Frontiers in Business, Economics and Management
ISSN: 2766-824X | Vol. 14, No. 1, 2024

Research on Real Estate Batch Evaluation Based on Spatial Econometric Model

-- A Case Study of Second-hand Housing Listing Prices in Yuzhong District, Chongqing

Yi Yang, Songmei Huang

School of Economics and Management, Guangxi University of Science and Technology, Liuzhou, Guangxi, 545006, China

Abstract: This article studies the real estate batch evaluation model in Yuzhong District, Chongqing City. By analyzing 350 real second-hand housing listing data in October 2023 as sample data, a feature price model is established and a spatial econometric model is constructed based on the regression results to seek the most suitable model for batch evaluation of second-hand housing in Yuzhong District, Chongqing City. The results show that there is a significant spatial correlation between sample data, and both the spatial error model and spatial lag model in the spatial econometric model have improved the evaluation results of the traditional feature price model. However, the spatial lag model has the best performance in all indicators. Applying the spatial econometric model to real estate batch evaluation can improve the accuracy and fairness of the evaluation results.

Keywords: Real estate batch evaluation; traditional feature price model; spatial error model; spatial lag model.

1. Introduction

Since the concept of "housing is for living, not for speculation" was first proposed at the Central Economic Work Conference in 2016, the development of the real estate market has deeply affected the daily lives of the vast majority of households in China. In recent years, the Chinese real estate market has attracted increasing attention, with measures such as "recognizing houses but not loans", reducing down payment ratios, and lowering mortgage interest rates being successively introduced to optimize the real estate market. With significant changes in the supply-demand relationship in the Chinese real estate market, the demand for real estate valuation has been increasing. In the era of information technology, we need to use information technology to timely analyze market changes. Traditional valuation methods are only suitable for evaluating individual real estate properties [1], while the technology for batch evaluation is gradually maturing and improving. The International Valuation Standards (IVS) defines batch evaluation as the activity of applying systematic, uniform, statistically tested, and results-analyzed evaluation methods and techniques to determine the value of multiple properties on a given date. Real estate batch evaluation can be defined as the process of evaluating a large number of real estate properties at a given time using standardized methods, common data references, statistical testing, and other methods [2].

This paper will consider the spatial factors between real estate properties, introduce the impact of spatial correlation on real estate prices into the evaluation model, and establish both traditional characteristic price models and batch evaluation models based on spatial econometrics. A comparative analysis will be conducted to identify the real estate batch evaluation technology that is closer to the actual transaction price. The mainstream batch evaluation techniques include multiple regression analysis, adaptive valuation techniques, multilayer neural networks, and time series analysis. Scholars both domestically and internationally have utilized various methods for real estate batch evaluation analysis. In recent years, most research on real estate batch evaluation technology in China has adopted traditional characteristic price models, as they believe that real estate prices are influenced by factors such as building characteristics, location features, and neighborhood characteristics [4]. However, traditional characteristic price models often suffer from omitted variables, biased assumptions, leading to estimation errors, or problems such as multicollinearity and mutual influence among various characteristic variables. Therefore, this paper will consider the spatial correlation between real estate properties and introduce spatial econometric models based on optimal characteristic variables to comprehensively analyze the batch evaluation technology that is closest to the actual transaction price.

2. Theoretical Research

1.Traditional Characteristic Price Models

Characteristic price models reflect the relationship between real estate prices and various characteristic factors influencing them. The models currently widely used mainly include: linear function form, semi-logarithmic function form, and logarithmic function form.

(1) Linear function form

$$\mathrm{P} = \alpha_0 + \Sigma_{i=1}^{I} \alpha_i \mathrm{X}_i + \varepsilon \tag{2.1}$$

(2) semi-logarithmic function form

$$\ln P = \alpha_0 + \Sigma_{i=1}^{I} \alpha_i \mathrm{X}_i + \varepsilon \tag{2.2}$$

(3) logarithmic function form.

$$\ln P = \alpha_0 + \Sigma_{i=1}^{I} \alpha_i \ln X_i + \varepsilon \tag{2.3}$$

In the above equation, P represents the housing prices in various real estate samples; α_0 is the constant term of the function; X_i represents the various characteristic variables, i.e., the many factors influencing real estate housing prices; α_i represents the coefficients of the characteristic variables on the housing price P; ε is the random error term [5]. The difference lies in whether the explanatory and dependent variables are in logarithmic form. When the dependent variable is in logarithmic form, the function interprets as: the percentage change in housing prices for a one-unit change in the characteristic variable, scaled by its coefficient; when both the dependent and explanatory variables are in logarithmic form, the function interprets as: each regression coefficient represents the price elasticity of the corresponding characteristic variable, i.e., the percentage change in housing prices for a one percent change in the characteristic variable.

