Abstract: Since the open door policy declared in the late 1970s, the economic growth of China has been rapidly developed and accumulated. However, the widening regional economic development disparity has been brought to the concern by the government. It is doubtful that the regional economic growth would tend to be equivalent by spillover effects from some more-developed regions to other less-developed ones. The goal of this paper is to examine the long-run relationship of Chinese provincial economic performance with the consideration of the characteristic of cross-sectional dependence in the panel data covering 30 provinces in China over the period 1990-2012. We find a strong spatial dependence over in China’s regional production function. After a cointegrating relation is confirmed using the methodology of Kao (1999) and Pedroni (1999), a spatial error correction model is further applied. We find the local cointegration term is significantly negative, suggesting a long-run convergence relation for the Chinese regional economic growth.

Keywords: Regional Cointegration, Spatial Error Correction Model, Regional Convergence

1. Introduction

China’s economy has been experiencing a rapid growth rate since the economic reforms in the late 1970s that China’s GDP growth rate has maintained an annual average of 9% over the last two decades. Furthermore, recent empirical literatures on economic growth have also found that at cross-regional level the income disparity in China has been increasing across regions over the recent years (Xu and Zou, 2000; Maasoumi and Wang, 2008; Zou et al., 2008 Zhou, et al., 2014; Liao and Wei, 2015). The widening regional economic disparity has brought about an important issue in Chinese economic development. It shows the evidence of a significant cross-sectional dependence in the GDP distribution, which displays a clear spatial-clustered effect in the coastal provinces. LeSage (1999) suggests that observations associated with spatial units such as zip-codes, counties, states, census tracts and so on, should consider spatial correlated, or might reflect measurement errors.

Due to the shortcoming of the conventional tests of spatial dependence, which is usually exogenously given a connection or spatial matrix for regional relevance, this paper applies the test for spatial correlation in residuals of the estimated model and adopts the test for cross-sectional dependence (CD test) developed by Pesaran (2004), which allows the panel data with a short time span $T$ and a large number of cross sections $N$. Since our panel data consist of 30 provinces of China for the periods 1990 to 2007, it is suitable for the assumption of Pesaran (2004).

As mentioned above, the spatial correlation effect in panels should be considered. Indeed, the panel data might be nonstationary by allowing for a feedback effect from the current dependent variable to the future explanatory variables. However, the spatial econometric researches using panel data usually ignore the issue of non-stationarity by assuming panel data are the temporally static (Elhorst, 2003; 2009). Moreover, a general panel data model also usually assumes the spatial independence. In this paper we attempt to integrate the issue of space and time in panel data and argue that the spatial variables are cointegrated in the long-run. Beenstock and Felsenstein (2010) propose the method of spatial vector error correction model (SpVECM) to overcome these problems. The SpVECM concurrently contains both the temporal dynamics and spatial dynamics by including the local cointegration and spatial cointegration in the estimated model. Local cointegration indicates that the nonstationary panel data are cointegrated “within” spatial units but not “between” them, while spatial cointegration indicates the non-stationary variables are cointegrated “between” spatial units but not “within” them (Beenstock and Felsenstein, 2010; Baker,
The purpose of this paper seeks to apply the panel data with the characteristic of cross-sectional dependence and examine the long-run relationship of Chinese provincial economic performance. Using an annual panel dataset for 30 provinces in China over the period 1990-2007, we provide the procedure of statistical and systematical tests for the panel data with the characteristic of cross-sectional dependence and apply the spatial error correction model provided by Beenstock and Felsenstein (2010) to measure the long-run effect on provincial GDP growth in China.

