies, Kobe University

Revised October 22, 2012

Abstract

This paper estimates the elasticity of labor productivity with respect to employment density, a widely used measure of the agglomeration effect, in the Yangtze River Delta, China. A spatial Durbin model is presented that makes explicit the influences of spatial dependence and endogeneity bias in a very simple way. Results of Bayesian estimation using the data of the year 2009 indicate that the productivity is influenced by factors correlated with density rather than density itself and that spatial spillovers of these factors of agglomeration play a significant role. They are consistent with the findings of Ke (2010) and Artis, et al. (2011) that suggest the importance of taking into account spatial dependence and hitherto omitted variables.

Keywords: agglomeration economies, endogeneity, omitted variables, Bayesian, spatial Durbin model

JEL classification: C21, C51, R10, R15

*Corresponding author.
Address: 3-2-2 Wakaba, Mihama-ku, Chiba-shi, Chiba 261-8545, Japan.
E-mail: Yoshihiro_Hashiguchi@ide.go.jp

1 Introduction

It is well known that the remarkable growth of China after adopting their open door policy has not been geographically uniform. Although causing the problem of regional inequality, it has generated large industrial agglomerations. At present, China has three major areas of industrial agglomeration: the Bohai Economic Rim centered on Beijing and Tianjin; the Yangtze River Delta area extending across Shanghai, Jiangsu, and Zhejiang; and the Pearl River Delta area located in Guangdong.

Considering the rapid growth of these agglomeration areas, it would be natural to expect that agglomeration economies have overwhelmed associated diseconomies and have produced a strong positive net effect. The expectation, however, is not well supported by the widely used measure of Ciccone and Hall (1996) and Ciccone (2002), the elasticity of labor productivity with respect to employment density.[1] While Fan (2007) estimated the elasticity to be significantly positive using the data of 261 prefecture-level regions in 2004, Ke (2010) found from the data of 617 cities in 2005 that it was insignificant and rather negative when spatial spillovers of productivity and the size of the industrial sector were controlled for. Why are their estimates so divergent? Does Ke's finding indicate that China failed to benefit from agglomeration economies?

Ke's finding seems to suggest that the elasticity estimate is sensitive to endogeneity bias due not only to the well-known problem of reverse causality, the problem that density could be an effect rather than a cause of productivity, but also to omitted variables.[2] Artis, et al. (2011) did report that their estimates of British elasticity dropped dramatically when spatial dependence and intangible assets were taken into account.

In this paper, we estimate the elasticity in the Yangtze River Delta with county-level data and a model that can make explicit the influences of spatial dependence and endogeneity bias in a very simple way. Specifically, we estimate the spatial Durbin model used by Chen and Hashiguchi (2010) with the Bayesian method and the results show a substantial influence of omitted variables on own and nearby regions.[3]

The rest of this paper is organized as follows: Section 2 describes the model, Section 3 explains the estimation method and data, Section 4 reports the results, and Section 5 concludes.

2 Model

We assume a production function of the Ciccone-Hall type:

$$\frac{Y_i}{A_i} = z_i \left[\left(\frac{L_i}{A_i} \right)^{\beta} \left(\frac{K_i}{A_i} \right)^{1-\beta} \right]^{\alpha} \left(\frac{Y_i}{A_i} \right)^{(\lambda-1)/\lambda}, \tag{1}$$

[1] Recent examples of its use are Brülhart and Mathys (2008) and Broersma and Oosterhaven (2009). Comprehensive reviews of agglomeration effects and their measurement are provided by Eberts and McMillen (1999), Rosenthal and Strange (2004), Graham (2008), Cohen and Paul (2009), and Puga (2010).

[2] Combes, et al. (2011) give a detailed discussion of bias caused by the endogeneity of employment density.

