Analyzing the spatial autocorrelation of regional urban datum land price

JIAO Limin* and LIU Yaolin

School of Resource and Environment Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China; Key Laboratory of Geographic Information System Ministry of Education, Wuhan University, 129 Luoyu Road, Wuhan 430079, China

(Received 1 April 2012; final version received 11 July 2012)

This study focuses on spatial autocorrelation and the spatial distribution of urban land prices from a regional perspective. Taking Hubei province, China, as a case study area, spatial autocorrelation degree, spatial autocorrelation pattern, and the mechanism of its formation were discussed. The study employs Moran’s I, local Moran’s I, and Moran’s I correlogram to analyze spatial autocorrelation degree and its change along with contiguity order. Some local clustering hot spots are found. This paper uses semi-variance statistic for land price based on route distance to find the spatial autocorrelation scale. We also adopt spatial clustering based on a kind of composite distance to probe into the clustering characteristic of land prices. By Moran’s I and Moran’s I correlogram, we find that datum price of the cities in Hubei province has faint spatial autocorrelation degree at the first and the second-order contiguity. Spatial variance hints that the scale of the autocorrelation is about 200 km in route distance. Spatial clustering result indicates that the spatial distribution of city land price is a kind of hierarchy structure similar to administrative regions. From principal factors analysis and stepwise linear regression, we find that the value added of city secondary and tertiary industry and the urban population are two of the most influential factors to urban datum land price. The value added of city secondary and tertiary industry has higher spatial autocorrelation than urban datum land price and has a bigger autocorrelation scale. But urban population has little spatial autocorrelation. It can be inferred that the spatial autocorrelation of urban land price is mainly caused by economic spatial autocorrelation. But its spatial autocorrelation degree is lower than economic factors because urban datum land price is also influenced by other special local factors, such as population, city infrastructure, land supply, etc.

Keywords: spatial autocorrelation; spatial clustering; spatial variation; urban datum land price

1. Introduction

Urban land price is an autocorrelated spatial variable at both city-level or region-level. A lot of research has verified the autocorrelation of urban real estate value at city-level. Researchers have also made efforts to estimate the urban housing price using spatial autocorrelation analysis, spatial statistics, and spatial autoregression. Most of this work, however, focuses on the interior area of a specific city. This paper will discuss inter-city autocorrelation of urban real estate prices, and try to estimate its degree of spatial autocorrelation, autocorrelation scale, and spatial autocorrelation pattern.

Spatial autocorrelation is an assessment of the correlation of a variable in reference to spatial location of the variable (1, 2). Spatial autocorrelation statistics, such as Moran’s I (3), Geary’s C (4), etc. are usually employed to judge whether a variable is spatially autocorrelated. There are several methods can be used to analyze spatial autocorrelation characteristics such as geological statistics (5), local indicators of spatial association (6), a spatial correlogram, spatial clustering, etc. Basu and Tibodeau examined the spatial autocorrelation in transaction prices of single-family properties in Dallas, Texas, and find strong evidence of spatial autocorrelation of housing prices (7). Liu and Li (8) found there is a high spatial autocorrelation for urban land price in Changzhou, a city in southeast of China, using Moran’s I analysis. The technical literature on spatially statistical estimation of real estate prices has been rapidly increasing (9–11). Commercial land price gradients for an emerging real estate market are estimated using spatial regression techniques (12). Many classic land evaluation models are improved and revised based on spatial autocorrelation analysis, for example, spatial hedonic models (13–17). Spatial interpolation of spatially autocorrelated variables, using interpolation methods such as Kriging, has also been introduced into the valuation of real estate (17–19). This research indicates that it is necessary and useful to introduce spatial autocorrelation analysis into the study of price of real estate. Most of this research was conducted in a single city, but what will happen when we examine the real estate price from a regional perspective while taking cities as point samples? This gap in the literature will be addressed in this article. In China, Urban land grading and datum land price appraisal is regulated by the Ministry of Land and Resource and usually implemented by real estate appraisal agencies, companies, or corre-
sponding institutes. It evaluates the urban land grades and average prices for each grade based on a land market survey, this has been done and is routinely updated in most cities and county seats in China. Most previous research on urban datum land prices in China concentrated on the interior of a city or county seat. In recent years, a few studies aiming to regulate urban datum land prices in a region have been conducted in China (8, 20). In these studies, researchers estimate the relationship between urban land prices and urban land classification, and are evaluated based on selected influential factors including economic indices, location, and infrastructure. This research, however, neglects spatial distribution characteristics and the pattern of spatial autocorrelation for urban datum land prices. Probing into the spatial characteristics of urban land prices in a region will contribute to regional land price regulation and knowledge discovery.