2. Methodology and Data

2.1. Determinants of Regional Production Function

In order to examine the long-run impact on regional comparisons of Chinese economic performance, we apply a dynamic spatial panel production function, which is augmented from the growth model of Solow (1957) and contemporaneously consider the feature of human capital (Mankiw et al., 1992). An augmented Cobb-Douglas production function for Chinese provincial output is employed in this study as follows.

\[ Y_i = A_i K_i^\alpha L_i^\beta H_i^\gamma e_i \]  

where \( Y \) is provincial output, \( K \) is physical capital stock, \( L \) is labor force, and \( H \) is human capital. The subscripts \( i \) and \( t \) refer to the province \( i \) and the current year \( t \). In general, we anticipate that the impacts of physical capital stock, labor force, and human capital on provincial output are positive. \( A \) is the technical factor. Based on the previous related studies, we know that the technological channels come from two sources: (1) inward FDI and international technology import; (2) internal leading research and development (R&D) activities. The international technology sources indicate that multinational enterprises (MNCs) usually have advantaged technologies, which can spill over to local firms through technological externalities. Moreover, because the in-house R&D activities were not quite spirited in the early 1990s, China’s spectacular economic growth is believed to be largely contributed by the implementation of its “open door policy” in the late 1970s. It implies the foreign-owned enterprises play a more important role for China’s economic growth relative to the in-house R&D. Therefore, we emphasize more on the role of FDI to determine its influence on regional economic performance. We assume the technical factor \( A \) is associated with the provincial inward FDI. Finally, the error term which reflects the effects of unknown factors. We rewrite equation (1) as a log-linear form of the augmented Cobb-Douglas production,

\[ \ln GDP_i = \beta_0 + \beta_1 \ln CAP_i + \beta_2 \ln LAB_i + \beta_3 \ln HUM_i + \beta_4 \ln FDI_i + \epsilon_i \]  

where \( \ln GDP_i \) is the logarithm of regional gross domestic product (GDP) in region \( i \) at year \( t \). \( CAP \) and \( LAB \) are the two core physical inputs that denote fixed capital stock and labor force, respectively. We follow Yao (2006)’s setting of perpetual inventory method for fixed capital stock, assuming that the depreciation rate is 7.5% and the growth rate of capital is 15%, which imply the average life of capital equipment is 13.3 years. Labor force is measured as the labor population who are older than 15 years old in region \( i \) and at year \( t \). \( HUM \) is measured as the number of students per 10,000 population enrolled in the higher education level. The error term \( \epsilon \) has the characteristic of heterogeneity, which is assumed to include a common regional factor that does not vary over time. That is, \( \epsilon_i = u_i + \nu_i \), where \( u \) is a region-specific effect that varies across regions but is fixed within a region over time, and \( \nu \) is a “white noise” error term.

2.2. Testing for Cross-Sectional Dependence in Panel Data

The conventional tests of spatial dependence are usually exogenously given a connection or spatial matrix for regional relevance. But the terminal inference and analysis of the spatial dependence test could critically depend on the choice of the spatial matrix. We adopt a cross-sectional dependence (CD) test in panel data proposed by Pesaran (2004). The test statistic is described as below.

\[ CD = \frac{2T}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} \hat{\rho}_{ij} \sim N(0,1) \]  

where \( \hat{\rho}_{ij} \) is the sample estimate of the pair-wise correlation of the residuals, \( \hat{\rho} \), \( \sum_{i=1}^{N} \epsilon_i^2 = \epsilon_i^2 \), \( \sum_{j=1}^{N} \epsilon_j^2 = \epsilon_j^2 \), \( \sum_{i=1}^{N} \epsilon_i^2 \), \( \sum_{j=1}^{N} \epsilon_j^2 \)

The null hypothesis of CD test assumes no cross-sectional or spatial dependence among the sample. The CD test can provide sufficient evidence whether the cross-sectional correlations are present between pairs of regions for all variables. Using the data we collect for the paper, the statistics of CD test is 53.23 (p-value=0.000) in the fixed effect panel model of equation (2) and 56.16 (p-value=0.000) in random effect panel model of equation (2), respectively. The results strongly indicate the existence of cross-sectional dependence in Chinese provincial production function model.