[3] Chen and Hashiguchi (2010) estimated the elasticity in Zhejiang, the southern part of the Yangtze Delta region.

where Y_i is output, A_i is land area, L_i is labor input, K_i is capital input of region i; z_i is a parameter representing total factor productivity, $\beta \in (0,1)$ and $\alpha \in (0,1)$ are distribution parameters, and λ is a parameter of density externality. α and λ measure the effects of congestion and of agglomeration, respectively.

Solving Equation (1) for Y_i/L_i yields

$$\frac{Y_i}{L_i} = z_i^{\lambda} \left(\frac{L_i}{A_i}\right)^{\gamma-1} \left(\frac{K_i}{L_i}\right)^{(1-\beta)\gamma}. \tag{2}$$

$\gamma = \alpha\lambda$ measures the net effect of agglomeration. $\gamma > 1$ if agglomeration economies are more than offset by congestion effects.

Due to the unavailability of capital data, we follow Ciccone and Hall and assume that the rental price of capital is constant at r in all regions, we then have the demand function of capital:

$$K_i = \frac{\alpha(1-\beta)}{r} Y_i.$$

Substitution into Equation (2) yields

$$\begin{aligned} \log\frac{Y_i}{L_i} &= \frac{\lambda}{1-(1-\beta)\gamma}\log z_i + \frac{(1-\beta)\gamma}{1-(1-\beta)\gamma}\log\frac{\alpha(1-\beta)}{r} + \frac{\gamma-1}{1-(1-\beta)\gamma}\log\frac{L_i}{A_i} \\ &= u_i + \phi + \theta\log\frac{L_i}{A_i}, \end{aligned} \tag{3}$$

where ϕ is a constant and

$$\theta = \frac{\gamma-1}{1-(1-\beta)\gamma}$$

is the elasticity of labor productivity with respect to employment density. Because $\partial\theta/\partial\gamma > 0$ and

$$\theta \gtreqless 0 \text{ when } \gamma \gtreqless 1, \tag{4}$$

θ can be used to assess the net agglomeration effect.[4]

Letting

$$u_i = \frac{\lambda}{1-(1-\beta)\gamma}\log z_i$$

associated with total factor productivity be the disturbance term enables the estimation of Equation (3). A standard way of estimation is to instrument $\log(L_i/A_i)$ because: (i) the density could be an effect rather than a cause of productivity and hence correlates with the TFP; and (ii) the model probably is underspecified and suffers from the omitted variable problem.[5]

Instead of instrumenting $\log(L_i/A_i)$, we assume: (i) the TFP and omitted variables depends on geography; and (ii) they are spatially autocorrelated as a result. We are

[4] θ is a hyperbolic function of γ with asymptotes at $\theta = -(1-\beta)^{-1}$ and $\gamma = (1-\beta)^{-1}$. Equation (4) holds only when $\theta > -(1-\beta)^{-1}$, and a paradoxical situation emerges where the employment elasticity $\theta < 0$ under the net agglomeration effect $\gamma > 1$ if $\theta < -(1-\beta)^{-1}$. We assume $\theta \geq -1$ to rule out this situation.

[5] In fact, the original models of Ciccone and Hall (1996) and Ciccone (2002) have a variable representing the quality of labor.

then able to take explicit account of the endogeneity problem with the following specification:

$$\log \frac{Y_i}{L_i} = \phi + \theta \log \frac{L_i}{A_i} + v_i,$$
$$v_i = \rho \sum_j w_{ij} v_j + \delta \log \frac{L_i}{A_i} + \varepsilon_i, \tag{5}$$

where v_i is the error term including the effects of omitted variables, $\sum_j w_{ij} v_j$ is the spatial lag of v_i, with w_{ij} being the (i, j)th element of a raw-standardized spatial weight matrix (Anselin 1988), ε_i is the "true" disturbance term; ρ is the autocorrelation parameter of v_i, and δ is the correlation parameter between v_i and $\log(L_i/A_i)$.[6]