The rest of this paper is organized as follows: first, we introduce our study area and data, and introduce our methodology based on Moran’s I, local Moran’s I, and spatial clustering. Then we present our results, including degree of spatial autocorrelation, spatial autocorrelation scale, and spatial clusters of urban land prices in the study area. Finally, we discover influential factors for urban land pricing using principle component analysis (PCA), and discuss the mechanisms of the spatial autocorrelation formation and the spatial pattern of urban land prices by comparing their spatial characteristics.

2. Data and methods
2.1. Case study area and data
The study area is Hubei province located in central China. There are 84 cities and county seats which have finished their urban land grading and datum land price appraisal, see Figure 1. Urban commercial datum land price (UCDLP) is defined as the price of commercial land use of 40 years/m² at 31 December 2004, for average land conditions. The average condition includes water supply, drainage condition, power supply, transportation infrastructure, telecommunication, land flatness, and floor area ratio. The UCDLP of all cities and county seats are regulated according to this definition. We take the high value of UCDLP as the focused spatial variable, the area weighted average of the top two land grade prices in a city.

Furthermore, we also collected the data for possible influential factors in urban land valuation, including city’s location, size, infrastructure, public services, land profit, environment, economic development, and land supply potential, for a total of 8 classes and 67 indices, see Table 1.

2.2. Methodology
2.2.1. Measurement of spatial autocorrelation
We employ Moran’s I (3) to measure the spatial autocorrelation of UCDLP. Cities are represented as points.

\[
I(d) = \frac{n \cdot \sum_{i \neq j} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{(\sum_{i} \sum_{j \neq i} w_{ij}) \cdot (\sum_{i} (y_i - \bar{y})^2)}
\]

where \(I(d)\) is the Moran’s I measure of spatial autocorrelation of UCDLP under \(d\)-order neighborhood, \(d\) is spatial lag and represents the type of neighborhood (\(d = 1\), spatial adjoining; \(d = 2\), 2-order neighborhood that means two spatial units are bridged by another one. And so on). \(y_i\) and \(y_j\) represent the UCDLP of object \(i\) and \(j\), respectively, and \(\bar{y}\) is the average of the UCDLP of all spatial objects. \(n\) is the number of spatial objects. \(w_{ij}\) is the matrix of spatial neighboring weights (if spatial unit

Figure 1. Cities, county seats, and main roads in Hubei province.
and \(j\) are \(d\)‐order neighboring, the weight equals 1, otherwise, 0).

The value of a spatial neighboring weight indicates if a pair of polygons is adjacent or not. Usually the adjacency of a point can be determined using Voronoi diagram, whereas we implement adjacency analysis based on administrative area of the cities because of their social and economic relations with their hinterland.

### 2.2.2. Pattern of spatial autocorrelation

We employ exploratory spatial analysis to illustrate the spatial autocorrelation pattern of UCDLP after estimating its degree of spatial autocorrelation. The methods include spatial correlogram, spatial variability analysis, and spatial clustering.

(1) **Spatial correlogram.** A spatial correlogram is a plot of \(I(d)\) against the number of spatial lags, \(d\) for \(d = 1, 2, ..., n\).