2.3. Testing for Stationarity in Panel Data

Pesaran (2007) proposes an individual cross-sectional augmented Dickey-Fuller (CADF) test to examine stationary of variables in panel data, which modifies the usual Im, Pesaran and Shin (2003)’s unit root test (IPS) with the lagged cross-sectional mean and its first difference to capture the cross-sectional dependence arising through a single factor model. The CADF test is described as follows.

\[ \Delta Y_{it} = a_t + b_t \gamma_{it} + c_t \delta_{it} + \sum_{j=0}^{p} \Delta_{it} Y_{j,t} + \sum_{k=0}^{q} \Delta_{it} \Delta_{j,t} + \epsilon_{it} \]  

where \( \Delta_{it} \) is the first difference of \( Y_{it} \) for each region \( i \) and each period \( t \). The CADF test is performed by regressing the first difference of \( Y_{it} \) on the lagged levels of \( Y_{it} \) and its first differences.
where $\bar{y}_t = N^{-1} \sum_{j=1}^{N} y_{jt}$, which is the average $y_{jt}$ of all provincial observations at time $t$. $p$ is the lagged order of the model. The null hypothesis of the test can be expressed as $H_0: \beta_i = 0$ for all $i$ against the alternative hypothesis $H_1: \beta_i < 0$ for some $i$. We denote $t_i(N,T)$ as the $t$-statistic of $\beta_i$ in the CADF regression. The $t$-statistic is used to construct a cross-sectionally augmented version of the IPS test:

$$CIPS(N,T) = N^{-1} \sum_{i=1}^{N} t_i(N,T) \quad (5)$$

where $t_i(N,T)$ is the cross-sectional augmented Dickey-Fuller statistics for the province $i$. In addition, to avoid the problem of an extreme statistic caused by a small sample, Pesaran (2007) constructs a truncated version of the CIPS test, which is denoted as a $CIPS^*$ test.

$$CIPS^*(N,T) = N^{-1} \sum_{i=1}^{N} t_i^*(N,T) \quad (6)$$

where

$$t_i^*(N,T) = \begin{cases} t_i(N,T), & \text{if } -K_1 < t_i(N,T) < -K_2 \\ -K_1, & \text{if } -t_i(N,T) \leq K_1 \\ -K_2, & \text{if } -t_i(N,T) \geq K_2 \end{cases}$$

The parameters $K_1$ and $K_2$ are positive constants based on the simulated values of $CADF$. For models with an intercept and trend, Pesaran (2007) suggests using $K_1 = 6.42$ and $K_2 = 1.71$.

### 2.4. Testing for Panel Cointegration

To avoid the problem of spurious regression, we must investigate the cointegration correlation of variables. If $y$ and $x$ are spatially panel unit root variables with the same order of integration, SI(1), in general any linear combination of $y$ and $x$ is also SI(1). If a linear combination $y - \beta x$ is stationary, it implies the linear combination of $y$ and $x$ has a cointegrated relationship.

Kao (1999) and Pedroni (1999), which are two well-known cointegration tests, propose residual-based cointegration tests. Kao (1999) develops the DF and ADF-type unit root tests for the error term of the panel regression model as a test for the null of no cointegration. There are four DF-type tests, $DF_p$, $DF_f$, $DF_{p^*}$ and $DF_{f^*}$, while $DF_p$ and $DF_f$ are based on the strong exogeneity of the regressors and errors, $DF_{p^*}$ and $DF_{f^*}$ are for the cointegration with endogenous relationship between regressors and errors. The null hypotheses of Kao (1999)'s cointegration tests assume a cointegration in a panel data model.

Pedroni (1999, 2004) proposes several tests for the null hypothesis of no cointegration in a panel data model, allowing for considerable heterogeneity. Seven tests follow asymptotically a standard normal distribution after a suitable normalization. Among them, four of these statistics, called panel cointegration statistics, are within-dimension based statistics. The other three statistics, called group mean panel cointegration statistics, are between-dimension panel statistics. The former statistics pool the autoregressive coefficients across different members for the unit root tests on the estimated residuals, while the latter are based on estimators that simply average the individually estimated coefficients for each member. The null hypothesis of no cointegration is tested ($H_0: \rho_i = 1 \forall i$), while the alternative hypothesis of between-dimension cointegration statistics is $H_1^b: \rho_i < 1 \forall i$.