In vector notation, Equations (5) are:

$$\mathbf{y} = \phi\, \mathbf{i} + \theta\, \mathbf{x} + \mathbf{v},$$
$$\mathbf{v} = \rho\, \mathbf{W} \mathbf{v} + \delta\, \mathbf{x} + \boldsymbol{\varepsilon}, \tag{6}$$

where $\mathbf{i}$ is an $n \times 1$ vector of ones, $\mathbf{I}$ is an $n \times n$ identity matrix, and

$$\mathbf{y} = \begin{bmatrix} \log \frac{Y_1}{L_1} & \log \frac{Y_2}{L_2} & \dots & \log \frac{Y_n}{L_n} \end{bmatrix}',$$
$$\mathbf{x} = \begin{bmatrix} \log \frac{L_1}{A_1} & \log \frac{L_2}{A_2} & \dots & \log \frac{L_n}{A_n} \end{bmatrix}',$$
$$\mathbf{v} = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}',$$
$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \dots & \varepsilon_n \end{bmatrix}',$$
$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nn} \end{bmatrix}.$$

Derived from Equations (6) is the spatial Durbin model (Anselin 1988):

$$\mathbf{y} = \rho\, \mathbf{W}\mathbf{y} + (1 - \rho)\, \phi\, \mathbf{i} + (\theta + \delta)\, \mathbf{x} - \rho\, \theta\, \mathbf{W}\mathbf{x} + \boldsymbol{\varepsilon}. \tag{7}$$

We estimate its parameters using the Bayesian method.

3 Bayesian estimation

3.1 Likelihood function and prior distribution

The Bayesian method uses the posterior distribution of unknown parameters for estimation. The posterior is proportional to the product of the likelihood function and the prior distribution.

[6]The assumption of the raw-standardized weight matrix implies that we specify the spatial lag $\sum_j w_{ij} v_j$ to be the average of nearby regions.

Assuming that $\boldsymbol{\varepsilon}$ in Equation (7) has a multivariate normal distribution $N(\mathbf{0}, \sigma^2\mathbf{I})$, we have the likelihood function:

$$\begin{aligned} f(\mathbf{y}\,|\,\Theta) &= (2\pi\sigma^2)^{-n/2}|\mathbf{I}-\rho\mathbf{W}| \\ &\quad \times \exp\left\{-\frac{1}{2\sigma^2}[(\mathbf{I}-\rho\mathbf{W})\,\mathbf{y} - h(\mathbf{x}, \mathbf{W}, \Theta_{-\sigma^2}]'[(\mathbf{I}-\rho\mathbf{W})\,\mathbf{y} - h(\mathbf{x}, \mathbf{W}, \Theta_{-\sigma^2})]\right\}, \end{aligned} \tag{8}$$

where Θ denotes the set of parameters ϕ, θ, δ, ρ, and σ^2; $\Theta_{-\sigma^2}$ denotes the set excluding σ^2, and

$$h(\mathbf{x}, \mathbf{W}, \Theta_{-\sigma^2}) = (1-\rho)\,\phi\,\mathbf{i} + (\theta+\delta)\,\mathbf{x} - \rho\,\theta\,\mathbf{W}\mathbf{x}.$$

The prior distribution is assumed to be

$$p(\Theta) = p(\phi)\, p(\theta)\, p(\delta)\, p(\rho)\, p(\sigma^2), \tag{9}$$

with ϕ and δ having normal distributions:

$$\begin{aligned} \phi &\sim N(\tilde{\phi}, \tilde{\sigma}^2_{\phi}), \\ \delta &\sim N(\tilde{\delta}, \tilde{\sigma}^2_{\delta}), \end{aligned}$$