As \(d\) increases, the value of the spatial autocorrelation efficient \(I(d)\) should decrease, since increasing lags diminishes correlation. The correlogram of UCDLP suggests at what lags the spatial autocorrelation is significant, thus we can estimate the spatial autocorrelation scale.

(2) **Spatial variability.** A semi-variogram is often used to reflect spatial variability of regional variables in geospatial statistics. Theoretically, since UCDLP is not a spatially continuous variable, it is not appropriate to employ semi-variogram to describe its spatial variability. We propose an analog called semi-variance to measure the spatial variability of UCDLP.

\[
D(h_i) = \frac{1}{2N(h_i)} \sum_{n=1}^{N(h_i)} (y_n - y_{ni})^2
\]

where \(h_i\) represents the \(i\)th separation distance between ordered data, which equals \(h \ast i\). \(D(h_i)\) is the semi-variance of UCDLP of the paired cities whose distances are between \(h \ast (i - 1)\) and \(h \ast i\). \(N(h_i)\) is the number of the paired cities. \(y_n\) and \(y_{ni}\) represent the UCDLP of the \(n\)th paired cities.

Shortest path distance instead of linear distance is used in the calculation of \(D(h_i)\), since the socio‐economic connections among cities, which influence urban land price, are formed by transportation system. Shortest paths are computed on the network composed by main forms of transportation, including highways, national roads, and province roads. Then we plot the scatter plot of the semi-variance of UCDLP against \(h_i\) and implement semi-variance function fitting if the scatter plot shows a significant trend of spatial variance.

(3) **Spatial clustering.** We employ K‐means clustering method based on a hybrid distance measure composed of spatial distance and attribute similarity to analyze the clustering characteristics of UCDLP. The hybrid distance is defined as
where \( D_{ij} \) is the hybrid distance between city \( i \) and \( j \), \( s_{ij} \) is the shortest path between city \( i \) and \( j \), and \( y_i, y_j \) are UCDLP of city \( i \) and \( j \), respectively.

3. Results

3.1. Moran’s I based spatial autocorrelation analysis

Figure 2 illustrates the Moran scatter plot of UCDLP and Table 2 shows the results of Moran’s I analysis. In Figure 2, the horizontal axis represents the standardized value of UCDLP and the vertical axis is the spatial lag. The four quadrants in the graph provide a classification of four types of spatial autocorrelation: high–high (upper right) and low–low (lower left) indicate positive spatial autocorrelation; high–low (lower right) and low–high (upper left), negative spatial autocorrelation. The slope of the regression line is Moran’s I. There are two regression lines in Figure 2. The lower slope (0.0620) represents the Moran’s I of the total data-set and the higher slope (0.3057) is the spatial autocorrelation index of the subdata-set excluding the outlier, Wuhan, capital of Hubei province, which is marked with a small rectangle in Figure 2. UCDLP of Wuhan city is 11,160 Yuan per square meter, and it is so much higher than the other cities’ with UCDLP between 432 and 2217 Yuan per square meter that the value seems like an outlier in Moran scatter plot. The scatter plot infers that UCDLP in Hubei province has a slight spatial autocorrelation and it becomes higher when the outlier is excluded.

Figure 3 shows the Moran’s I autocorrelogram of UCDLP, built by many Moran’s I scatter plots with different contiguity orders. For total data-set, the Moran’s I almost equals zero when the spatial lag is between 2 and 5, while 3 and 5 for the subdata-set excluding Wuhan. It indicates that UCDLP is slightly autocorrelated at first-order contiguity for the whole data-set, while it is autocorrelated at first- and second-order contiguity when excluding Wuhan.