### 2.5. The Estimation of Regional Production Function

Assume that $y$ and $X$ are the variables with a spatial panel unit root (non-stationary), and therefore the combination of $y$ and $X$ could be cointegrated. Moreover, according to the result of $CD$ test, we evidence the existence the cross-sectional correlation in equation (2). We can rewrite equation (2) and present the regional production function with spatial lag terms as below.

$$y_{it} = \alpha + \psi y_{it} + \theta \bar{y}_{it} + \sum_{k=1}^{4} \beta_k X_{it} + \sum_{k=1}^{4} \delta_k X_{it} + u_{it} \quad (7)$$

where $i$ denotes the cross-sectional dimension, $t$ denotes the time-series dimension. $y_{it}$ is the logarithm of provincial GDP. $Z$ is a vector of observed common factors that are assumed to affect all spatial units. $X$ denotes a matrix of other explanatory variables, presented as the logarithm type, including physical capital stock ($CAP$), labor force ($LAB$), human capital ($HUM$), and foreign direct investment ($FDI$). The superscript asterisked variables refer to spatial lags terms which are calculated by the sum of spatially weighted values of neighboring locations. They are defined as below.

$$y_{it}^s = \bar{w} y_{it} = \sum_{j=1}^{N} w_{ij} y_{jt} \quad ; \quad X_{it}^s = \bar{w} X_{it} = \sum_{j=1}^{N} w_{ij} X_{jt} \quad (8)$$

where $w_{ij}$ is the spatial weight, $w_{ij} = 1/distance_{ij}$, $i \neq j$; $w_{ij} = 0, i = j$. We use the longitude and latitude of provincial capital cities to calculate the straight distance from region $i$ to region $j$. The spatial weight matrix can be defined as below.

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\
 w_{21} & w_{22} & \cdots & w_{2n} \\
 \vdots & \vdots & \ddots & \vdots \\
 w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix} \quad (9)$$
The spatial weight matrix intends to capture the spatial spillover effect of other neighboring provinces and gives each spatial unit the relative weight. However, to ensure the sum of weight in each row equal to one, we adjust the spatial weight matrix to a row-standardized matrix. The row-standardized spatial weight matrix is presented as follows.

\[
\bar{w}_{ij} = \frac{w_{ij}}{\sum w_{ij}}, \quad i, j = 1, 2, \ldots, n.
\]

In equation (3), as the variables are cointegrated, implies that at least exists an one-way error correction in the spatial panel error correction model (SpECM). We follow the setting of Beenstock and Felsenstein (2010)'s SpECM. The estimation equations are described as below.

\[
\Delta y_t = \gamma_0 + \gamma_1 \Delta y_{t-1} + \sum_{i=2}^{4} \gamma_i \Delta y_{t-i} + \sum_{i=2}^{4} \Delta \Delta y_{t-i} + \gamma_d x_{t-1} + \gamma_h x_{t-1} + \gamma_e z_{t-1} + v_t
\]

where \( u \) and \( u^* \) are derived from the estimated residuals of equation (3) \( (u^* = \bar{W} u) \). \( v_t \) is the residual of SpECM. We denote \( u \) is the “local error correction” term, while \( u^* \) is the “spatial error correction” term. Using the SpECM, we can find the dynamic long-run equilibrium in the model and carry out the causal relationship tests.

When \( \gamma_5 \) is non-zero, there is local error correction; and when \( \gamma_6 \) is non-zero, there is spatial error correction. When both \( \gamma_5 \) and \( \gamma_6 \) are non-zero, i.e. both types of error correction occur, we refer to this as “global error correction”. In equation (11), \( \gamma_5 \) is expected to be negative since \( \gamma_6 \) is expected to decrease as it corrects itself toward to its equilibrium value (i.e. so-called convergence), while \( \gamma_6 \) is expected to be positive since a positive neighboring spillover is expected to increase the local GDP.

The data used in this study contain 30 provinces in China for 23 consecutive years, 1990-2012, yielding 690 observations. The period is sampled because China’s economic activity has risen significantly since the early 1990s. We ignore the Tibet autonomous province from our dataset because Tibet lacks for several data in control variables. Therefore, the dataset used in exploring the long-run relationship among the regional production function with cross-sectional dependence can provide valuable insights for Chinese provincial production function cointegrated relation.

