Spatial dependence of the level of urbanization and its economic mechanism- A spatial analysis in demography

Wang Weijin, Ph. D student; Lu Jiehua, Professor
Department of Sociology, Peking University, China
Email: Wangweijin311@163.com, lujiehua@pku.edu.cn

Abstract: Through exploratory spatial analysis, this paper found strong spatial dependence in levels of urbanization among 287 prefecture-level cities in China, and the high-high and low-low were the dominant spatial agglomeration types among the neighbor cities. The major socio-economic factors also showed obvious spatial dependence, which might contribute to the spatial agglomeration of the urbanization level. Spatial error regression model further verified and controlled this kind of spatial dependence, and indicated that the degree of openness relatively more significantly increased the urbanization level than the degree of industrialization due to household registration system, industrialization strategy as well as industrialization process in China.

Keywords: prefecture-level city, urbanization, spatial dependence, Moran’s I, spatial error model.

Introduction

For a long time, China’s urbanization shows significant regional variation (Zhang, 2002; Liu et al., 2007). From the east coast to the interior area, the spatial distribution of the level of urbanizations is extremely unbalanced (Zheng, 2007), while city agglomerations always exist and dynamically change.

Then an immediate question is: Which factors lead to the disparities of the level of urbanization and city agglomerations? Or, what’s the intrinsic mechanism which shapes the spatial distribution pattern of the level of urbanization in China?

Theoretical frame

In response to this issue, rather than from the spatial relationship perspective,
most of previous studies are limited to discussion of the differentials of regional economic development in a broad sense. However, a spatial relationship is fundamental to understand the urban systems in the human ecology (Logan & Molotch, 1987). The growth poles theory proposed the unbalanced economic growth, which argued that a region does not grow economically at the same rate over all of its areas but rather tends to grow more rapidly at one point (growth pole) and more slowly in another point (lagging region), this happens because of establishment of propulsive industry or growth pole that maintains interdependence linkages with other industries or economic activities and stimulates economic growth (Perroux, 1955). The spatial diffusion theory argued contagious spread of things produces “neighborhood effect” of clustered growth (Hudson, 1972), in other words, population growth may diffuse around the neighbor areas, and likewise urbanization may show spatial autocorrelation. Spatial demography follows the first law of geography that “everything is related to everything else, but near things are more related than distant things” (Tobler, 1970).

So, the first objective of this study was to introduce the spatial perspective into the analysis of variation of the urbanization level in China, and test whether the spatial dependence of urbanization level existed or not.

Some studies have discussed the differentials of the urbanization level at provincial or above levels, while neglected the great variation of many important aspects within provinces, such as natural conditions, socio-economic development, regional policies on reforms and opening up as well as urbanization strategies. So, the second objective of this paper was to explore the differentials of urbanization in prefecture-level cities rather than provinces in China.

Besides spatial autocorrelation, we also wanted to test the effects of economic factors, such as degree of industrialization and openness, and explore whether they functioned as intrinsic mechanism of spatial dependence, which few previous studies did.

**Data and measurement**
1. **Data source**

The data source of this research was *2008 China City Statistical Yearbook* (National Bureau of Statistics of China, 2008). In total, there were 287 prefecture-level or above cities with aggregate data in 2007, which included 4 municipalities (Beijing, Shanghai, Tianjin and Chongqing), 15 sub-provincial cities (Guangzhou, Hangzhou, Nanjing, Wuhan, Chengdu, Xi’an, Jinan, Shenyang, Changchun, Harbin, Shenzhen, Xiamen, Dalian, Qingdao and Ningbo) and 268 prefecture-level cities. From here on, we will call all of them prefecture-level cities.

