# AN APPLICATION OF SPATIAL - PANEL ANALYSIS - PROVINCIAL ECONOMIC GROWTH AND LOGISTICS IN CHINA

**YANG SHAO** *

**ABSTRACT:** *This paper introduces the spatial panel autocorrelation model, utilizes C-D production functions, constructs the spatial econometric model and finally studies the spatial correlativity between provincial economic growth and logistics. By using the spatial package of Matlab software, it verifies the possibility if there is the remarkable autocorrelation of the Chinese provincial economic growth and local logistics. On the base of building the spatial panel model, we research the spatial quantitative autocorrelation of the Chinese provincial economic growth and local logistics.*

**KEY WORDS:** *economic growth; logistics; spatial panel autocorrelation*

**JEL CLASSIFICATION:** *C10, O40*

Modern economic growth depends strongly on logistics. Logistics has become one of the most important factors to promote economic growth, adjust industrial layout and drive the evolution of economic spatial structure. Previous studies of the relationship between economic growth and logistics, limited in time series, which ignored the differences between locations. This paper introduces the spatial factor into a unified analytical framework, considers not only the spatial heterogeneity but also spatial correlation between economic growth and logistics. This paper uses individual fixed-effect model as the basic panel-data model, and uses latest spatial panel-data model to study the correlation between provincial economic growth and local logistics in China.

---

* *Ph.D., School of Economics and Management, Changchun University of Science and Technology, China and Management College, Donghua University, China,* shaoyangliangliang@yahoo.com.cn

*The research is financed by National Natural Science Foundation of China. No. 70971020 (Sponsoring information)*

# 1. SPATIAL-PANEL MODEL AND CORRELATION TEST

## 1.1 Spatial-panel Models

Spatial effects of the spatial econometrics include spatial autocorrelation and spatial differences. The former is the correlation of the observations between a regional sample and other regional samples. The latter is the spatial-effect non-uniform at the regional level caused by the heterogeneity of spatial units (Anselin, 1988a). Spatial autocorrelation in the spatial autoregressive model is reflected in the error term and the lagged item of dependent variable. Therefore there are two basic spatial econometric models, one is Spatial Auto Regressive Model (SAR), the other is Spatial Error Model (SEM), and the basic formulas of two models are:

Spatial Auto Regressive Model (SAR):

$$y = \rho W_N y + X'\beta + \varepsilon \tag{1}$$

Spatial Error Model (SEM):

$$y = X'\beta + \mu$$
$$\mu = \lambda W_N \mu + \varepsilon \tag{2}$$

y is the dependent variable, X is the vector of independent variables (including constant term), β is variable factors, $\rho$ is spatial regression coefficients, $\lambda$ is spatial autocorrelation coefficients, $\varepsilon$ is the error components obeying the normal distribution, $W_N$ is the spatial matrix of $n \times n$ (n is the number of region), the weight coefficient can defined on actual conditions.

The above-mentioned model is a model for the cross-sectional data. In order to apply it to panel data, we need to change the model to meet the basic formula of panel data model. This paper uses individual fixed-effect model (Elhorst, 2003). The model controls two kinds of non-observable effects: spatial fixed-effect and time fixed-effect, the former is the effect of background variables which changed with the location, but no changed with time (such as economic structure and natural endowments, etc.) on steady-state level; the latter is the effect of background variables which changed with time, but no changed with location (such as the business cycle and temporary shock, etc.) on steady-state level.

To assume sF is N-dimensional column vector of spatial fixed-effect; tF is T-dimensional column vectors of time fixed-effect, the form is showing as follows:

$$sF = (\alpha_1, \alpha_2, \cdots, \alpha_N)^T,\ tF = (\delta_1, \delta_2, \cdots, \delta_N)^T$$

The column vectors of spatial and time fixed-effect of each observation are showing as follows: $\eta = i_T \otimes s_F, \delta = tF \otimes i_N$, where $i_T$ is T-dimensional column vector and $i_N$ is N-dimensional column vector, all elements of these two column vectors are 1. Then the equation (1) and (2) can be transformed into the following model (3) and (4):

$$y = \rho(I_T \otimes W_N)y + \eta + \delta + X'\beta + \nu \tag{3}$$

$$y = X'\beta + \eta + \delta + \mu$$

$$\mu = \lambda(I_T \otimes W_N)\mu + \nu \quad (4)$$

In the one-dimensional error decomposition model, $\varepsilon = \eta_i + \upsilon_{it}$ or $\varepsilon = \delta_i + \upsilon_{it}$; In the two-dimensional error decomposition model, $\varepsilon = \eta_i + \delta_i + \upsilon_{it}$ , $\eta_i \sim IID(0, \omega_i^2)$、 $\delta_i \sim IID(0, \xi_i^2)$ and $\upsilon_i \sim IID(0, \sigma_i^2)$. t is the time dimension, i is cross-section dimension, IT is an unit matrix of T-dimensional time matrix.