Table 1 shows the the definitions of the variables used in our analysis and the basic statistic description.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Definition</th>
<th>Mean (Std. Dev)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>Gross domestic Product for i province (millions RMB)</td>
<td>344,871.7 (447,108.5)</td>
</tr>
<tr>
<td>CAP</td>
<td>Capital stock: it is calculated using the perpetual inventory method (millions RMB)</td>
<td>1.63e+07 (1.25e+08)</td>
</tr>
<tr>
<td>LAB</td>
<td>Labor population who are older than 15 years old (10,000 persons)</td>
<td>4,937.283 (9,304.88)</td>
</tr>
<tr>
<td>HUM</td>
<td>Human capital: Number of students enrolled in high education per 10,000 population (persons).</td>
<td>74,44084 (88,887)</td>
</tr>
<tr>
<td>FDI</td>
<td>Foreign direct investment inflow (RMB million), adjusted by annual average exchange rate.</td>
<td>145,768.3 (274,608)</td>
</tr>
</tbody>
</table>

Notes: The means and standard errors are calculated by pooling data for the 1990-2012 period. All monetary figures are at 2002 prices.

3. Empirical Results

In the preliminary step we apply the panel unit root test provided by Pesaran (2007), which allows cross-sectional dependence in the panel data model. The model used to test the unit root hypothesis is the one with intercept and trend. We set the lag lengths order at the maximum equal to 2 \((p=0, 1, 2)\). The results indicate all of the variables exhibit non-stationarity in both CIPS and CIPS* statistics. We further take the first differencing for the series of these variables and the results show they become stationary, implying that these variables could be integrated of order one. Given that regional GDP and the inputs of regional GDP have a panel unit root, we further examine two types of panel cointegration tests developed by Kao (1999) and Pedroni (1999, 2004) to evidence the existence of a long-run relationship among the variables included in the regional production function model. The results of panel cointegration tests suggest the rejection of null hypothesis of no cointegration among the variables in the regional production function at the 5% statistical level except Kao (1999)'s DF and DF statistics. The estimated results imply that the variables are clearly cointegrated and hence exists an error correction relationship, suggesting that an error correction model is necessary to be used in the estimation.

To obtain a comparison base, we report the ordinary least square (OLS) estimation results for the benchmark model, which are presented in Table 2. The empirical results obtained by using fixed effect (FE) model are reported in Models 1 and 2. We perform the cross-sectional dependence (CD) test provided by Pesaran (2007), AR(1) test, and also adopt Breusch and Pagan (1979)'s Lagrangian multiplier (LM) test for random effects to test the heteroskedasticity in error terms. Furthermore, in order to check the robustness of estimated results, we adopt the feasible generalized least square (FGLS) estimation in Model 3, which is to capture the heteroskedasticity with cross-sectional correlation and common AR(1) coefficient for all panels.
In Model 3, the estimated results show the variables $CAP$, $LAB$, and $HUM$ are significantly positive, and the variables $CAP^*$, $LAB^*$, and $HUM^*$ are significantly negative effects at 1% statistic level. We consider that the impacts of spatial lagged physical capital, labor force, and human capital on Chinese regional GDP exist a crowding out effect. As the neighboring provincial GDP increases, it will crowd out the input resource of local production function, which in turn worsens the local economic performance. The estimated coefficients of $FDI$ and $FDI^*$ represent the effects of international and regional technologies spillovers, respectively. The empirical results indicate that the estimated coefficient of $FDI$ is 0.013 and significant at 1% statistical level and the estimated coefficient on the spatial lagged term of $FDI$ is 0.0951 and significant at 1% statistical level. It clearly indicates that there are strong spillover effects on international and regional technologies. Moreover, the estimated coefficient on variable $GDP^*$ in Model 3 is positive and significant at 1% statistical level, supporting the existence of spatial spillover effect of regional growth.

Table 3 reports the estimated spatial error correction model (SpECM) for regional economic performance in China. The estimation of SpECM is utilized by a maximum likelihood (ML) approach (Anselin, 1988).