2. **Variables and measurements**

The urbanization rate and non-agricultural rate are two most common indicators measuring urbanization level in China. Basically, the urbanization rate is the proportion of population living in urban administrative areas of a region, while the non-agricultural rate is the proportion of population with a non-agricultural household registration status of a region, some of whom may live in rural. It is a common sense in China that it would overestimate the proportion of urban population (urbanization level) to some extent if we use de facto population living in urban administrative areas because many China's urban administrative areas contains some portions of rural areas where rural population lives. On the contrary, it could substantially underestimate the urbanization level because non-agricultural population measures de jure population, which does not count both short-term and long-term migrants without a local household registration status. Generally, the non-agricultural rate tends to be lower than the urbanization rate. Considering the data availability and measurement consistency in all cities, the proportion of non-agricultural population was chosen to be the final dependent variable, which equals to the total number of population with non-agricultural household registration status of a prefecture-level city divided by its total population at the end of 2007. Note the new definition by the National Bureau of Statistics of China which was used in the sixth census in 2010 may be more accurate, since it counted all persons as citizens if they had been living in the city districts for more than 6 months when they were interviewed.

This research included the following independent variables: the annual GDP
growth rate (%) reflecting the speed of the overall economic development of the city in 2007, the proportion of the second industry accounting for the total GDP (%) which showed the degree of industrialization, the percent of the foreign-funded (including funding from Taiwan, Hong Kong and Macao) enterprises in the total enterprises (%) which indicated the degree of the openness. The degree of industrialization and openness are of most interest in this study.

According to literature research, the control variables included per capita land area (km²) and annual per capita meat yield (kg) measuring the basic geographic/agricultural environment, the number of doctors and beds in hospital in 10 thousand people as well as the number of ordinary secondary school students in 10 thousand people to indicate the degree of social development.

**Analysis strategy and methods**

To deal with spatial autocorrelation issue which lead to violation of the independent condition of normal OLS regression (Wu, 2007), spatial methods were used, and it was conducted in four steps as follows.

First, we examined spatial distributions of the main variables, calculated Moran’s I and created Moran plot to see the overall (global) spatial interdependent trend of the level of urbanization. Moran’s I is a global indicator reflecting the overall trend of spatial autocorrelation. The formulas is

\[
I = \frac{\sum \sum w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum \sum w_{ij}y_iy_j}
\]

While \( n \) is the number of the spatial regions, \( y_i \) is the value of variable \( y \) in region \( i \), \( y_j \) is the value of variable \( y \) in region \( j \), \( \bar{y} \) is the mean of \( y \). \( w_{ij} \) is a \( n \times n \) neighborhood matrix, if two regions are neighbors, then \( w_{ij} = 1 \), otherwise \( w_{ij} = 0 \). Just like Pearson’s correlation coefficient \( r \), Moran’s I ranges from -1 to 1, if the statistic is close to 1, it indicates a strong positive spatial autocorrelation, and if the statistic is close to -1, it indicates a strong negative spatial autocorrelation. As a visual tool to demonstrate the global spatial trend, the Moran plot is just a lagged \( y \) vs. \( y \) two-way scatterplot.
Because lagged $y_i = \left( \sum_j w_{ij} y_j \right)$ is the mean of $y$ values of the neighbors of region $i$ weighted by spatial neighborhood matrix, so we can see the correlation between the average urbanization levels of neighbors of regions and themselves.

Then, we used Local Indicators of Spatial Association (LISA) to classify the four kinds of urbanization agglomerations, figured out the local spatial dependence of the level of urbanization if any and tried to give an explanation.

Finally, we specified one appropriate spatial regression model to control the found spatial autocorrelation if any to test the net effects of degree of industrialization and openness on the level of urbanization. Spatial lag model and spatial error model are two common spatial models, they can be written as:

$$y_i = X \beta + \rho \sum_j w_{ij} y_j + \epsilon_i$$

and

$$y_i = X \beta + \lambda \sum_j w_{ij} \xi_j + \epsilon_i$$

, respectively. While $X \beta$ are the effect of covariates, $\epsilon_i$, $\xi_i$ are the error terms. $\sum_j w_{ij}$ and $\sum_j w_{ij} \xi_j$ are the spatial lagged variables based on dependent variable and the error term. Note $\epsilon_i$ in the spatial error model is the error term in the OLS model ($y_i = X \rho + \epsilon_i$). $\rho$ and $\lambda$ is the coefficient for the spatial lagged variable in the two models, respectively. Finally, we compared their coefficients in the normal OLS regression and the spatial regression model to discuss the economic mechanism of spatial dependence of urbanization.