3.2. Pattern of spatial autocorrelation of UCDLP

3.2.1. Spatial variance of UCDLP

We estimate the spatial variance of UCDLP using semi-variance analysis. The main forms of inter-city transportation in Hubei are express ways, national ways, and provincial ways, see Figure 1. The shortest path between each pair of cities is calculated, and the values are between 10.98 and 823.88 km. whereas the shortest path between contiguous cities is between 10.98 and 206.97 km, and the average is 69.68. Figure 4 shows the scatter plot of shortest path based semi-variance against separation distance, using a separation step of 10 km.

The semi-variance trend cannot be fitted well by ordinary functions, such as the linear function, exponent function, and logarithm functions. The curve in Figure 4 represents the exponent function fitted with a low confidence level. Although we cannot find the appropriate semi-variance function with a high confidence level for the whole data-set, however, we can see from Figure 4 that the semi-variance of UCDLP increases along with the increasing separation distance in general. We infer that the autocorrelation of UCDLP is nonsignificant and descends with the increasing separation distance. The semi-variance becomes random when the separation distance increases.

| Items                        | Moran’s I | Z-score | E(I) | Avg. Moran’s I | St. error of Moran’s I |
|------------------------------|-----------|---------|------|----------------|------------------------|
| The complete data-set        | 0.0620    | 2.1264  | -0.0120 | -0.0137        | 0.0348                 |
| The subdata-set excluding Wuhan | 0.3057    | 4.6477  | -0.0122 | -0.0115        | 0.0684                 |
tance is larger than 200 km. However, according to a Moran’s I analysis, UCDLP is autocorrelated at the first to second contiguity orders, so the autocorrelation scale of UCDLP is between 139.36 and 209.04 km, since the average shortest path distance between contiguous cities is 69.68 km. Hence, we can conclude that the autocorrelation scale of UCDLP is about 200 km due to the high consistency between the semi-variance analysis and Moran’s I analysis.

3.2.2. Spatial clustering of UCDLP

Figure 5 shows the result of spatial clustering based on the hybrid distance measure described in Section 2.2.2. Figure 5 illustrates the clustering distribution of UCDLP. When we examine Figure 5 in relation to Figure 1, we find: (1) The UCDLP of Wuhan, capital of the province, and the capital cities of prefectures are significantly higher than other cities; (2) the UCDLP of the county cities in the same prefecture are similar but less than the capital city of the prefecture; and (3) the UCDLP in Hubei province has a spatial structure similar to the regional administrative hierarchy.

4. Discussion

Theoretically, the spatial autocorrelation of UCDLP is derived from the spatial autocorrelation of related socio-economic factors, such as economic development level, land use profit, fixed assets investment, etc. On the surface, the spatial pattern of UCDLP is similar with the administrative hierarchy. Actually, this is caused by the similarity between the administrative structure and the distribution pattern of dependent factors of UCDLP. But what are the influential factors which cause the spatial pattern of UCDLP? We employed PCA and stepwise regression analysis to identify the most influential factors. We selected 52 factors related to UCDLP, belonging to nine categories: transportation and location of city, scale of economic aggregation, infrastructure, urban public service, profit and investment of land, ecology and environment, regional economy development, regional service ability, and regional potentiality of land provision. Figure 6 shows the scatter plot of all principal components.

The accumulative contribution ratio of the first component is over 30%, far beyond the others. The first component reflects the urban economic development and UCDLP. The most influential factors were selected according to correlation coefficients in the first component: the value added of city secondary and tertiary industry, regional total retail sales of consumer goods, regional gross domestic product, regional total value of social fixed assets per capita, regional percentage of professionals and technical workers, regional local financial revenue, regional social investment of fixed assets, balance of savings deposit per capita, urban population, freight ability, passenger transport ability, regional retail sales of consumer goods per capita, roads area per capita, regional gross domestic product per capita, and regional local financial revenue per capita.

These factors are highly related to UCDLP, but are not independent and some of them are highly correlated.
We conducted a stepwise linear regression analysis to further identify the independent and influential factors for UCDLP. The final result of stepwise regression is shown in Table 1.