θ having a truncated normal distribution:

$$\theta \sim TN_{[-1,\infty)}(\tilde{\theta}, \tilde{\sigma}^2_{\theta}),$$

ρ having a uniform distribution:

$$\rho \sim U(\tilde{a}, \tilde{b}),$$

and σ^2 following an inverse gamma distribution:

$$\sigma^2 \sim IG(\tilde{\nu}/2, \tilde{\omega}/2).$$

3.2 Full conditional posterior distributions

We used the Markov chain Monte Carlo method for posterior inference. MCMC samples were generated from the following full conditional posteriors derived from Equations (8) and (9):

$$\begin{aligned} \phi\,|\,\Theta_{-\phi}, \mathbf{y} &\sim N(\hat{\phi}, \hat{\sigma}^2_{\phi}), \\ \delta\,|\,\Theta_{-\delta}, \mathbf{y} &\sim N(\hat{\delta}, \hat{\sigma}^2_{\delta}), \\ \theta\,|\,\Theta_{-\theta}, \mathbf{y} &\sim TN_{[-1,\infty)}(\hat{\theta}, \hat{\sigma}^2_{\theta}), \\ \sigma^2\,|\,\Theta_{-\sigma^2}, \mathbf{y} &\sim IG(\hat{\nu}/2, \hat{\omega}/2), \end{aligned}$$

where

$$
\begin{aligned}
&\hat{\phi} = \hat{\sigma}^2_\phi \left\{\sigma^{-2}(1-\rho)\,\mathbf{i}'\,[(\mathbf{I}-\rho\mathbf{W})\,\mathbf{y} - (\theta+\delta)\,\mathbf{x} + \rho\,\theta\,\mathbf{W}\mathbf{x}] + \tilde{\phi}/\tilde{\sigma}^2_\phi\right\},\\
&\hat{\sigma}^2_\phi = \left[\sigma^{-2}(1-\rho)^2\,\mathbf{i}'\mathbf{i} + \tilde{\sigma}^{-2}_\phi\right]^{-1},\\
&\hat{\delta} = \hat{\sigma}^2_\delta \left\{\sigma^{-2}\mathbf{x}'\,[(\mathbf{I}-\rho\,\mathbf{W})\,\mathbf{y} - (1-\rho)\,\phi\,\mathbf{i} - \theta\,\mathbf{x} + \rho\,\theta\,\mathbf{W}\mathbf{x}] + \tilde{\delta}/\tilde{\sigma}^2_\delta\right\},\\
&\hat{\sigma}^2_\delta = \left(\sigma^{-2}\mathbf{x}'\mathbf{x} + \tilde{\sigma}^{-2}_\delta\right)^{-1},\\
&\hat{\theta} = \hat{\sigma}^2_\theta \left\{\sigma^{-2}\mathbf{x}'(\mathbf{I}-\rho\,\mathbf{W})'\,[(\mathbf{I}-\rho\,\mathbf{W})\,\mathbf{y} - (1-\rho)\,\phi\,\mathbf{i} - \rho\,\mathbf{x}] + \tilde{\theta}/\tilde{\sigma}^2_\theta\right\},\\
&\hat{\sigma}^2_\theta = \left[\sigma^{-2}\mathbf{x}'(\mathbf{I}-\rho\,\mathbf{W})'(\mathbf{I}-\rho\,\mathbf{W})\,\mathbf{x}) + \tilde{\sigma}^{-2}_\theta\right]^{-1},\\
&\hat{\nu} = \tilde{\nu} + n,\\
&\hat{\omega} = \tilde{\omega} + [(\mathbf{I}-\rho\,\mathbf{W})\,\mathbf{y} - h(\mathbf{x},\mathbf{W},\Theta_{-\sigma^2})]'\,[(\mathbf{I}-\rho\,\mathbf{W})\,\mathbf{y} - h(\mathbf{x},\mathbf{W},\Theta_{-\sigma^2})],
\end{aligned}
$$

and

$$
p(\rho\,|\,\Theta_{-\rho},\mathbf{y}) \propto |\mathbf{I}-\rho\,\mathbf{W}|\,\exp\left[-\frac{1}{2\hat{\sigma}^2_\rho}(\rho-\hat{\rho})^2\right] I_{(\tilde{a},\tilde{b})}(\rho),
$$