**The neighborhood relationships and matrix**

The first order Queen Neighborhood\(^1\) matrix based on border connection rather than geography distance was used due to the irregular shapes of the prefecture-level cities of China. As the 287 prefecture-level cities did not cover all the administrative areas and all population of China in 2007, some neighborhood relationships were missing according to the above neighborhood matrix criterion, and most of those cities without neighbors were ethnic autonomous areas such as Yanbian Korean Autonomous Prefecture, Yushu Tibetan Autonomous Prefecture.

\(^1\) If two regions are adjacent directly to each other at some point, then the two regions are treated as neighbors.
Meantime, we manually created some neighborhood relationships for the left city islands without any neighbors, if we could make clear that they were very near geographically and could affect each other economically in reality. These neighbors included Haikou with Sanya and Zhanjiang, Xi’ning with Zhangye, Lanzhou and Lhasa, Urumqi with Karamay, Zhourshan with Shanghai and Nantong, Kunming with

Figure 1 the neighborhood relationships of prefecture-level cities in China in 2007

Footnotes: the lines between two cities in green indicated neighborhood relationships created manually. The map was an incomplete form of administrative division map of People’s Republic of China 2007, which should include Kau line in South China Sea (same below). Panzhihua, etc. (see figure 1). The neighborhood relationships (a n by n neighborhood matrix $w_{ij}$) can be either expressed in binary form or row standard form. For the binary form, if two cities i and j were neighbors, then $w_{ij}=1$, otherwise $w_{ij}=0$; for row standard form, if city i had m neighbor cities, then $w_{ij}=1/m$ or 0 correspondingly.

Main findings
1. The spatial distribution of main variables

Figure 2 showed those regions with high level of urbanization were mostly located in Northeast, Bohai Bay, Shandong Peninsula, the Yangtze River Delta, Pearl River Delta and Hetao Plain, which were either the coastal areas or places rich in resources in China (see Figure 2). While cities in Bohai Bay and the Yellow River area were more characterized by traditional industry (such as manufacturing) than their southern counterparts due to their resource advantage such as iron and coal, the coastal cities tended to be more open than inland cities in 2007, which also make sense since they have been enjoying the advantage policies earlier.

Figure 2 the spatial distribution of major variables
2. The global spatial autocorrelation of level of urbanization in China

Both Moran’s I (Table 1) and the Moran Plot (Figure 3) showed spatially positive correlation in all the four variables. The Moran’s I was 0.365 for urbanization proportion of non-agricultural population (\%)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Moran’s I</th>
<th>expectation</th>
<th>variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>proportion of non-agricultural population (%)</td>
<td>0.365***</td>
<td>-0.003</td>
<td>0.002</td>
</tr>
<tr>
<td>proportion of second industry in GDP (%)</td>
<td>0.215***</td>
<td>-0.003</td>
<td>0.002</td>
</tr>
<tr>
<td>percent of the foreign-funded enterprises (%)</td>
<td>0.721***</td>
<td>-0.003</td>
<td>0.002</td>
</tr>
<tr>
<td>annual GDP growth rate (%)</td>
<td>0.247**</td>
<td>-0.003</td>
<td>0.002</td>
</tr>
<tr>
<td>land area in 10 thousand people (km²)</td>
<td>0.221**</td>
<td>-0.003</td>
<td>0.001</td>
</tr>
<tr>
<td>annual per capita meat yield (kg)</td>
<td>0.561***</td>
<td>-0.003</td>
<td>0.002</td>
</tr>
<tr>
<td>number of doctors in hospital per thousand people</td>
<td>0.386***</td>
<td>-0.003</td>
<td>0.002</td>
</tr>
<tr>
<td>number of beds in hospital per thousand people</td>
<td>0.415***</td>
<td>-0.003</td>
<td>0.002</td>
</tr>
<tr>
<td>number of ordinary secondary school students in 10 thousand people</td>
<td>0.508***</td>
<td>-0.003</td>
<td>0.002</td>
</tr>
</tbody>
</table>