### 1.2. Spatial Correlation Test

Spatial correlation test bases mainly on the hypothesis testing of the maximum likelihood estimate, Wald, LR and LM statistics and spatial-related indices" Moran'S I", the null hypothesis $H_0$:$\rho$=0 or$\lambda$=0. However, Moran'S I (Moran,1948), LMerr (Burridge,1980), LMsar, Lratios, Walds (Anselin,1988b) and other spatial-related tests are applied for a single cross-section regression model, and can not be directly used by panel-data model. In this paper, block-diagonal matrix"$C = I_T \otimes W_N$"replace spatial-weight matrix of Moran'S I statistics, etc. So we can easily extend these tests to panel-data analysis. In the selection of model, we use firstly the LSDV (Least Square Dummy Variables) method estimation, do not consider the bound model of spatial correlation, and then carry out the spatial-related test. If LMsar (or LMerr) estimation is more significant than LMerr (or Lmsar) estimation, then the spatial lag model (or spatial error model) is more appropriate than the spatial error model (or spatial lag model). Anselin and Rey (1991) use Monte Carlo experiments method to show that this method can provide a good guidance on the selection of spatial econometric models.

### 1.3. Parameter Estimation

Usually we use the maximum likelihood method (ML) to estimate spatial econometric models (Anselin, 1988a; Anselin and Hud1992). ML estimation program can not be used directly by the spatial panel-data model, because it is applied to the cross-section regression model. In addition, when the dimension of spatial-weight matrix is large, there is a problem (Kelejian and Prucha, 1999) in the usual ML estimation procedures in spatial econometrics. At present, a solution can be used, it is the Monte Carlo method to approximate the log-likelihood function, the Jacobian determinant of natural logarithm (Barry & Pace, 1999). This method can be implemented in the spatial package of Matlab, and can be used to estimate model (3), (4).

## 2. MEASUREMENT MODEL

The local production function can be denoted by the homogeneous equation of Cobb-Douglas:

$$Y_t = f(K_t, L_t) = AK_t^{\alpha} L_t^{\beta} \tag{5}$$

Logistics become more and more important factor in the process of production, logistics has been seen as "the third profit source", same as the factors of capital and labor, to promote economic growth. Therefore, the C-D production function is improved. As the logistics(W) is independent of capital and labour, on Solow production function, the production function which include the elements of logistics can be described as:

$$Y_t = f(K_t, L_t, W_t) = AK_t^{\alpha} L_t^{\beta} W_t^{\gamma} \tag{6}$$

Y is the output, A is combination of technological advance, K is capital investment, L is labour input, W is logistics, $\alpha$、$\beta$、$\gamma$ are elasticity coefficient of capital, labour, logistics on economic growth, respectively. Considered the comparison of data and economic significance, all variables should be logarithmic, as follows:

$$\ln Y_{it} = \ln A_i + \alpha \ln K_{it} + \beta \ln L_{it} + \gamma \ln W_{it} + \mu_{it} \tag{7}$$

Subscript i is the province name, t is time series, is $\mu_{it}$ the random disturbance.

## 3. EMPIRICAL ANALYSIS

### 3.1. The Selection of Factors and Data Collection

Panel data from 1978 to 2007of 30 provinces in China are used to be empirical analysis. Data is mainly from the "New China, Compiling Statistical Information on Fifty-five Years" and "China Statistical Yearbook" (2006-2008). In order to compare data and reduce heteroscedasticity, all data should be changed into logarithm.

