International Journal of Strategic Property Management
ISSN: 1648-715X / eISSN: 1648-9179
2023 Volume 27 Issue 5: 304–316
https://doi.org/10.3846/ijspm.2023.20376

PATH SELECTION OF SPATIAL ECONOMETRIC MODEL FOR MASS APPRAISAL OF REAL ESTATE: EVIDENCE FROM YINCHUAN, CHINA

Yu ZHAO, Xuejia SHEN*, Jian MA, Miao YU

School of Management, Shenyang Jianzhu University, Shenyang, Liaoning, China

Received 19 May 2023; accepted 1 November 2023

Abstract. Urbanization, national economic growth, and China's changing population structure have elevated the importance of real estate assessment in various contexts, including mortgage financing, secondary housing market transactions, and real estate tax reform. To address this need, this study employs a time-spatial double-fixed spatial cross-section data model as a mass appraisal tool to analyze the transaction price data of 429 ordinary residential houses in Xixia District, Yinchuan, China on April 1, 2022. Specifically, this study analyzes 7 spatial cross-section data models, discerning their interconnections. It devises an assignment technique that merges distance and characteristic variable rank into a unified indicator. The results explore spatial lag effects in real estate transaction price generation and assess the descriptive capabilities of different spatial cross-section data models.

Keywords: mass appraisal of real estate, spatial section data model, SDM, empirical analysis.

Online supplementary material: Supporting information for this paper is available as online supplementary material at https://doi.org/10.3846/ijspm.2023.20376

Introduction

The COVID-19 outbreak has affected China's real estate market, leading to reduced new construction and a rise in second-hand housing transactions (Song, 2021). This has posed a challenge to the implementation of the local government's "land finance" model. On October 23, 2021, the "Decision to carry out pilot work on real estate tax reform in some areas" was ratified during the 31st meeting of the Standing Committee of the National People's Congress (NPC). Subsequently, since 2022, numerous cities, Yinchuan City included, have introduced policies aimed at easing property purchase restrictions. In light of the demand for property tax collection and the continuous growth of the real estate market, the accurate assessment of real estate values has gained heightened significance. As a result, there has been a growing need for efficient and scientific real estate valuation in the financial market, tax reform, and the real estate trading market. The mass appraisal of real estate is a key area of current real estate research, with a focus on providing a value reference and pricing basis for buyers and sellers in the secondary housing market. It also serves as a framework for appropriate investment, mortgage, and pricing for developers and financial institutions, and a basis for rational and scientific urban planning by local governments. Additionally, it provides a source for evaluating the real estate tax base. Mass appraisal, as defined by the International Association of Appraisal Officers (IAAO, 2012), has been widely used to evaluate a collection of properties at a certain date using shared data, standardized methodologies, and statistical testing. However, the mass appraisal of real estate faces challenges due to inconsistent criteria for data use and the inadaptability of the setting and assignment of characteristic variables. The absence of a unified database and system of characteristic variables in the hedonic price model and other approaches exacerbates these issues (Chen et al., 2020). Additionally, the geographical characteristics of real estate as a commodity are often ignored by the hedonic price model, which emphasizes the variety of commodities (Feng et al., 2019). Therefore, there are concerns about the accuracy and flexibility of the estimation of house costs using mass appraisal methods. Several studies have utilized the average transaction price of residential units or the transaction price of a single unit as a sample in the empirical examination of real estate mass appraisal. However, the data distribution is often unequal due to the varied transaction activity of different residential areas at a given time, which directly affects the accuracy of regression analysis. In addition, the impact of high-quality

*Corresponding author. E-mail: *shenxuejia6@163.com*

landscape, developer brand effect, high-quality enterprises, and small and medium-sized business districts on real estate prices are often ignored if the selection of characteristic factors is not exhaustive enough and the assignment of variables is too subjective to accurately reflect the utility of variables in a specific region. Therefore, the inadaptability of the setting and assignment of characteristic variables can lead to inconsistent criteria for the use of data in real estate mass appraisal. This study follows the principles of science, objectivity, fairness, and applicability by selecting transaction data at a specific time in the study area. Spatial cross-section data models are then applied to verify the spatial spillover effect in the data generation process, and the most explanatory evaluation model is selected. Specifically, (1) multiple spatial cross-section data models are constructed using actual transaction price, property data, location, and geographic information collected in the study area. (2) An ideal valuation approach is presented based on the regression analysis findings to address the issue of unrepresentative samples caused by transaction activity. (3) The potential of different spatial cross-section data models is explored and validated to evaluate the spatial spillover effect in the data generating process.