There are five factors which are maintained in the stepwise linear regression, the value added of city secondary and tertiary industry, urban population, regional local financial revenue, regional local financial revenue per capita, and roads area per capita. It is obvious that the first two independent variables are most influential factors, here we will analyze the spatial autocorrelation of these two variables and compare with the UCDLP’s to probe into the cause of spatial autocorrelation of UCDLP.

Figure 7 shows the scatter plot of Moran’s I of the value added of city secondary and tertiary industry. A zoom-in graph is shown in a dash rectangle and Wuhan, the outlier, is highlighted by a little rectangle. The Moran’s I, slope of the regression line of the data-set excluding the outlier is 0.5023, and is significant. Figure 8 shows the Moran’s I for urban population and indicates that the variable is not spatially autocorrelated significantly. Figure 9 shows the scatter plot of the semi-variance of the value added of city secondary and tertiary industry with a step of 10 km based on shortest path distance. The semi-variance of the variable has a trend; increasing with the spatial interval, until the spatial interval is equal or more than 400 km and then becomes almost random. In other words, the spatial autocorrelation scale of the variable is about 400 km.

Table 3 shows the comparison of the spatial autocorrelation of three variables, UCDLP, the value added of city secondary and tertiary industry, and urban population. The value of city secondary and tertiary industry is highly correlated with UCDLP. It has a more significant spatial autocorrelation and a larger spatial scale than UCDLP. Although urban population has a high correlation to UCDLP but it is not significantly spatially autocorrelated itself.

Spatial autocorrelation of UCDLP is caused by the spatial autocorrelation of economic development, especially, the value of city secondary and tertiary industries. But UCDLP is also influenced by other factors with non-
significant spatial autocorrelation, such as urban population, urban infrastructure, supply and demand of land, land management, etc. These factors increase the local randomness of UCDLP. It explains why the spatial autocorrelation scale of UCDLP is smaller than the value for city secondary and tertiary industries.

5. Conclusion

In this study, we employ Moran’s I, Moran autocorrelogram, semi-variance analysis, spatial clustering, and statistics methods to study the degree of spatial autocorrelation, the UCDLP spatial autocorrelation pattern and its formation mechanism from a regional perspective, taking Hubei province as a case. The main conclusions are as follows: (1) The results of Moran’s I analysis indicates that UCDLP has a low spatial autocorrelation. (2) A Moran autocorrelogram shows that UCDLP is spatial autocorrelated at the first and the second-order contiguity. Incorporating the semi-variance analysis, we infer that the spatial autocorrelation scale is about 200 km. (3) Spatial clustering analysis of UCDLP shows that it has a hierarchy structure in the spatial domain similar to the administrative hierarchy. (4) The spatial autocorrelation of UCDLP is caused mainly by the autocorrelation of economic development, especially the value added of city secondary and tertiary industries, but UCDLP has a lower degree of spatial autocorrelation and a smaller spatial autocorrelation scale because UCDLP is also influenced by other factors with low or no spatial autocorrelation, such as urban infrastructure, land supply and demand, land management, etc. Moran’s I, spatial variance, and spatial clustering are powerful tools for the analysis of UCDLP or related spatial variables. Urban datum land price is a basic economic element in China, and further study on its spatial characteristics at the more macro scale (the whole country) is necessary.

Acknowledgment

This research was funded by the National Natural Science Foundation of China (Nos. 41171312 and 40901188).

Notes on contributor
Limin Jiao, PhD, is an associate professor in the school of resource and environment science, Wuhan University. He is engaged in research on geo-spatial analysis and spatial data mining.

Yaolin Liu is a professor in the school of resource and environment science, Wuhan University. He is engaged in research on geographic information science.