***p<0.001.

Footnotes: x axis represents variables y, y axis indicates the spatial lagged variable. While dashed lines stand for the mean levels of the y and lagged y.

level, which indicated significant spatial dependence for urbanization among
prefecture-level cities in China. Spatial autocorrelations were also seen in the degree of openness, degree of industrialization and annual GDP growth rate, with the Moran’s I 0.721, 0.215 and 0.247, respectively. As the spatial dependences in the level of urbanization and the degree of openness were pretty strong, this might suggest the important effect of openness of a city on level of urbanization.

3. The local autocorrelation of level of urbanization in China

The LISA map (see Table 2 and Figure 4) showed the High-High and Low-Low were the major spatial agglomeration types of level of urbanization. The High-High type meant the cities with high level of urbanization tended to have neighbors with high level of urbanization, these hot spots were characterized by advantages in foreign trade, nature resource and national policy, such as cities in the bordering regions with Russia in Heilongjiang province, major resource-rich cities around Shenyang city, part of Yangtze River delta and almost whole Pearl River delta areas, Wuhai –
Shizuishan and Urumqi – Karamay in northwest China. The Low-Low cities meant the cities with low level of urbanization tended to have neighbors with low level of urbanization. These areas were either mountainous with poor communication with neighbors or economically dominated by agriculture, such as central plain area and eastern areas of Sichuan province.

Table 2 Spatial agglomeration types of level of urbanization in 2007

<table>
<thead>
<tr>
<th>types</th>
<th>cities</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-High</td>
<td>Dongguan, Foshan, Guangzhou, Huizhou, Jiangmen, Shenzhen, Zhonshan, Zuhai, Hegang, Heihe, Jiamusi, Jixi, Mudanjiang, Qitaihe, Shangyashan, Yichun, Anshan, Benqi, Fushun, Shenyang, Wuhai, Shizuishan, Suzhou, Shanghai, Karamay, Urumqi</td>
</tr>
<tr>
<td>High-Low</td>
<td>Lanzhou</td>
</tr>
<tr>
<td>Low-Low</td>
<td>Fuyang, Lu’an, Longnan, Pingliang, Qingyang, Baise, Guigang, Laibing, Qinzhou, Yulin, Nanyang, Shangqiu, Xinyang, Zhoukou, Zhumadian, Guyuan, Bazhong, Dazhou, Luzhou, Nanchong, Baoshan, Lingcang, Pu’er, Qujing, Zhaotong</td>
</tr>
<tr>
<td>Low-High</td>
<td>Xuancheng</td>
</tr>
</tbody>
</table>

High-Low and Low-High areas were cities with negative correlation with their neighbors, in our data only Lanzhou and Xuancheng were significant correspondingly. This further ascertained the strong spatial positive correlation. Although the natural condition and social-economic environment might have a quite uneven distribution, the possibility of extreme heterogeneity of level of urbanization was small.

4. The economic mechanism of the level of urbanization

LMerr and LMLag tests as well as their robust forms (RLMerr and RLMlag test) of error from OLS regression are alternatives of Moran test. What’s more, they together can offer a criterion on which spatial model is better. If both LMLag and LMerr tests are significant, and LMLag is more significant than LMerr, meantime, RLMlag is significant and RLMerr is not significant, the spatial lag model is better, and vice versa (Anselin & Florax, 1995). In our study, tests showed spatial error model was appropriate.