Under this background, this study uses the geographical cross-section data model to estimate the cost of second-hand ordinary residential properties in Xixia District of Yinchuan, China, as an example. Several spatial cross-section data models are compared to select the best-fitting model for the region's needs in describing the generation process of second-hand ordinary residential real estate price data. The regression analysis results of the chosen model are then utilized to assess the price distribution of used ordinary residential real estate in the area.

1. Literature review

1.1. Mass appraisal

The Surveyors Club was established in 1792 in response to the growing demand for real estate valuation studies and related activities in Britain following the Industrial Revolution. Real estate appraisal research focused primarily on single appraisal methods such as the cost method, income method, and market comparison method from the 18th century to the beginning of the 20th century. With the increasing use of statistics from the 1920s to the 1970s, the mass appraisal of contemporary real estate began to emerge. Williams (1955) discovered that housing prices are not solely dependent on cost, as construction price and cost fluctuate in different directions with distinct amplitudes. Blettner (1969) proposed that multiple regression analysis can be used to evaluate property values. Kanji (1975) examined the effects of various factors on home prices using a multiple linear regression model. Starting in the 1970s, the characteristic price was gradually included in mass real estate valuation due to advances in computer technology, the development of new consumer theory, and market supply and demand equilibrium theory. Carbone and Longini (1977) proposed the automatic assessment hypothesis and created a feedback model from the taxpayer's perspective. Fletcher et al. (2000) demonstrated that heteroscedasticity in the hedonic model could be eliminated using the generalized least squares method of estimation.

1.2. Econometrics

The use of econometrics in real estate appraisal has evolved significantly in the latter half of the 20th century, with researchers developing new ideas and strategies to address the challenges associated with data gathering and modeling. For example, Nellis and Longbottom (1981) used an error correction estimating method to analyze econometric studies on housing prices and generate global norms, while Pace (1995) examined parametric, semi-parametric, and non-parametric estimates of characteristic values in the context of mass appraisal and hedonic pricing models. Other researchers have explored the use of space technology, such as Dubin et al. (1999), who described how it can increase the precision of market value estimations derived via multiple regression analysis. Researchers have also developed new mathematical tools and algorithms to improve the accuracy of real estate appraisal models. For instance, Gloudemans (2002) discussed the benefits and drawbacks of addition, multiplication, and nonlinear models, while D'amato (2004) established rough set theory to help address the lack of data. Anderson and West (2006) found that there is indeed spatial dependence in the observation sample by analyzing the sample data when modeling HPM. Doszyń (2020) developed a technique to increase the accuracy of results by placing limits on the parameters of the econometric model to guarantee the non-negativity and monotonicity of the impacts of real estate attributes.

The integration of computer technology and artificial intelligence has further advanced real estate appraisal technology, with researchers exploring the use of GIS technology, artificial neural networks, and the random forest model. McCluskey et al. (1997) suggested that GIS technology might increase the effectiveness of mass real estate evaluation, while Antipov and Pokryshevskaya (2012) showed that the random forest model handled "noise" well. Brankovic (2013) examined how market and institutional factors affect real estate value by fusing cadastral and spatial databases with GIS, and Dimopoulos et al. (2018) looked at the random forest model's prediction power in the context of a mass appraisal of apartments in the Cypriot city of Nicosia. Spatial econometric models have also emerged as a promising approach to real estate appraisal, with researchers such as Wilhelmsson (2002) and Osland (2010) demonstrating how they can effectively account for spatial effects in real estate pricing. Yasnitsky et al. (2021) created a complex model with static and dynamic modeling properties to compensate for the fact that neural networks do not adapt to changes in the economic situation or cannot be applied to other regions.

Additionally, supervised regularized regression methods have been used to forecast home prices while accounting for spatial autocorrelation, as shown by McCord et al. (2022). Lo et al. (2022) explored horizontal and vertical spatial autocorrelation among residential properties in Hong Kong through the creation of multiple spatial autoregressive models. Meanwhile, Hermans et al. (2022, 2023) proposed an innovative approach to model development, emphasizing the synthesis rather than mere comparison of various models. Their work extends the mass appraisal model and enhances its overall performance.

