Research Article
Spatial Differentiation and Influencing Factors of Second-Hand Housing Prices: A Case Study of Binhu New District, Hefei City, Anhui Province, China

Song Xu and Zhen Zhang
School of Economics and Management, Anhui Jianzhu University, Hefei, Anhui, China
Correspondence should be addressed to Song Xu; songxuxs@gmail.com

Received 14 July 2021; Revised 31 August 2021; Accepted 11 September 2021; Published 30 September 2021

Academic Editor: Shenggang Li

Copyright © 2021 Song Xu and Zhen Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The multiscale geographic weighted regression (MGWR) model obtains different influence scales of various variables better than the classical geographic weighted regression (GWR) model. This paper studies the price characteristics of second-hand residential transactions in Binhu New District taking advantage of the hedonic price model and MGWR model and draws the following conclusions. (1) There are obvious spatial positive correlation and spatial heterogeneity in the price of second-hand housing in Binhu New District. (2) The number of bedrooms, area, age of the house, and the distance to the nearest school have small effect on the scale, so they have strong spatial heterogeneity. The decoration status and floor are global scale variables, and their spatial heterogeneity is weak. (3) The number of bedrooms, orientation, decoration status, floor, and building structure all positively affect house prices, while area, house age, the distance to the nearest subway station, and the distance to the nearest school negatively affect house prices. Among all factors, the distance to the nearest school is the most important factor affecting house prices, followed by the number of bedrooms and then followed by the distance to the nearest subway station and area, while the orientation, floor, building structure, and decoration conditions have less impact, and the house age has the weakest impact.

1. Introduction
Housing is an important livelihood issue, and housing prices are a common topic of our concern. Since the reform and opening up, the national economy has grown rapidly, and the income of domestic residents has also increased significantly. However, the rate of increase in housing prices is far greater than the rate of growth of the national economy in Hefei city, and there is greater pressure to purchase houses for those in need of housing. Excessive rise in housing prices will lead to some realistic social problems such as imbalances in resource allocation and excessive household income gap and will easily affect the healthy development of the regional economy. Binhu New District is located in the southeast of the main urban area of Hefei City, Anhui Province, China. It has a planned land area of 196 square kilometers and is an important part of Hefei City’s “1331″ urban spatial development plan. In the past ten years, housing prices in Hefei city have increased rapidly, and housing prices in Binhu New District are up from its record highs. The impact of high housing prices involves all aspects of society, and the influencing factors and formation mechanisms of urban residential prices need to be studied urgently.

The hedonic price model can well analyze the heterogeneity of housing prices. Lancaster elaborated on the heterogeneity of hedonic prices based on consumption theory [1], Rosen brought the implicit market into hedonic price analysis and proposed a framework for empirical research [2], and both of them laid the theoretical foundation for the hedonic price model. Freeman theoretically explained the application of the hedonic price model in the residential market [3]. Haizhen and Jia selected 15 factors in Xihu District of Hangzhou as residential characteristics, established a hedonic price model, and analyzed and
explained local residential prices [4]. Liang Zhao established a hedonic price model to explore the commercial housing index in Shenyang and determined the variables and interrelationships that affect commercial housing [5]. According to the hedonic price theory, Zheng and Lia established a hedonic price model of the value of the surrounding residential buildings in Shenzhen subway stations and quantitatively analyzed the impact of rail transit on the price of real estate [6].

Geographically weighted regression (GWR) model is a spatial linear model that adds a spatial weight matrix to the linear regression model. Its purpose is to better show the difference in spatial structure [7]. Tang et al. used the GWR model to explore the impact of different influencing factors on the Shanghai community and its spatial differentiation [8]. Jinjin et al. selected the second-hand housing community in Jinan and established a geographically weighted regression model to analyze the spatial distribution and impact of housing prices factor [9]. But the classic GWR cannot reveal the spatial scale of different influencing factors to some extent. There are still few studies on the scale differences of spatial heterogeneity of different influencing factors. The multiscale geographically weighted regression (MGWR) model was proposed by Fotheringham in 2017 which can subdivide the global scale and local scale of each factor [10]. Yu et al. have supplemented and improved the statistical inference of MGWR, so that the MGWR model can be applied to empirical analysis [11]. Shen et al. discussed the spatial scale and spatial analysis of the influencing factors of the second-hand housing price in Beijing using the MGWR model. The results showed that location is the most important factor affecting housing prices and the influencing scale is the smallest. MGWR regression results are also better than classic GWR [12].