where

$$
\begin{aligned}
&\hat{\rho} = \hat{\sigma}^2_\rho\,\sigma^{-2}(\mathbf{W}\mathbf{y} - \phi\,\mathbf{i} - \theta\,\mathbf{W}\mathbf{x})'\,[\mathbf{y} - \phi\,\mathbf{i} - (\theta+\delta)\,\mathbf{x}],\\
&\hat{\sigma}^2_\rho = \left[\sigma^{-2}(\mathbf{W}\mathbf{y} - \phi\,\mathbf{i} - \theta\,\mathbf{W}\mathbf{x})'(\mathbf{W}\mathbf{y} - \phi\,\mathbf{i} - \theta\,\mathbf{W}\mathbf{x})\right]^{-1},\\
&I_{(\tilde{a},\tilde{b})}(\rho) = \begin{cases} 1 & \text{if } \tilde{a} < \rho < \tilde{b} \\ 0 & \text{elsewhere} \end{cases}.
\end{aligned}
$$

The sampling algorithm is described in Appendix 1.

3.3 Data, spatial weights, and hyperparameters

We used county-level data from the municipality of Shanghai and the provinces of Jiangsu and Zhejiang for the year 2009.[7] They were gross regional products (Y_i), numbers of employed persons (L_i), and land areas (A_i) obtained from the statistical yearbooks of Shanghai, Jiangsu, and Zhejiang.[8] The sample size is $n = 134$.

We used the (raw-standardized) spatial weight matrix $\mathbf{W}$ of the queen contiguity type. Appendix 2 gives the details of our neighborhood identification.

The hyperparameters of prior distributions were given as follows:

$$
\begin{aligned}
&\tilde{\phi} = \tilde{\delta} = \tilde{\theta} = 0,\\
&\tilde{\sigma}^2_\phi = \tilde{\sigma}^2_\delta = \tilde{\sigma}^2_\theta = 100,\\
&\tilde{a} = \lambda^{-1}_{\min},\ \tilde{b} = \lambda^{-1}_{\max},\\
&\tilde{\nu} = 3,\ \tilde{\omega} = 0.01,
\end{aligned}
$$

[7]County-level regions in this area are: (i) city districts and a county (Chongming) in Shanghai, and (ii) city districts of prefecture-level cities, counties, and county-level cities in Jiangsu and Zhejiang. Due to the unavailability of data, we aggregated: (i) all the regions in Shanghai, and (ii) city districts of prefecture-level cities into respective cities.

[8]We averaged the end-of-year numbers of 2008 and 2009 for L_i.

where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest eigenvalues of $\mathbf{W}$, respectively.[9]

4 Estimation results

The estimation was performed separately for secondary industry, tertiary industry, and non-primary industries (both secondary and tertiary industry).[10] Table 1 summarizes the results.

Table 1: Estimation results

	Mean	SD	95% CI
Secondary industry			
ϕ	2.989	0.445	[2.109, 3.870]
θ	-0.193	0.148	[-0.527, 0.062]
δ	0.192	0.114	[0.010, 0.462]
ρ	0.582	0.086	[0.408, 0.740]
σ^2	0.133	0.017	[0.103, 0.170]
Tertiary industry			
ϕ	3.386	0.649	[2.126, 4.691]
θ	-0.169	0.112	[-0.408, 0.034]
δ	0.176	0.090	[0.020, 0.376]
ρ	0.654	0.070	[0.512, 0.783]
σ^2	0.133	0.017	[0.104, 0.171]
Non-primary industries			
ϕ	2.993	0.421	[2.159, 3.830]
θ	-0.035	0.101	[-0.248, 0.151]
δ	0.119	0.074	[-0.009, 0.282]
ρ	0.655	0.072	[0.508, 0.791]
σ^2	0.094	0.012	[0.073, 0.121]

Note: Mean, SD, and 95% CI denote the posterior mean and standard deviation, and 95% credible interval, respectively.