In Model 3, we also test the autoregressive coefficient of estimated residuals, AR(1), and the Lagrangian multiplier (LM) test for random effects developed by Breusch and Pagan (1979). The estimated results indicate the existence of heteroskedasticity with cross-sectional correlation and AR(1) problem in panels. Therefore, we adopt the generalized least square estimation to control the problems of heteroskedasticity and AR(1) in Model 3.

### Table 2. The benchmark models of regional production function.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 1 (FE) Coeff. (Std. err.)</th>
<th>Model 2 (FE) Coeff. (Std. err.)</th>
<th>Model 3 (FGLS) Coeff. (Std. err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAP</td>
<td>0.2398*** (0.0199)</td>
<td>0.0059 (0.0055)</td>
<td>0.0339*** (0.0053)</td>
</tr>
<tr>
<td>LAB</td>
<td>0.0042 (0.0316)</td>
<td>0.0721** (0.0337)</td>
<td>0.9303*** (0.0408)</td>
</tr>
<tr>
<td>HUM</td>
<td>0.5348*** (0.0301)</td>
<td>0.0146 (0.0301)</td>
<td>0.4398*** (0.0404)</td>
</tr>
<tr>
<td>FDI</td>
<td>0.2069*** (0.0142)</td>
<td>0.0275*** (0.0041)</td>
<td>0.0130*** (0.0005)</td>
</tr>
<tr>
<td>GDP*</td>
<td>1.0020*** (0.0317)</td>
<td>1.0782*** (0.0317)</td>
<td>0.9020*** (0.0317)</td>
</tr>
<tr>
<td>CAP*</td>
<td>-0.0169 (0.0261)</td>
<td>-0.2498*** (0.0261)</td>
<td>-0.0169 (0.0261)</td>
</tr>
<tr>
<td>LAB*</td>
<td>-0.0701 (0.0341)</td>
<td>-0.9321*** (0.0341)</td>
<td>-0.0701 (0.0341)</td>
</tr>
<tr>
<td>HUM*</td>
<td>0.0061 (0.0320)</td>
<td>-0.4051*** (0.0320)</td>
<td>0.0061 (0.0320)</td>
</tr>
<tr>
<td>FDI*</td>
<td>-0.0207 (0.0105)</td>
<td>0.0951*** (0.0105)</td>
<td>-0.0207 (0.0105)</td>
</tr>
<tr>
<td>intercept</td>
<td>4.7354*** (0.2617)</td>
<td>0.541 (0.1108)</td>
<td>0.6142*** (0.446)</td>
</tr>
<tr>
<td>R2</td>
<td>0.893</td>
<td>0.949</td>
<td>0.902</td>
</tr>
<tr>
<td>CD test</td>
<td>53.225***</td>
<td>-2.575</td>
<td></td>
</tr>
<tr>
<td>AR(1) test</td>
<td>982.177***</td>
<td>1194.89***</td>
<td></td>
</tr>
<tr>
<td>B-P LM test</td>
<td>608.45***</td>
<td>2849.24***</td>
<td></td>
</tr>
<tr>
<td>Hausman test</td>
<td>599.77***</td>
<td>3735.94***</td>
<td></td>
</tr>
<tr>
<td># of obs.</td>
<td>630</td>
<td>630</td>
<td>630</td>
</tr>
</tbody>
</table>

Notes: All variables are expressed in logarithms. *, **, and *** indicate the parameters that are significance at 10%, 5%, and 1% probability level respectively.
by estimating a spatial error correction model, we find the production function are cointegrated, which are tested using the methodologies of Kao (1999) and Pedroni (1999). Finally, we evidence that the variables among the regional dependence in panel and given the variables among the local cointegration term is significantly negative, suggesting a long-run convergence relation for the Chinese regional economic growth.

### 4. Conclusion

The empirical findings can be summarized as follows. First, the estimated results clearly indicate that a strong spatial dependence over the variables in China’s regional production function. Second, based on the presence of cross-sectional dependence in panel and given the variables among the regional production function have the same order of integrated, we evidence that the variables among the regional production function are cointegrated, which are tested using the methodologies of Kao (1999) and Pedroni (1999). Finally, by estimating a spatial error correction model, we find the local cointegration term is significantly negative, suggesting a long-run convergence relation for the Chinese regional economic growth.

### References