2. Data Source and Variable Selection

2.1. Data Source. This paper selects the unit prices of second-hand houses for sale in Binhu New District of Hefei City in April 2021 as the research object. The research data come from the second-hand website of Lianjia (https://hf.lianjia.com/ershoufang/binhuinxinqu/). After excluding incomplete data and data with incorrect information, there are 507 samples from 116 communities. The coordinate information of the second-hand housing sample comes from the Baidu Map API coordinate picker.

2.2. Variable Selection. The basic expression form of the feature price model is

\[ P = f(Z) = f(S, L, N), \]  

where \( P \) represents the house price, \( Z \) is the relevant attribute feature, \( S \) is the architectural feature of the house, and \( L \) is the location feature of the house. \( N \) is the neighborhood characteristic of the house. In the previous studies, the hedonic price model has linear, logarithmic, semilogarithmic, and other functional forms, and the estimation method mainly adopts the method of ordinary least square (OLS). This article refers to the linear form, and its expression is as follows:

\[ P = a_0 + \sum a_i b_i + \epsilon. \]  

Among them, \( P \) represents the house price, \( a_0 \) represents the constant term, \( a_i \) represents the corresponding regression coefficient, \( b_i \) represents the explanatory variable, and \( \epsilon \) is the error term.

This paper takes the unit price of second-hand housing in Binhu New District as the dependent variable [12]. The independent variables are shown in Table 1. In order to consider the local collinearity of the data, the independent variable variance inflation factor (VIF) and tolerance are used to diagnose the multicollinearity between variables. There is serious multicollinearity after adding tourist attractions, commercial centers, and tertiary hospitals in the model, so we do not study them here.

The explanatory variables selected in this paper are shown in Table 1. The maximum value of VIF coefficient is 2.834 and the corresponding tolerance is 0.353, indicating that there is no serious multicollinearity between explanatory variables.

3. Research Methods

3.1. Spatial Autocorrelation

(1) Global spatial autocorrelation analysis: it mainly analyzes the overall spatial autocorrelation degree of housing price [13]. The global Moran’s index is calculated as follows:

\[ I = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \overline{x})(x_j - \overline{x})}{S^2 \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}}. \]  

Among them, \( S^2 = (1/n) \sum_{i=1}^{n} (x_i - \overline{x})^2 \) is the sample variance, \( n \) is the total number of spatial units, \( x_i \) and \( x_j \) represent the housing price of the \( i \)th spatial unit and the \( j \)th spatial unit, respectively, \( \overline{x} \) is the mean value of the attribute values of all spatial units, and \( w_{ij} \) is the spatial weight. Z-test value is

\[ Z(I) = \frac{I - E(I)}{\sqrt{Var(I)}}. \]  

Among them, \( I \) is the value of Moran’s I index, \( E(I) \) is the expected value of Moran’s I index, and \( Var(I) \) is the variance of Moran’s index.

(2) Local spatial autocorrelation analysis: it describes the degree of similarity between a spatial unit and its domain and explores the heterogeneity of spatial differences [14]. The local Moran’s index is calculated as follows:

\[ I_i = \sum_{i} w_{ij}Z_iZ_j. \]  

Among them, \( w_{ij} \) is the row standardization of the spatial weight matrix \( w_{ij} \), and \( Z_i \) and \( Z_j \) are the standardized
values of second-hand housing prices. Under the significance level of less than 0.05, if \( I_i \) and \( Z_i \) are both positive, it is high-high agglomeration; if \( I_i \) is negative and \( Z_i \) is positive, it is high-low agglomeration; if \( I_i \) is positive and \( Z_i \) is negative, it is low-low agglomeration; when both \( I_i \) and \( Z_i \) are negative values, it is low-high agglomeration [15].