2. Spatial Econometric Models

Spatial econometric methods fully consider the spatial dependence between cross-sectional units, breaking the assumption of mutual independence between samples in standard econometrics. Therefore, when constructing spatial econometric models, traditional regression methods in econometrics cannot be directly used to reveal the spatial effects of sample data, and adjustments need to be made to the analysis methods of traditional models in the actual process.

(1)Spatial Weight Matrix

Before using spatial econometric models for analysis, it is necessary to measure the form of spatial dependence between sample data, which is usually done using a spatial weight matrix to represent the spatial dependence between samples [6].

$$W = \begin{pmatrix} 0 & w_{12} & \cdots & w_{1n} \\ w_{21} & 0 & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & 0 \end{pmatrix} \qquad (2.4)$$

In the spatial weight matrix in the above equation, w_{ij} represents the elements in the matrix, indicating the spatial distance between different samples. The diagonal elements w_{11}, w_{22}, etc., are all 0, indicating that there is no spatial dependence among the samples themselves. The spatial weight matrix mainly includes the neighboring spatial weight matrix and the distance spatial weight matrix. The neighboring spatial weight assigns a value of 1 if samples share a common boundary and 0 otherwise, while the distance spatial weight matrix uses the geographical distance between samples, assigning a value of 0 only when (i=j), otherwise 1. Due to the small area selected in this study, the Euclidean distance is used to define the distance spatial weight matrix.

(2)Spatial Autocorrelation Test

After creating the spatial weight matrix, it is necessary to confirm whether there is spatial autocorrelation among the samples. The commonly used method is to conduct a Moran's I test (Moran, 1950)[7]. When there is spatial autocorrelation among the samples, spatial econometric models can be used. If there is no spatial autocorrelation, only traditional econometric methods can be used for evaluation.

$$\text{Moran's I} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i-\bar{y})(y_j-\bar{y})}{s^2 \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \qquad (2.5)$$

In the calculation formula for Moran's I index in the above equation, y_i represents the sample observation value of region i, y_j represents the sample observation value of region j, $\bar{y} = \frac{\sum y}{n}$, $s^2 = \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n}$ measures the sample variance, and w_{ij} represents the geographical distance between region i and region j in the spatial weight matrix.

The Moran's I index ranges from -1 to 1. Generally, a value greater than 0 indicates positive spatial correlation among the samples, while a value less than 0 indicates negative spatial correlation. If the Moran's I index is closer to 0, the sample observations are more randomly distributed in space, indicating no spatial correlation among them, and spatial econometric methods are not needed for evaluation, only general econometric methods can be used for calculation.

(3)Spatial Econometric Models

Spatial Lag Model

$$y = \lambda Wy + X\beta + \varepsilon \qquad (2.6)$$

(y) represents the dependent variable, (W) is the spatial weight matrix, (Wy) represents the spatial lag of the dependent variable, (λ) denotes the spatial regression coefficient, (X) is the matrix of explanatory variables, (β) represents the influence coefficient of (X) on (y), and (ε) is the random error vector.

Spatial Error Model

$$y = X\beta + \mu \qquad (2.7)$$

In the equation, the random error term $\mu = \rho W u + \varepsilon, \varepsilon \sim N(0, \sigma^2 I_n)$, where (W) is the spatial weight matrix, (ρ) represents the spatial correlation coefficient of (μ), and ($\varepsilon = (I - \rho M)\mu$) is the error term.

Due to potential endogeneity or autocorrelation issues among sample variables, using OLS for regression may lead to estimation errors and other problems. Therefore, spatial econometric models typically employ Maximum Likelihood Estimation (MLE) for regression.

3. Empirical Research

1.Sample Selection

Yuzhong District is located in the central area of Chongqing, as the main city of Chongqing, it is one of the nine main districts of Chongqing. It is located at the confluence of the Yangtze River and Jialing River, with a total area of 23.24 square kilometers, under the jurisdiction of 11 streets and 79 communities. This study selected 350 real second-hand housing listing data from Yuzhong District in Chongqing in October 2023 through real estate information websites such as Anjuke and Fang.com. The sample data are from 78 residential communities, with residential areas ranging from 60m2 to 150m2, ensuring at least meeting the requirements of one-bedroom and one-living room, and considering the construction year limited to communities built after 2000.