The posterior means of ρ, the parameter of spatial dependence, are 0.582–0.655 and all the credible intervals do not include zero, supporting our use of the spatial model. The means of δ are 0.119–0.192, with creditable intervals for secondary and tertiary industries not including zero and that of non-primary industries only slightly overlapping zero, indicating a large probability that the employment density $\log(L_i/A_i)$ and the error term containing omitted variables v_i are correlated.

The means of θ, the elasticity of productivity with respect to employment density, are all negative, ranging between -0.196 and -0.035. All the credible intervals overlap

[9] $\lambda_{\min}^{-1}$ and $\lambda_{\max}^{-1}$ of our $\mathbf{W}$ are -1.189 and 1, respectively.

[10] Computation was implemented with *Ox* version 6.20 (Doornik 2009).

zero, but, as Figure 1 shows, the probability that $\theta < 0$ is greater than 90% in secondary and tertiary industries. It would be safe to estimate that the elasticity is almost zero in non-primary industries and is negative in secondary and tertiary industries.

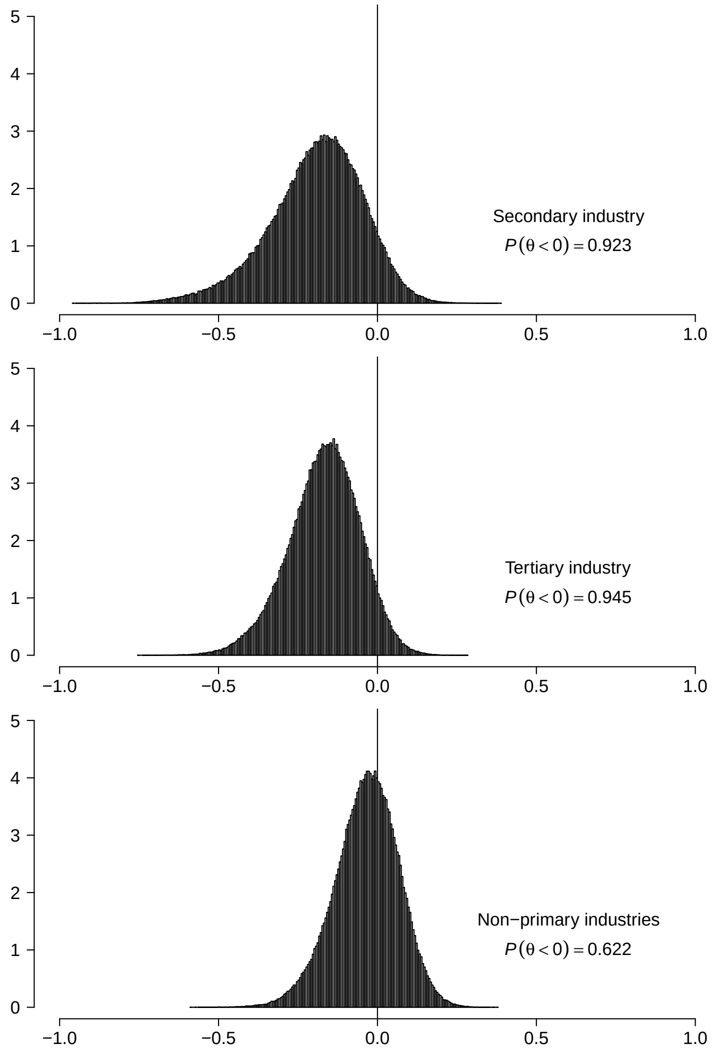

Figure 1: Posterior distribution of θ

5 Conclusion

We estimated the agglomeration effect, the elasticity of labor productivity with respect to employment density, in the Yangtze River Delta, using the spatial Durbin model, which makes explicit the influences of spatial dependence and endogeneity bias in a very simple way. The elasticity was estimated to be almost zero in non-primary industries and negative in secondary and tertiary industries.