### 3.2. Multiscale Geographically Weighted Regression (MGWR)

The multiscale geographically weighted regression (MGWR) model is one of the frontier methods trying to reveal the multiscale motivations and processes behind various social phenomena of humanistic economy. The main difference from the classical geographically weighted regression (GWR) model is that the MGWR model considers the heterogeneous scale of the difference between the coefficients, calculated as follows:

\[
y_i = \sum_{j=1}^{n} \beta_{buj}(u_i, v_i)x_{ij} + \epsilon_i.
\]

Among them, \( y_i \) is the residential price of the \( i \)-th district, \( \beta_{buj}(u_i, v_i) \) represents the regression coefficient of the independent variable, \( buj \) represents the bandwidth used by the regression coefficient of the \( j \)-th variable, \( x_{ij} \) represents the \( j \)-th independent variable of the \( i \)-th sample point, \( u_i, v_i \) represent the geographic location coordinates of sample \( i \), and \( \epsilon_i \) represents a random error term. Compared with the classical geographically weighted regression model, each regression coefficient of the multiscale geographic weighted regression model is obtained based on local regression. The kernel function and bandwidth selection criterion of the MGWR model use the adaptive Gaussian function and AICc criterion commonly used in classic GWR [12].

The MGWR model determines the difference bandwidth and spatial weight matrix for each independent variable. The GWR model uses the least squares method, and MGWR can be regarded as a generalized additive model (GAM) [16], the MGWR model can also be expressed as follows:

\[
y = \sum_{j=1}^{n} f_j + \epsilon.
\]

Among them, \( f_j = \beta_{buj}x_j \) represents the \( j \)-th additive term and \( \epsilon \) represents the residual term.

Multiscale geographic weighted regression is mainly based on linear additive model and backward fitting method for parameter estimation iteration [10]. Using the classic GWR estimation result as the initial value of MGWR, the difference between the measured value of \( y \) and the initial estimated value is obtained, which is the initial residual \( \bar{\epsilon} \); the formula is as follows:

\[
\bar{\epsilon} = y - \sum_{j=1}^{n} \hat{f}_j.
\]

Perform classical geographic weighted regression on the initial test residual \( \bar{\epsilon} \), the additive term \( \hat{f}_1 \), and the dependent variable \( X_1 \), find the optimal bandwidth \( bu1 \) and update \( \bar{\epsilon} \) and \( \hat{f}_1 \), and then regress and update the additive term \( \hat{f}_2 \) and residual \( \bar{\epsilon} \) of the second variable \( X_2 \), until the last variable \( X_n \) is updated. This process is a complete iteration. Repeat the iteration until the estimate converges to the convergence criterion. Generally, the classic residual square sum change ratio is used as the convergence criterion [17].

### 4. An Empirical Study on the Spatial Differentiation of the Influencing Factors of Second-Hand Housing Prices in Binhu New District

#### 4.1. Analysis of the Spatial Distribution of Second-Hand Housing Prices

It can be seen from Figure 1 that there are obvious differences in the spatial distribution of unit prices of second-hand houses in Binhu New District. The spatial distribution of unit prices is high in the middle, low in the surrounding area, high in the southwest, and low in the northeast; high values are distributed in the central region;
and the unit price is around 40,000. The northeast region is a low-value area with a unit price of around 10,000 yuan.

This paper uses ArcGIS 10.7 software to perform global autocorrelation analysis on 507 sample points. The results are shown in Table 2. The global Moran’s I index is 0.600295, indicating that the unit price of second-hand houses has a strong spatial positive correlation in space. The Z-score is 6.368341, which is greater than the positive critical value of 2.58 at the 1% significance level, indicating that the price of second-hand housing has a significant positive spatial autocorrelation.

In order to further explore the spatial heterogeneity of housing unit prices, judge the local spatial autocorrelation characteristics of housing prices based on local Moran’s I index. The local Moran’s I index is 0.440724, the results show that there is a positive spatial correlation in the region, and it shows a state of high-value clustering or low-value clustering in space. The LISA index can identify the spatial agglomeration of house prices, and the LISA agglomeration is visualized to show the spatial agglomeration of house prices, as shown in Figure 2. It can be seen from the figure that there are significant “high-high” agglomeration areas in the central area and the area around the lake, and “low-low” agglomeration areas appear in the northeast and west regions.