2.Selection of Feature Variables and Quantification

Methods

Before establishing the feature price model, it is necessary to comprehensively consider the feature variables included in the model. Referring to the studies of Luo Tingting, Zhang Luanyang, and others, this paper selected a total of 16 feature variables from three aspects: architectural features, location features, and neighborhood features. The explanation, quantification, and expected impact of each feature variable are shown in Table 1.

Table 1. Explanation, Quantification, and Expected Impact of Each Feature Variable

Feature Classification	Variable Names	Quantitative Methods	Expected impact
Housing Prices	PRICE	Average Listing Price of Residential Properties (RMB/m^2)	
Building Features	FAR	Community Plot Ratio	-
	FEE	Community Property Management Fee	+
	GREEN	Community Greening Rate	+
	AREA	Actual Building Area of Houses	+
	AGE	Number of Years from the Year of House Completion to 2023	-
	ORIENTATION	South-North Orientation: 1, Other Orientations: 0	+
	FLOOR	Low Floor: 1; Middle Floor: 2; High Floor: 3	unknown
	ROOM	1 Bedroom: 1; 1 Living Room: 1; 1 Bathroom: 0.5	+
	DECORATION	Rough Finish: 1; Simple Decoration: 2; Medium Decoration: 3; Fine Decoration: 4; Luxury Decoration: 5	+
Neighborhood Features	S_BUS	Number of Bus Routes within 1km	+
	S_METRO	Number of Metro Stations within 2km	+
	D_SUPERMARKT	Distance to the Nearest Large Supermarket	-
	D_COMMERCE	Distance to the Nearest Mall	-
	D_HOSPITAL	Distance to the Nearest Tertiary Hospital	-
	D_SCHOOL	Average Distance to the Nearest Key Primary and Secondary Schools	-

Firstly, descriptive statistical analysis was conducted on the sample data. From Table 2, it can be seen that the average price of second-hand housing in residential areas of Yuzhong District is 17,547 yuan/m2, with a highest price of 49,651 yuan/m2 and a lowest price of 9,489 yuan/m^2, indicating significant price differences in the second-hand residential market in Yuzhong District.

Table 2. Descriptive Statistical Analysis of Each Feature Variable

Variable	Mean	SD	Min	p50	Max
PRICE	17547	5010	9489	17341	49651
FAR	4.245	1.028	2.500	4.300	5.890
FEE	1.745	1.885	0	0.900	8
GREEN	0.279	0.0541	0.193	0.250	0.356
AREA	99.27	19.80	60.21	96.98	149
AGE	6.891	3.765	2	6	23
ORIENTATION	0.549	0.498	0	1	1
FLOOR	2.029	0.808	1	2	3
ROOM	5.250	0.997	2.500	5.500	7
DECORATION	3.794	0.774	1	4	5
S_BUS	30.29	9.551	17	31	50
S_METRO	9.363	1.045	7	9	13
D_SUPERMARKT	0.677	0.379	0.166	0.649	1.325
D_COMMERCE	0.993	0.572	0.0410	0.934	1.826
D_HOSPITAL	1.162	0.457	0.248	1.251	1.822
D_SCHOOL	2.471	1.057	0.515	2.950	3.850

2.Traditional Feature Price Model Establishment

This paper uses Stata software to perform parameter regression estimation on the model using a linear form basic equation, and the estimation results are shown in Table 3.

Table 3. Analysis of Traditional Feature Price Models

Variable	Coef.	Std.Err	t	p>|t|
C	82854.351	6359.843	13.028	0.000
FAR	-5204.281	415.141	-12.536	0.000
FEE	17.105	254.786	.067	0.947
GREEN	385.503	956.074	0.39	0.006
AREA	14.684	11.751	1.250	0.212
AGE	66.691	52.705	1.265	0.207
ORIENTATION	-546.214	239.502	-2.281	0.023
FLOOR	175.932	139.025	1.265	0.207
ROOM	-412.997	234.730	-1.759	0.079
DECORATION	48.886	151.393	.323	0.747
S_BUS	-1043.554	91.571	-11.396	0.000
S_METRO	1951.702	279.614	6.980	0.000
D_SUPERMARKT	-5109.581	812.204	-6.291	0.000
D_COMMERCE	-4182.790	827.872	-5.052	0.000
D_HOSPITAL	-5046.484	714.448	-7.063	0.000
D_SCHOOL	-6646.321	749.700	-8.865	0.000

From Table 3, among the 16 feature variables, six variables including FEE (property fee), AREA (floor area), AGR (construction year limit), FLOOR (floor level), ROOM (number of rooms), DECORATION (decoration level) are not significant. In the subsequent spatial econometric model analysis, the feature price model will be gradually regressed to obtain the optimal regression subset, and finally, the following nine variables will be used for subsequent spatial econometric analysis: FAR (plot ratio), GREEN (greening rate), ORIENTATION (orientation), S_BUS (number of bus stops within 1 km), S_METRO (number of metro stations within 2 km), D_SUPERMARKET (nearest large supermarket distance), D_COMMERCE (nearest shopping mall distance), D_HOSPITAL (medical facilities), D_SCHOOL (educational facilities).