Has China failed to benefit from agglomeration economies? Our results do not necessarily imply failure, but they do not support the idea that density improves productivity on its own. We have from Equation (7)

$$\mathbf{y} = \phi\,\mathbf{i} + \theta\,\mathbf{x} + \delta\,\mathbf{x} + \rho\,\mathbf{W}(\mathbf{y} - \phi\,\mathbf{i} - \theta\,\mathbf{x}) + \boldsymbol{\varepsilon}.$$

Our parameter estimates indicate: (i) productivity was influenced by factors correlated with density, $\delta\,\mathbf{x}$, rather than density itself; and (ii) spatial spillovers of these factors of agglomeration, $\rho\,\mathbf{W}(\mathbf{y} - \phi\,\mathbf{i} - \theta\,\mathbf{x})$, played a significant role.

Our results are consistent with the findings of Ke (2010) and Artis, et al. (2011) that suggest the importance of taking into account spatial dependence and hitherto omitted variables. What then are indispensable variables? The list is incomplete. There seems to be no consensus other than labor quality. Further research is required.

Appendix 1 MCMC Sampling

The MCMC samples are generated as follows:

1. Choose arbitrary initial values of parameters $\Theta_{(0)} = \{\phi_{(0)}, \theta_{(0)}, \delta_{(0)}, \rho_{(0)}, \sigma^2_{(0)}\}$.
2. Draw $\Theta_{(t)}, t = 1, 2, \ldots, M$ in the following order:
   (i) Draw $\phi_{(t)}$ from $p(\phi\,|\,\theta_{(t-1)}, \delta_{(t-1)}, \rho_{(t-1)}, \sigma^2_{(t-1)}, \mathbf{y})$.
   (ii) Draw $\delta_{(t)}$ from $p(\delta\,|\,\phi_{(t)}, \theta_{(t-1)}, \rho_{(t-1)}, \sigma^2_{(t-1)}, \mathbf{y})$.
   (iii) Draw $\theta_{(t)}$ from $p(\theta\,|\,\phi_{(t)}, \delta_{(t)}, \rho_{(t-1)}, \sigma^2_{(t-1)}, \mathbf{y})$.
   (iv) Draw $\sigma^2_{(t)}$ from $p(\sigma^2\,|\,\phi_{(t)}, \theta_{(t)}, \delta_{(t)}, \rho_{(t-1)}, \mathbf{y})$.
   (v) Draw $\rho_{(t)}$ from $p(\rho\,|\,\phi_{(t)}, \theta_{(t)}, \delta_{(t)}, \sigma^2_{(t)}, \mathbf{y})$.
3. Discard the first M_0 draws and save the remaining $M - M_0$.

Since $\rho\,|\,\Theta_{-\rho}, \mathbf{y}$ follows a non-standard distribution:

$$p(\rho\,|\,\Theta_{-\rho}, \mathbf{y}) \propto |\mathbf{I} - \rho\,\mathbf{W}|\;\exp\left[-\frac{1}{2\hat{\sigma}^2_\rho}\,(\rho - \hat{\rho})^2\right]\,I_{(\tilde{a},\tilde{b})}(\rho),$$

the Metropolis-Hastings algorithm is used to draw $\rho_{(t)}$:

1. Generate a proposal ρ^* from a truncated normal distribution $TN_{(\tilde{a},\tilde{b})}(\hat{\rho}, \hat{\sigma}^2_\rho)$.
2. Calculate the acceptance probability:

$$\alpha = \min\left[1,\, \frac{|\mathbf{I} - \rho^*\mathbf{W}|}{|\mathbf{I} - \rho_{(t-1)}\mathbf{W}|}\right].$$

3. Generate $u \sim U(0, 1)$ and let

$$\rho_{(t)} = \begin{cases} \rho^* & \text{if } u \leq \alpha \\ \rho_{(t-1)} & \text{else} \end{cases}.$$

We let $M = 500,000$ and $M_0 = 50,000$, and used the samples of 450,000 draws for posterior inference.