4.2. Model Comparison. The regression results of the multiscale geographic weighted regression model and the classical geographic weighted regression model are compared, as shown in Table 3. Goodness-of-fit $R^2$ of MGWR is 0.905, while goodness-of-fit $R^2$ of classical GWR is 0.888, and the AICc value of MGWR (513.112) is lower than the AICc value of classic GWR (664.471). In terms of the number of effective parameters, MGWR is smaller than classic GWR, indicating that MGWR obtains more accurate regression results than classic GWR using fewer parameters. The residual sum of squares of MGWR is also smaller than that of classic GWR. Therefore, based on the regression analysis results of MGWR and classic GWR in this case, it shows that the MGWR model has better regression results than the classic GWR model.
4.3. Scale Analysis. Since the regression coefficients of the MGWR model are obtained in a different way from the classic GWR model, the bandwidth used by the variable regression coefficients of the MGWR is based on local regression and has a specific type [12]. The different bandwidths of MGWR reflect the scale of action of different variables, while the bandwidth of classic GWR cannot reflect the differences between different variables. The comparison of MGWR bandwidth and classic GWR bandwidth in this case is shown in Table 4. The bandwidth of the classic GWR is 70, accounting for 13.8% of the total sample number. By calculating MGWR, the bandwidth of each variable is different, and it can be seen that the scale of action of different variables is very different. From the perspective of the significance level of the regression coefficient, the regression coefficients for the number of bedrooms, area, orientation, age of the house, distance to the nearest subway station, and distance to the nearest school are overall significant, while the other three variables are not significant. The bandwidth results of MGWR show that the number of bedrooms, area, age, and the distance to the nearest school all have a scale of 44, indicating that the impact of these variables on housing prices varies greatly in space. The function scale of the distance to the nearest subway station is 60, indicating that the spatial heterogeneity of the distance to the nearest subway station is large. The functional scales of building structure and orientation are 126 and 194, respectively, indicating that the spatial heterogeneity of building structure and orientation is general. The decoration status and the function scale of the floor are relatively large, and the spatial heterogeneity is relatively small.

4.4. Analysis of the Spatial Pattern of Coefficients. The statistical description of each coefficient of MGWR is shown in Table 5. The distance to the nearest school negatively affects house prices. The farther the house is from the school, the lower the house price will be, as shown in Figure 3(a). The coefficient of the distance to the nearest school ranges from −1.847 to 0.375, with a mean value of −0.478 and a standard deviation of 0.434. It shows that for every 1 km increase in the distance to the school, the average house price decreases by 478 yuan. The reason for this phenomenon is that people prefer to be close to the school and pay attention to the quality of education. Judging from the absolute value of the regression coefficient, it is the largest among all variables, and its influence is the strongest.

The number of bedrooms reflects the structural characteristics of the house. Generally speaking, the number of bedrooms positively affects the price of second-hand housing, and the size of this influencing factor decreases from the lake to the inland, as shown in Figure 3(b). The coefficient of the number of bedrooms is between −0.016 and 0.674, with an average value of 0.228 and a standard deviation of 0.170. It shows that under a given area, each increase in the number of bedrooms will increase the unit price of the house by an average of 228 yuan. From the perspective of the absolute value of the regression coefficient, its influence is relatively strong.

The distance to the nearest subway station negatively affects the housing price. The closer the distance to the subway station, the higher the housing price. The distance coefficient to the subway station is between −1.049 and 0.324, with a mean value of −0.137 and a standard deviation of 0.282. It shows that for every 1 km increase in the distance to the subway station, the average house price decreases by 137 yuan. Judging from the absolute value of the regression coefficient, its influence is in the middle.

The area factor negatively affects the housing price, indicating that the larger the area, the lower the unit price, which reflects that people have lower demand for houses with larger areas, because the total price of houses with larger areas will be higher, which leads to decrease of the unit price of houses. The area coefficient ranges from −0.690 to 0.304, the average value is −0.124, and the standard deviation is 0.189. It means that for every increase of 1 m² in the area of second-hand residential buildings, the unit price will decrease by 124 yuan on average. Judging from the absolute value of the regression coefficient, its influence is in the middle.