Table 4. Traditional Feature Price Statistical Test Table

Model Testing	Test Indicators	Statistical Value
Goodness of Fit Test	R^2	0.836
	Adjusted R^2	0.829
Overall Significance Test	F	113.64
	Prob>F	0.000
Error Term Independence Test	D-W	1.393

Table 4: Traditional Feature Price Statistical Test Table 4 shows the results of statistical tests on the linear price model using Stata 16.0. It can be observed that the goodness of fit and adjusted goodness of fit of the traditional feature price model are 0.836 and 0.829 respectively, indicating a relatively significant fitting effect of the model, and there is a strong explanatory relationship between the explanatory variables and the explained variable.

(3) Establishment of Spatial Econometric Model

Table 5. Moran's I Table

Variables	I	E(I)	Sd(I)	Z	P-value
price	0.218	-0.005	0.034	8.547	0.000

From Table 5, it can be seen that the Moran's I of the sample data is 0.218, which is greater than 0 and has passed a significance test at the 1% level. This indicates that there is obvious positive spatial dependence among second-hand residential properties in the sample, specifically manifested as high-high or low-low clustering between residential communities. It is necessary to use a spatial econometric model to adjust during the batch evaluation process of second-hand houses to obtain more accurate assessment results.

Table 6. LM Test

TEST	Statistic	df	p-value
Moran's I(error)	0.074	1	0.009
LM(error)	0.008	1	0.930
R-LM(error)	8.618	1	0.003
LM(lag)	1.047	1	0.000
R-LM(lag)	9.657	1	0.002

Table 6 presents the likelihood ratio tests for the spatial error model and spatial lag model. From the test results in the table, it can be observed that the LM(error) statistic is 0.008, which does not pass the significance test at the 10% level, indicating that this test cannot reject the null hypothesis, suggesting the possible presence of spatial autocorrelation among the samples. In the two tests conducted for the spatial lag model, the LM(lag) and R-LM(lag) statistics are 1.047 and 9.657 respectively, both passing the significance test at the 1% level. This indicates that establishing a spatial lag model can better correct the traditional feature price model.

Based on the previous analysis of feature variables, a spatial error model and spatial lag model of housing prices with nine variables were established. Comparing the significance of the two models, all selected feature variables from the basis of the feature price model have a significant impact on housing prices. The spatial error model (SEM) has a spatial error coefficient ρ of 0.3025 with a p-value of 0.008, while the spatial lag model (SLM) has a spatial correlation coefficient λ of 0.107 with a p-value of 0.021. Therefore, the SEM regression result is more significant, which conflicts somewhat with the analysis in Table 6. It shows that solely analyzing the significance of the model may not accurately determine which model has an advantage in evaluating bulk real estate properties. The comparison analysis based on Log-likelihood (LogL) and Akaike Information Criterion (AIC) suggests that the SLM model provides better explanatory power for second-hand housing prices in Yuzhong District, Chongqing.

Table 7. Spatial Econometric Model Statistical Results

Variable	SEM		SLM	
	Coef.	p>|t|	Coef.	p>|t|
FAR	-4547.469 (232.9007)	0.000	-4559.226 (225.350)	0.000
GREEN	7064.433 (4600.378)	0.005	7110.926 (4441.968)	0.009
ORIENTATION	-200.9162 (266.6931)	0.451	-196.84 (265.948)	0.459
S_BUS	-917.7326 (38.94232)	0.000	-913.916 (37.969)	0.000
S_METRO	1071.202 (160.889)	0.000	1064.3 (155.613)	0.000
D_SUPERMARKT	-4113.861 (576.193)	0.000	-4061.809 (560.9337)	0.000
D_COMMERCE	-3593.868 (411.1574)	0.000	-3745.155 (450.367)	0.000
D_HOSPITAL	-5104.613 (816.5963)	0.000	-5022.94 (796.616)	0.000
D_SCHOOL	-6192.626 (293.9904)	0.000	-6248.102 (294.222)	0.000
C	79785.52 (3133.697)	0.000	78101.24 (3844.242)	0.000
ρ	0.3025	0.008		
λ			0.407	0.001
R2	0.922		0.923	
LogL	-1475.868		-1475.623	
AIC	2977.736		2975.245	