The Moran’s I of the OLS model error was 0.2396 while the coefficient of the spatial lagged error term was 0.5286, and both of them were significant, which verified the spatial positive dependence of level of urbanization as the global and
local spatial autocorrelation analyses demonstrated. In the spatial error model, the Moran’s I became insignificant, which meant the model has successfully controlled the spatial autocorrelation of level of urbanization.

The coefficient of the spatial lagged error term was quite large (0.5286), in other words, if a city at prefecture-level in China has neighbors with higher level of urbanization, the city tended to get a higher level of urbanization. This is usually termed as a “spillover effect” or “neighborhood effect” as spatial diffusion theory proposed. However, the “spillover effect” here might not affect each city with equal quantity in reality. As we have seen, the directions of the effect could be negative in some areas, such as Lanzhou and Xuancheng. Three major macro conditions might contribute to this process: 1. natural conditions such as location and resource distribution of a region, for example, the urbanization of some cities may benefit from abundant resource and convenient transportation; 2. the role of the government, which matters through the stimulus of development strategy and regional policy, such as the northeast old industrial base revitalizing plan and the western development strategy; 3. the market economy system itself might also accelerate the diffusion of urbanization in a region differently according to its development level.

Of the three economic independent variables, the effect of annual GDP growth rate was neither significant in the OLS model nor in the spatial model (see Table 3). The possible explanation is that annual GDP growth rate was only an indicator of the economy situation in a single year, while urbanization was a cumulative outcome, the effect of GDP growth rate should be studied in a long period with time serial data.

The coefficient of proportion of second industry in GDP became significantly positive in the spatial error model, which was not significant in the OLS model (see Table 3). Controlling for other confounding factors, one point increase in proportion of second industry in GDP led to 0.2 percent increase in level of urbanization, which was consistent with other findings. Industrialization benefits urbanization through the transformation of industry structure (Chenery et al., 1975). However, the effect of industrialization is China was limited, to some degree, there are three possible reasons for this. Firstly, The China household registration system has hindered the
transformation of the non-agricultural occupation and the conversion of the citizen status from the farmer status, and some scholars even argued that the household registration system could damage the function of push and pull process in population immigration by reshaping the psychological expectations of rural workers in China (Li, 2003). Secondly, when the urbanization of a region climbs to a certain level, the relative influence of industrialization compared to urbanization decreases (the project Table 3 OLS regression and spatial regression model (SEM) of level of urbanization of prefecture-level cities in China in 2007

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>6.5338(6.1117)</td>
<td>13.7618(6.1642)*</td>
</tr>
<tr>
<td>Proportion of second industry in GDP (%)</td>
<td>0.0666(0.6727)</td>
<td>0.1855(0.0599)**</td>
</tr>
<tr>
<td>Percent of the foreign-funded enterprises (%)</td>
<td>0.3588(0.0670)***</td>
<td>0.2708(0.0849)***</td>
</tr>
<tr>
<td>Annual GDP growth rate (%)</td>
<td>0.0823(0.2796)</td>
<td>-0.2202(0.2544)</td>
</tr>
<tr>
<td>Land area in 10 thousand people (km2)</td>
<td>0.0023(0.0049)</td>
<td>0.0040(0.0045)</td>
</tr>
<tr>
<td>Annual per capita meat yield (kg)</td>
<td>-0.0070(0.0186)</td>
<td>-0.0361(0.0205)</td>
</tr>
<tr>
<td>Number of beds in hospital per thousand people</td>
<td>9.7367(1.0141)***</td>
<td>8.2790(1.0134)***</td>
</tr>
<tr>
<td>Number of doctors in hospital per thousand people</td>
<td>-0.1784(1.4250)</td>
<td>2.7339(1.3965)</td>
</tr>
<tr>
<td>Number of ordinary secondary school students in 10 thousand people</td>
<td>-0.0116(0.0056)*</td>
<td>-0.0202(0.0059)***</td>
</tr>
<tr>
<td>Loglik</td>
<td>-1113.414</td>
<td>-1090.659</td>
</tr>
<tr>
<td>AIC</td>
<td>2246.828</td>
<td>2205.3</td>
</tr>
<tr>
<td>BIC</td>
<td>2283.42</td>
<td>2247.159</td>
</tr>
<tr>
<td>Observations</td>
<td>287</td>
<td>287</td>
</tr>
</tbody>
</table>

Note: Figures in parentheses are standard errors. * p <0.05, ** p <0.01, *** p <0.001.