The orientation of the house positively affects the housing price. The better the orientation of the house, the higher the housing price. The degree of its influence decreases from along the lake to the inland, as shown in Figure 3(c). The orientation coefficient of the house is between −0.047 and 1.351, with a mean value of 0.062 and a standard deviation of 0.060. Judging from the absolute value of the regression coefficient, the intensity of its influence is relatively weak.

The floor has a positive effect on the housing price, and the size of the influencing factor gradually decreases from southeast to northwest. As shown in Figure 3(d), the
Figure 3: Continued.
The coefficient range is between 0.039 and 0.070, with a mean value of 0.060 and a standard deviation of 0.010. It shows that the unit price of the upper floors is 60 yuan higher than the average price of the middle floors. The possible reason for this phenomenon is that the closer to the Chaohu area and the higher the floor, the better the scenery and the view. In addition, the high floors have better lighting and ventilation, and people are more inclined to high floors. Moreover, most of the buildings in Binhu New District have elevators to facilitate people to go upstairs and downstairs, so people prefer high-rise floors. Judging from the absolute value of the regression coefficient, its influence is relatively weak.

Building structure positively affects housing prices. The coefficient of the building structure is between −0.027 and 0.209, with a mean value of 0.052 and a standard deviation of 0.051. Judging from the absolute value of the regression coefficient, the intensity of its influence is relatively weak.

The decoration condition also positively affects the housing price, and the size of this influencing factor decreases from south to north, as shown in Figure 3(e). The coefficient of decoration condition ranges from 0.032 to 0.055, with an average value of 0.047 and a standard deviation of 0.008. It shows that the unit price of hardcover houses is 47 yuan higher than simple ones on average. Judging from the absolute value of the regression coefficient, its influence is relatively weak.

House age has a negative influence on the unit price of second-hand housing, and house age reflects the useful life of the house. The size of the influencing factors varies greatly in size and space, as shown in Figure 3(f). The house age coefficient ranges from −0.711 to 2.041, with an average value of −0.041 and a standard deviation of 0.468, indicating that people are more inclined to second-hand houses with lower house age. For every year of house age, the unit price of second-hand houses decreases by an average of 41 yuan. Judging from the absolute value of the regression coefficient, it is the smallest among all variables, and its influence is the weakest. The possible reason for this phenomenon is that most houses in Binhu New District are newly built in recent years.

5. Conclusion

This paper takes the unit price of second-hand housing in Binhu New District as the research object, considers the school district problem on the basis of previous studies, studies the spatial differentiation and spatial scale of the second-hand housing unit price in Binhu New District by using the characteristic price model and MGWR model, and draws the following conclusions. (1) There is obvious spatial positive correlation and spatial heterogeneity in the price of second-hand housing in Binhu New District. (2) Compared with the regression results of classic GWR, the regression results of MGWR are better than classic GWR, and the...
results are better. The main reason is that the multiscale geographic weighted regression model can obtain different influence scales of different variables better than the classical geographic weighted regression model. (3) The number of bedrooms, area, age of the house, and the distance to the nearest school have a small effect on the scale. These factors have strong spatial heterogeneity. The decoration status and floor are variables of global scale, and their spatial heterogeneity is weak. (4) The number of bedrooms, orientation, decoration status, floor, and building structure all positively affect house prices, while area, house age, the distance to the nearest subway station, and the distance to the nearest school negatively affect house prices. Among all these factors, the distance to the nearest key school is the most important factor affecting house prices, followed by the number of bedrooms. The distance to the nearest subway station and area have a greater impact, while the orientation, floor, building structure, and decoration conditions have less impact, and the age of the house has the weakest impact.

Due to the limited data, the influence factors of other location characteristics and neighborhood characteristics are not considered. And the research area is small, the second-hand houses in the main urban area of Hefei City are not researched in this paper. The characteristics of residential transaction prices in old urban areas, as well as the impact of tourist attractions, business centers, and the grade A tertiary hospitals on house prices, need to be further studied.