According to Table 7, except for ORIENTATION which did not pass the significance test and ORIENTATION and S_BUS variables showing unexpected effects, other variables in the model passed the significance test and aligned with the expected effects. The insignificant impact of the number of bus stops within 1 km on residential housing prices can be attributed to two reasons: firstly, with the development of society and strong government support for the development of new energy vehicles, the increasing number of private cars might lead to families choosing private cars or the subway to avoid losses caused by traffic congestion due to limited time; secondly, most people prefer living in a quiet environment, where an increase in nearby bus stops could potentially affect the quality of life, hence the number of bus stops within 1 km does not influence residential prices.

The empirical results from Table 7 indicate a significant spatial effect on second-hand housing prices in Yuzhong District, implying that the residential housing prices in this area are influenced by neighboring residential community prices. The regression result of the SLM model in Table 7 shows a λ value of 0.407, rejecting the null hypothesis at the 1% significance level, which means that a change of 1 yuan/m^2 in neighboring residential community prices leads to a change of 0.407 yuan/m^2 in the housing price in this area.

4. Ratio Analysis

Table 8. Comparison of Ratio Analysis Results between Traditional Feature Price Models and Spatial Lag Models

	Test Criteria	HPM	SLM
Median Ratio	0.90-1.10	1.165	1.016
Coefficient of Variation	5-20	20.068	18.344
Price Correlation Difference	0.98-1.03	1.012	1.009

Table 8 presents the comparison results of ratio analysis based on the criteria established by the International Association of Assessing Officers. A sample of 10%-20% that was not included in the parameter regression analysis mentioned above was selected for testing, and the results are shown in the table above. The spatial lag model (SLM), considering spatial influencing factors, shows advantages in various aspects compared to the traditional feature price model. Specifically, the median ratio of the SLM model is 1.016, closer to 1 compared to the traditional feature price model, indicating that the results evaluated by the spatial lag model have smaller discrepancies with the actual results. Furthermore, the SLM model's coefficient of variation is 18.344, closer to 15 compared to the traditional feature price model, while the coefficient of variation for the HPM model is 20.068, slightly exceeding the test standard. Lastly, the price correlation difference of the SLM model is 1.009, closer to 1 compared to the traditional feature price model, indicating that its evaluation results have smaller discrepancies with the real results.

4. Conclusion

Through the empirical analysis above, it can be concluded that: first, compared to the traditional feature price model, the spatial econometric model incorporating spatial factors provides more accurate results in mass real estate evaluations, indicating that using spatial econometric methods for real estate bulk assessment can enhance precision and fairness; second, there exists a significant spatial effect between second-hand residential housing prices in Yuzhong District, as evidenced by a Moran's I value of 0.218, indicating clustering phenomena of high-priced communities near high-priced communities and low-priced communities near low-priced communities. Subsequent spatial econometric model analysis and tests further confirmed the significant spatial correlation among residential prices in this region. Moreover, ratio comparison analysis demonstrated that the spatial lag model yields evaluation values closer to the actual values, suggesting superior accuracy compared to the traditional feature price model that does not consider spatial factors, which may lead to significant deviations between the final evaluation results and actual listing prices. Therefore, in the selection of technical methods for mass real estate assessments, the inclusion of spatial factors through spatial lag models offers comprehensive and precise advantages over traditional feature price models.

References

[1] Cui, S. J. (2023). Research on batch evaluation of commercial real estate based on GIS correction. Chongqing University of Technology.

[2] Liu, C. X. (2019). Research on batch evaluation of commercial real estate prices based on internet data. Chongqing University.

[3] Liu, Y. (2021). Application of spatial econometric models in batch evaluation of urban second-hand houses. China Management Informatization, 24(17), 16-17.

[4] Luo, T. T. (2021). Research on the application of spatial econometric models in bulk evaluation of real estate. Yunnan University of Finance and Economics.

[5] Han, J. (2023). Application research of spatial feature price model in batch evaluation of second-hand house values in Nanjing City. Harbin University of Commerce.

[6] Zhang, L. Y. (2019). Research on real estate bulk valuation based on spatial feature method. Jiangxi University of Finance and Economics.