Appendix 2 Neighborhood identification

Using the queen contiguity criteria, we defined regions sharing a common border, including a river border, or vertex as neighbors. We assumed in addition:

- (i) Shengsi adjoined Daishan, (ii) Daishan adjoined the city districts of Zhoushan, and (iii) Dongtou adjoined Yuhuan, to avoid leaving out island regions that had no neighbor; and
- (i) Shanghai adjoined Shengsi, and (ii) Cixi adjoined Haiyan, taking account of connections through Donghai Bridge and Hangzhou Bay Bridge, respectively.

Figure 2 shows the neighbor relations.

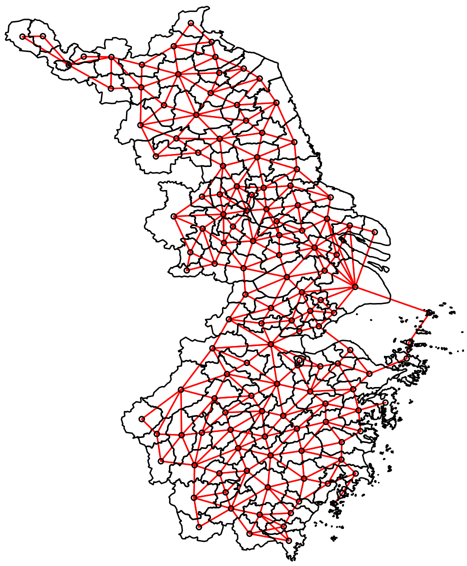

Figure 2: Neighbor relations

References

Anselin L (1988) *Spatial econometrics: Methods and models*. Kluwer Academic Publishers, Dordrecht.

Artis MJ, Miguélez E, Moreno R (2011) Agglomeration economies and regional intangible asset: An empirical investigation. *Journal of Economic Geography* 11: 1-23.

Broersma L, Oosterhaven J (2009) Regional labor productivity in the Netherlands: Evidence of agglomeration and congestion effects. *Journal of Regional Science* 49: 483-511.

Brülhart M, Mathys NA (2008) Sectoral agglomeration economies in a panel of European regions. *Regional Science and Urban Economics* 38: 348-362.

Chen K, Hashiguchi Y (2010) Agglomerations and agglomeration economies in Zhejiang, China. *Kokumin-Keizai Zasshi* 201.4: 53-64. (In Japanese.)

Ciccone A (2002) Agglomeration effects in Europe. *European Economic Review* 46: 213-227.

Ciccone A, Hall RE (1996) Productivity and the density of economic activity. *American Economic Review* 86: 54-70.

Cohen JP, Paul CJM (2009) Agglomeration, productivity and regional growth: Production theory approaches. In: Capello R, Nijkamp P (eds) *Handbook of regional growth and development theories*. Edward Elgar, Cheltenham, UK. pp. 101-117.

Combes P, Duranton G, Gobilln L (2011) The identification of agglomeration economies. *Journal of Economic Geography* 11: 253-266.

Doornik JA (2009) *An object-oriented matrix programming language Ox 6*. Timberlake Consultants, London.

Eberts RW, McMillen DP (1999) Agglomeration economies and urban public infrastructure. In: Cheshire P, Mills ES (eds) *Handbook of regional and urban economics* vol. 3. North-Holland, New York. pp. 1455-1495.

Fan J (2007) Industrial agglomeration and difference of regional productivity. *Frontiers of Economics in China* 2: 346-361.

Graham DJ (2008) Identify urbanisation and localisation externalities in manufacturing and service industries. *Papers in Regional Science* 88: 63-84.

Ke S (2010) Agglomeration, productivity, and spatial spillovers across Chinese cities. *Annals of Regional Science* 45: 157-179.

Puga D (2010) The magnitude and causes of agglomeration economies. *Journal of Regional Science* 50: 203-219.

Rosenthal SS, Strange WC (2004) Evidence on the nature and sources of agglomeration economies. In: Henderson JV, Thisse JF (eds) *Handbook of regional and urban economics* vol. 4. Elsevier, Amsterdam. pp. 2119-2171.