Data Availability

All data come from the website https://hf.lianjia.com/ershoufang/binhuxinqu/.

Conflicts of Interest

The authors declare that they have no conflicts of interest related to this work.

Acknowledgments

This work was supported by the Key Natural Science Research Projects in Anhui Universities (KJ2019A695), Anhui Jianzhu University Research Startup Project (2019QDZ61), Key Project of Educational Commission of Anhui Province (SK2020A0261), and Anhui Province Philosophy Social Science Project for Youths (AHSKQ2020D67).

Supplementary Materials

Analyze data as attachment: DATA.xlsx. (Supplementary Materials)

References

[1] K. J. Lancaster, “A New approach to consumer theory,” Journal of Political Economy, vol. 74, no. 2, pp. 132–157, 1966.
[2] S. Rosen, “Hedonic prices and implicit markets: product differentiation in pure competition,” Journal of Political Economy, vol. 82, no. 1, pp. 34–55, 1974.
[3] A. M. Freeman, "Hedonic prices, property values and measuring environmental benefits: a survey of the issues," The Scandinavian Journal of Economics, vol. 81, no. 2, pp. 154–173, 1979.
[4] H. Z. Wen and S. H. Jia, “Housing characteristics and hedonic price: analysis based on hedonic price model,” Journal of Zhejiang University (Engineering Science), vol. 38, no. 10, pp. 1338–1342, 2004.
[5] L. Zhao, “Shenyang commercial housing price index based on hedonic price model,” Journal of Northeastern University, vol. 36, no. 8, pp. 1208–1211, 2015.
[6] J. F. Zheng and H. Y. Liu, “The impact of URRT on house prices in shenzhen,” Journal of the China Railway Society, vol. 27, no. 5, pp. 11–18, 2005.
[7] M. W. Wang, Y. Liu, and X. H. Liu, “Spatial pattern and determinant factors of housing prices in polycentric mountainous cities——the case of chongqing city proper,” Journal of Southwest University (Natural Science Edition), vol. 38, no. 5, pp. 133–138, 2016.
[8] Q. Y. Tang, W. Xu, and F. L. Ai, “A GWR- based study on spatial pattern and structural determinants of shanghai’s housing price,” Economic Geography, vol. 32, no. 2, pp. 52–58, 2012.
[9] J. J. Guo, T. S. Xia, and J. C. Li, “Research on the spatial differentiation and influencing factors of urban housing prices," Statistics & Decisions, vol. 8, pp. 142–145, 2018.
[10] A. S. Fotheringham, W. Yang, and W. Kang, “Multiscale geographically weighted regression (MGWR),” Annals of the Association of American Geographers, vol. 107, no. 6, pp. 1247–1265, 2017.
[11] H. Yu, A. S. Fotheringham, Z. Li, T. Oshan, W. Kang, and L. J. Wolf, "Inference in multiscale geographically weighted regression," Geographical Analysis, vol. 52, no. 1, pp. 87–106, 2019.
[12] T. Y. Shen, H. C. Yu, and L. Zhou, “On hedonic price of second-hand houses in beijing based on multi-scale geographically weighted regression: scale law of spatial heterogeneity,” Economic Geography, vol. 40, no. 3, pp. 75–83, 2020.
[13] A. D. Cliff and J. K. Ord, Spatial Processes, Models and Applications, Pion, London, UK, 1981.
[14] L. Anselin, “Local indicators of spatial association: LISA,” Geographical Analysis, vol. 27, no. 2, pp. 93–115, 1995.
[15] S. J. Wang, Y. Wang, X. Q. Lin, and H. O. Zhang, “Spatial differentiation patterns and influencing mechanism of housing prices in China: based on data of 2872 counties,” Acta Geographica Sinica, vol. 71, no. 8, pp. 1329–1342, 2016.
[16] T. J. Hastie, “Generalized additive models,” Statistical Models in S Routledge, vol. 8, pp. 249–307, 2017.
[17] Q. Li, B. Xu, and J. F. Wei, “Urban science spatial-temporal evolution and influencing factors of economic development in yangtze river delta region: data analysis based on 307 county units,” Urban Problems, vol. 12, pp. 14–23, 2020.