The detail data of variables are as follows:

1. GDP: for the elimination of price change factors, we think 1952 year as the base period, and generate the real GDP according to GDP index (Unit: hundred million).
2. Logistics level: we use cargo turnover of various province to measure the level of logistics and logistics capacity (unit: 100 million ton-km).
3. Labour: we use employment numbers of the whole society (unit: ten thousand).
4. Capital stock: we use a perpetual inventory method (Goldsmith 1951), which is now used widely by OECD countries, and its basic formula is:

$$K_{it} = K_{i,t-1}(1 - \delta_{it}) + I_{it} \tag{8}$$

$K_{it}$ represents the capital stock of i-region's at t-year, $K_{i,\ t-1}$ represents the capital stock of i-region's at (t-1)-year, $I_{it}$ represents the investment of i-region's at t-year; $\delta_{it}$ is the t-year's economic depreciation rate. We use Zhang Jun's capital stock data which mentioned in the paper "China's Provincial physical Capital Stock Estimate: 1952-2000", and capital stock in the other period is calculated by the data of "China Statistical Yearbook". Depreciation rate $\delta_{it} = 5\%$. (Unit: hundred million).

### 3.2. The Determination of Economic Spatial-weight ($W_{ij}$)

Spatial-weight matrix ($W_{ij}$) embodies the regional spatial-effect. Obeying the rule of "Rook", the adjacent rule, the matrix $W_{ij}$ is:

$$w_{ij} = \begin{cases} 1 & \text{When the region } i \text{ and the region } j \text{ are adjace} \\ 0 & \text{When the region } i \text{ and the region } j \text{ are not adjacent} \end{cases}$$

the main diagonal elements are 0. $w_{ij}$ ( i=1,2,…,n,j=1,2,…,n ) should be standardized.

There are borders between neighboring regions, but the economic ties are not identical between neighboring regions. Relative to the backward regions, the driving impact of backward regions on developed regions is weak, while the developed regions can generate great driving impact on the backward around regions, which is intensive spatial influence. Therefore, we get economic weight-matrix based on the binary weight matrix (Lin Guang- Ping, 2005), the formula is:

$$W^* = W * E, E_{ij} = \frac{1}{|\bar{y}_i - \bar{y}_i|}, \text{ and, } \bar{y}_i = \frac{1}{t_1 - t_0 + 1}\sum_{t=t_0}^{t_1} y_{it} \tag{9}$$

W is the weight-matrix of spatial location, E is the matrix of economic strength. We calculate the mean of proportion which is the real GDP of every region accounted for real GDP of all regions, with the result of them, measure the regional economic level. And assuming that the economic strength of this region is strong, the spatial impact of it on surrounds is strong, contrary to the weak (Xiao-ping Chen, Guo-ping Li, 2006). Economic spatial-weight matrix ( $W_{ii}$ ) is the diagonal matrix which is the product of a geo-spatial-weight( $w_{ij}$ )and the mean of proportion of regional GDP, the formula is:

$$W_{ij} = w_{ij} * \text{diag}(\frac{\bar{y}_1}{\bar{y}}, \frac{\bar{y}_2}{\bar{y}}, \cdots, \frac{\bar{y}_n}{\bar{y}}) \tag{10}$$

and $\bar{y}_i = \frac{1}{t_1 - t_0 + 1}\sum_{t=t_0}^{t_1} y_{it}, \bar{y} = \frac{1}{n(t_1 - t_0 + 1)}\sum_{t=1}^{n}\sum_{t=t_0}^{t_1} y_{it}$ 。

### 3.3. Empirical Analysis

With these assumptions of the model and estimation methods, using sub-provincial panel data, we establish the individual fixed-panel regression model of 30 regions from 1978 to 2007, analyzed by Eviews 6.0 software, use LSDV method to estimate the individual fixed-effects model, and get the elasticity coefficients and associated test results of the regression model, and estimates of the individual fixed-effect coefficients $\eta_i$. The results are showing in the table 1.

In table1 the values of $R^2$ and Adjust $R^2$ is high in the regression models, it indicate that the result of the simulation fitting of the model data is very good. In table 2 judging from the fixed-effects estimate of various region, we can find the size of the value of fixed-effects is close in adjacent regions of Beijing and Tianjin, Jiangsu and Zhejiang, northeast, southwest and northwest provinces, it shows that there is significant regional relevance, it is necessary to do spatial-related test firstly, with the result of it, we can know if it is necessary to do spatial-panel analysis further.