Table 3. Comparison of the spatial autocorrelation of UCDLP and related factors.

| Variable | Moran’s I | Scale of spatial autocorrelation (km) |
|----------|-----------|--------------------------------------|
| UCDLP    | 0.3057    | About 200                             |
| The value added of city secondary and tertiary industry | 0.5023 | About 400 |
| Urban population | 0.0823 | – |

References

(1) Cliff, A.D.; Ord, J.K. Spatial Autocorrelation; Pion Press: London, 1973.
(2) Cliff, A.D.; Ord, J.K. Spatial Process: Models and Applications; Pion Press: London, 1981.
(3) Moran, P.A.P. The Interpretation of Statistical Maps. J. Royal Stat. Soc., Ser. B. 1948, 10, 243–251.
(4) Geary, R. The Contiguity Ratio and Statistical Mapping. The Incorporated Stat. 1954, 5, 115–145.
(5) Matheoron, G. Principles of Geostatistics. Economic Geol. 1993, 58, 1246–1266.
(6) Anselin, L. Local Indicators of Spatial Association – LISA. Geogr. Anal. 1995, 27, 93–115.
(7) Basu, S.; Thibodeau, T.G. Analysis of Spatial Autocorrelation in House Prices. J. Real Estate Finance Econ. 1998, 17 (1), 61–85.
(8) Liu, Z.G.; Li, M.C. Study on Spatial Autocorrelation of Urban Land Price Distribution in Changzhou City of Jiangsu Province. Chin. Geogr. Sci. 2006, 16 (2), 160–164.
(9) Militino, A.; Ugarte, L.; Garcia-Reinaldos, L. Alternative Models for Describing Spatial Dependence Among Residential Selling Prices. J. Real Estate Finance Econ. 2004, 29 (2), 193–209.
(10) Case, B.; Clap, J.; Dubin, R.; Rodriguez, M. Modeling Spatial and Temporal House Price Patterns: A Comparison of Four Models. J. Real Estate Finance Econ. 2004, 29 (2), 167–191.
(11) Lesage, J.; Pace, K. Models for Spatially Dependent Missing Data. J. Real Estate Finance Econ. 2004, 29 (2), 233–254.
(12) Dale-Johnson, D.; Brzeski, W.J. Spatial Regression Analysis of Commercial Land Price Gradients; Asian Real Estate Society Sixth Annual Conference: Tokyo, 2001.
(13) Kim, C.W.; Phipps, T.; Anselin, L. Measuring the Benefits of Air Quality Improvement: A Spatial Hedonic Approach. J. Environ. Econ. Manage. 2003, 45, 24–39.
(14) Beron, K.J.; Hanson, Y.; Murdoch, J.C.; Thayer, M.A. Hedonic Price Functions and Spatial Dependence: Implications for the Demand for Urban Air Quality. In Advances in Spatial Econometrics: Methodology, Tools and Applications; Anselin, L.; Florax, R.J.; Rey, S.J., Eds.; Springer-Verlag: Berlin, 2004; pp 267–281.
(15) Brasington, D.M.; Hite, D. Demand for Environmental Quality: A Spatial Hedonic Analysis. Regional Sci. Urban Econ. 2005, 35, 57–82.
(16) Anselin, L.; Le Gallo, J. Interpolation of Air Quality Measures in Hedonic House Price Models: Spatial aspects. Spatial Econ. Anal. 2006, 1, 31–52.
(17) Anselin, L.; Lozano-Gracia, N. Errors in Variables and Spatial Effects in Hedonic House Price Models of Ambient Air Quality. Emp. Econ. 2008, 34, 5–34.
(18) Panatier, Y. Variowin: Software for Spatial Data Analysis in 2D; Springer-Verlag: New York, NY, 1996.
(19) Pagourtzi, E.; Assimakopoulou, V.; Hatzichristos, T.; French, N. Real Estate Appraisal: A Review of Valuation Methods. J. Property Invest. Finance. 2003, 21 (4), 383–401.
(20) Wang, Q.G.; Zheng, X.Q. Research of Balancing of Urban Datum Land Price. Sci. Geogr. 2004, 24 (1), 37–41, (in Chinese).