In 2007, the average percent of second industry in GDP of the 287 prefecture-level cities was about 48.9%, which implied the effect of industrialization on urbanization could be limited. Thirdly, after the founding of People’s Republic of China, the strategy of giving high priority to the development of heavy industry had made industry could only absorb modest amount of agricultural population for a long time.

The effect of openness to urbanization was quite large, the coefficient was 0.3588 in the OLS model, and 0.2708 even after controlling for the spatial
autocorrelation in the spatial model. Clearly, one point increase of percent of foreign-funded enterprises could relatively more significantly increase the urbanization level than the degree of industrialization. The degree of the openness was important in several aspects: firstly, it could counteract the inhibition effect of household registration system, for example, many cities made quite flexible polices to attract labor, capital and technology; secondly, the inflow of capital and technology could directly improve the industrialization of a region, and then stimulate urbanization by industrialization.

Meantime, the coefficient of proportion of second industry in total GDP became significant in the spatial model, while the coefficient of percent of foreign-invested enterprises reduced by 25 percent from the OLS model to the spatial model, which implied that the degree of industrialization and the degree of openness might also indirectly affect urbanization through other spatial relevant factors. So both of them not only affected the level of urbanization, but also functioned as an economic mechanism of spatial dependence of the level of urbanization.

**Conclusion and discussion**

- Both the level of urbanization and main economic factors showed significant autocorrelation, of which the proportion of foreign-invented enterprise and the level of urbanization were particularly strong. This might suggest the influence of the degree of openness on level of urbanization.

- LISA analysis showed that, compared to the high-high and low-low spatial agglomerations (positive autocorrelation), the cities belonging to high-low and low-high spatial agglomeration types were relatively isolated and sparse, which verified the strong positive spatial dependence of the level of urbanization.

- The spatial error model further verified the strong spatial dependence of level of urbanization, and successfully controlled the spatial autocorrelation. It was very clear that the level of urbanization of one city was highly dependent on its neighbor cities. The natural environment, market, government policies and development strategy might together contribute to this.

- Both degree of industrialization and openness affected the level of
urbanization significantly. However, one point increase of latter could relatively more significantly increased the urbanization level than the degree of industrialization due to China household registration system, strategy or process of industrialization in China.

- Both degree of industrialization and openness might have an indirect effect on the level of urbanization, through which they played an important role in the spatial dependence of level of urbanization.

This paper tried to introduce spatial perspective and methods to explore the level of urbanization, which should make the estimate of the effect coefficients better and empirically demonstrate the spatial dependence of level of urbanization. Moreover, we analyzed the data in 287 prefecture-level cities in China, which enabled us to control the heterogeneity within the province better. Finally, we showed the economic mechanism of spatial dependence of the level of urbanization in China by comparing the results of the OLS regression and the spatial error model, and compared the marginal effects of degree of industrialization and degree of openness on urbanization.

There were also several limitations in this study. Firstly, non-agricultural rate tended to underestimate the level of urbanization, but this was the only consistent and available indicator we could use then. In terms of the measurements of independent variables, the degree of industrialization here only indicated the share of the GDP in the traditional secondary sector (e.g. mining/manufacture/electricity/construction), while transportation, information transmission, computer sciences, real estate, finance or other industries in the tertiary sector should be also considered as a key indicator of industrialization. Also, some important covariates such as resource and transportation conditions were not available, which might confound the result. Finally, using cross-sectional data rather than panel data might also affect the causal inference.
References