**Table 1. The empirical results of individual fixed-effect model of spatial-panel data of various province from 1978 to 2007**

| | $lnK_{it}$ | $lnL_{it}$ | $lnW_{it}$ | c |
|---|---|---|---|---|
| Coefficient | 0.5983 | 0.1388 | 0.0670 | -0.7626 |
| t-Statistic | 72.0240 | 4.9840 | 4.6823 | -4.1491 |
| Prob. | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $R^2$=0.9632，Adjusted $R^2$=0.9826，F-Stat.=521.147，DW-Stat=0.4665 | | | | |

**Table 2. The cross-sectional estimate of influence coefficients of various province from 1978 to 2007**

| Province | $\eta_i$ | Province | $\eta_i$ | Province | $\eta_i$ | Province | $\eta_i$ |
|---|---|---|---|---|---|---|---|
| Beijing | 0.1551 | Shanghai | 0.4440 | Hubei | 0.3215 | Yunnan | 0.2717 |
| Tianjing | 0.1316 | Jiangsu | 0.3564 | Hunan | -0.1819 | Shanxi | -0.2021 |
| Hebei | 0.3704 | Zhejiang | 0.1954 | Guangdong | 0.1238 | Gansu | -0.3782 |
| Shanxi | -0.2962 | Anhui | 0.1720 | Guangxi | -0.1373 | Qinghai | -0.3994 |
| Inner Mongolia | -0.0777 | Fujian | 0.3934 | Hainan | 0.3018 | Ningxia | -0.0905 |
| Liaoning | 0.3742 | Jiangxi | -0.1314 | Chongqing | -0.9655 | Xinjiang | -0.1368 |
| Jilin | 0.1558 | Shandong | 0.6441 | Sichuan | -0.3181 | | |
| Heilongjiang | 0.2090 | Henan | -0.3460 | Guizhou | -0.2232 | | |

According to the regression result, $DW_{1978\text{-}2007}$ = 0.4665, it shows that there is autocorrelation between the variables, and then we test autocorrelation of spatial - regression error terms, the following estimate of the model are used with Spatial Econometric Modules of Matlab7.0, the results of estimation are showing in the table 3.

**Table 3. The spatial correlation test**

| n =900 | Lmerr | Lmsar | Lratios | Moran'I | Walds |
|---|---|---|---|---|---|
| value | 66.5840 | 79.6035 | 86.6162 | 26.3590 | 21.6285 |
| chi(1) .01 value | 17.6110 | 6.6350 | 6.6350 | 1.9657 | 6.6350 |
| Prob. | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

With the test results, five test values (spatial dependence) are very significant (Prob. =0.0000), it prove that there is a significant spatial correlation between the logistics and regional. Thus the spatial factors must be taken into account in order to show the interaction between various regions GDP and logistics. The test value of spatial-panel lag term is bigger than the test value of spatial-panel error term, that is, $Lmerror_{1978\text{-}2007}$=66.5840 <$Lmsar_{1978\text{-}2007}$=79.6035;
$Lmerror_{1978\text{-}2007}$=66.5840 <$Lmsar_{1978\text{-}2007}$=79.6035.

Based on the criteria described previously, the Sar-panel lag model is the optimal model. The spatial-panel lag model is used to estimate the correlation between economic growth and the logistics. Results are showing in the table 4 and table 5:

**Table 4. The estimation results of Sar-panel model parameter from 1978 to 2007**

| | LnK | LnL | LnW | $\rho/\lambda$ |
|---|---|---|---|---|
| $\beta$ | 0.5609 | 0.3040 | 0.0848 | 0.1250 |
| t-Stat | 67.5058 | 15.3086 | 2.9924 | 8.8159 |
| Prob. | 0.0000 | 0.0000 | 0.0064 | 0.0000 |
| R-squared=0.9775，Rbar-squared=0.9758，sigma^2=0.0308，log-likelihood=219.95985 | | | | |

**Table 5. The estimate of spatial-fixed influence coefficient of various regions from 1978 to 2007**

| Province | $\eta_i$ | Province | $\eta_i$ | Province | $\eta_i$ | Province | $\eta_i$ |
|---|---|---|---|---|---|---|---|
| Beijing | 0.2365 | Shanghai | 0.5753 | Hubei | 0.4321 | Yunnan | 0.3234 |
| Tianjing | 0.2145 | Jiangsu | 0.5638 | Hunan | -0.1143 | Shanxi | -0.1761 |
| Hebei | 0.4582 | Zhejiang | 0.4631 | Guangdong | 0.4542 | Gansu | -0.3012 |
| Shanxi | -0.3758 | Anhui | 0.2315 | Guangxi | -0.1004 | Qinghai | -0.2874 |
| Inner Mongolia | -0.1436 | Fujian | 0.4327 | Hainan | 0.3570 | Ningxia | -0.1923 |
| Liaoning | 0.4653 | Jiangxi | -0.3421 | Chongqing | -0.8541 | Xinjiang | -0.1028 |
| Jilin | 0.2153 | Shandong | 0.6743 | Sichuan | -0.2181 | | |
| Heilongjiang | 0.3645 | Henan | -0.2653 | Guizhou | -0.1843 | | |

In table 4 and 5, with the results of model estimation, we can draw the following conclusions. Firstly, in the estimation results, the estimation of the parameters $\rho$ in the spatial-panel lag model significance test is by 1%. It indicates that there is a significant spatial correlation between GDP and logistics in 30 provinces. As the logistics has the network properties, the logistics can connect economic activity into a whole unit. Through the spatial overflow (diffusion) benefit, the rapid economic growth regions drive the economic development of slower economic growth regions. It demonstrates positive spillover effect. Meanwhile the logistics will have a negative spillover effect, production factors flow easily into developed regions, the economic growth in a region is likely to be on the expense of economic recession in other regions. Secondly, the fitting of $R^2$ value which we introduced the spatial and time fixed-effects into the spatial-panel lag regression model are better than that of the traditional fixed-effect model. It proves that it can explain the model better and show the actual situation better after we introduce the time and spatial fixed effects into the model. The elasticity coefficients of GDP with capital stock and labour are 0.56 and 0.30, respectively. The significant level is 1%, indicating that the effect of investment and labour on economic growth is still the most important factor, the elasticity coefficient of GDP on logistics is 0.08, the significant level is 0.64%. Indicating that the logistics has a significant impact on GDP, but the degree of influence on economic growth is limited, which is showing the current situation of China. i.e. modernization of Chinese logistics is not high, the logistics network system is imperfect, it is lack of the application of information technology and the level of logistics management is low.

Thirdly, with the estimate results of spatial-fixed influence parameter ($\eta_i$) of the various region, the fixed-effect parameters of different regions are showing significant difference. The logistics development level is better in the developed regions, worse in the developing regions. There are three levels of logistics development in China. The best one is developed coastal areas, such as Shanghai city, Jiangsu province, Zhejiang province and other developed coastal areas, second level is middle level which include central and north-eastern region of China, such as Jiangxi province, Hunan province, Hubei province, Liaoning province, the third level is the worst one include north-western region of China. It is consistent with the actual development situation in various regions of China.

## 4. CONCLUSION

Based on spatial-panel model, we estimated the correlation between the logistics and economic growth. We can draw the conclusion that there is significant spatial-correlation between GDP and logistics in various regions, the GDP and logistics has obvious spatial overflow (diffusion) benefit between adjacent regions. The logistics has a significant impact on local GDP, but the degree of influence on economic growth is limited, the reasons are low degree of logistics modernization in China, the imperfect system of logistics network, lack of application of information technology and the low level of logistics management. Fixed effect parameters of different regions are showing significant differences. The local economic development is better; the local logistics development level is higher.

## REFERENCES:

[1]. **Anselin, L.** (1988) *Spatial Econometrics. Methods and Models*. Dordrecht:Kluwer Academic (Chapter 4)

[2]. **Anselin, L.** (1988b) *Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity*, Journal of Geographical Analysis, 20, 1-17

[3]. **Baltagi, Badi H.** (2001) *Econometric. Analysis of Panel Data*. London: John Wiley&Sons, Chichester, United Kingdom

[4]. **Elhorst, J.** (2003) P-*Specification and Estimation of Spatial Panel Data Models*, Journal of International Regional Science Review, 26, 244-268

[5]. **He Jiang Zhang Xin Zhi** (2006) *Chinese Regional Economic Growth and Convergence: Spatial Panel Data Analysis*, South China Journal of Economics, 5, 44-52

[6]. **Li Xu Ying, Gu Lan** (2004) *The Spatial Autoregressive Model and the Relevant Estimation*, Journal of Statistical Research, 6, 48-51

[7]. **State Bureau of Surveying and Mapping** (2008) *The Map of the Republic of China,* China:Sino Map Press