In conclusion, the cross-stage development of econometrics and computer technology has contributed significantly to the advancement of real estate appraisal technology, with researchers developing new mathematical tools, algorithms, and support systems to improve the accuracy of models. The incorporation of spatial econometric models and GIS technology has also enhanced the ability to account for spatial effects in real estate pricing, offering new possibilities for mass appraisal of real estate.

1.3. Literature summary

By reviewing the literature on real estate, several shortcomings can be identified in mass evaluation of real estate based on the function model. Firstly, there is a lack of research on the use of spatial cross-section data models for mass real estate evaluation, and the suitability of different models has not been discussed adequately. Secondly, there are issues with the selection and assignment of characteristic variables in residential areas, including missing variables, a single assignment method, and an inability to reflect objective utility. Additionally, there is a lack of research on the validity of the valuation method that combines distance, quantity, and quality in terms of neighborhood features. Lastly, using sample data for average residential area pricing and single transaction price as dependent variables to train the model may result in skewed average transaction prices and an unrepresentative sample set.

To address these issues, this study integrates geographical distance and function, establishes benchmark real estate, and underscores the significance of distinct attributes. Subsequently, the research assesses the precision of various spatial cross-section data models and gauges the appraisal efficacy of the optimal model using universally recognized criteria.

2. Methods

2.1. Spatial cross-section data models

The mass evaluation of real estate must take into account the impact of the economic phenomenon resulting from the heterogeneity and geographical correlation of commodities on the cost or value of real estate. To address this issue, it is necessary to develop spatial econometrics that can study the geographical dependence of cross-sectional sample data. Traditional econometrics can only analyze the heterogeneity of cross-sectional sample data, making it insufficient for this purpose. The general functional form of the spatial section data model is as follows.

$$P=\rho W_1 P + X\beta + \theta W_2 X + \mu; \tag{1}$$

$$\mu=\lambda W_3 \mu + \varepsilon; \tag{2}$$

$$\varepsilon \sim N(0, \sigma^2 I_N), \tag{3}$$

where: P – the vector of dependent variable; ρ – the spatial regression coefficient of dependent variable; W_1, W_2 and W_3 – the spatial weight matrices corresponding to the dependent variable, the independent variable and the disturbance term respectively; X – the independent variable matrix; β – the regression coefficient of independent variable; θ – the spatial regression coefficient of independent variable; μ – a random disturbance vector; λ – the spatial regression coefficient of the disturbance term; ε – a random error vector.

The constraints of the model are $\rho \neq 0$, $\theta \neq 0$, $\lambda \neq 0$.

The functional forms of the spatial cross-section data model can be divided into seven types: Independent variable spatial lag model (SLX), dependent variable spatial lag model (SLM), spatial error model (SEM), spatial Durbin model (SDM), spatial Durbin error model (SDEM), generalized spatial model (SAC) and generalized nested spatial model (GNS) (general form of spatial cross-section data model). The premise assumptions of models are shown as Table 1.

The mutual transformation between the seven spatial cross-section data models and the classical linear regression model can be understood as two levels.

(1) From general to special, as shown in Figure 1 (Elhorst, 2010), that is, adding a single constraint ($\theta = 0$, $\rho = 0$, $\lambda = 0$) or a combined constraint ($\theta = \rho = 0$, $\rho = \lambda = 0$, $\theta = \lambda = 0$) to the generalized nested spatial model.

(2) From special to general (the reverse process in Figure 1), that is, on the basis of the classical linear regression model, the single constraint ($\theta \neq 0$, $\rho \neq 0$, $\lambda \neq 0$) or the combined constraint ($\theta \neq 0$ and $\rho \neq 0$, $\rho \neq 0$ and $\lambda \neq 0$) is released.

Table 1. Functional forms of spatial section data model

Models	Premise assumptions
SLX	Dependent variable has spatial lag effect
SLM	Independent variable has spatial lag effect
SEM	Only the error term has spatial lag effect
SDM	Both independent and dependent variables have spatial lag effect
SDEM	Both independent variable and error term have spatial lag effect
SAC	Both dependent variable and error term have spatial lag effect
GNS	Independent variable, dependent variable and error term all have spatial lag effect

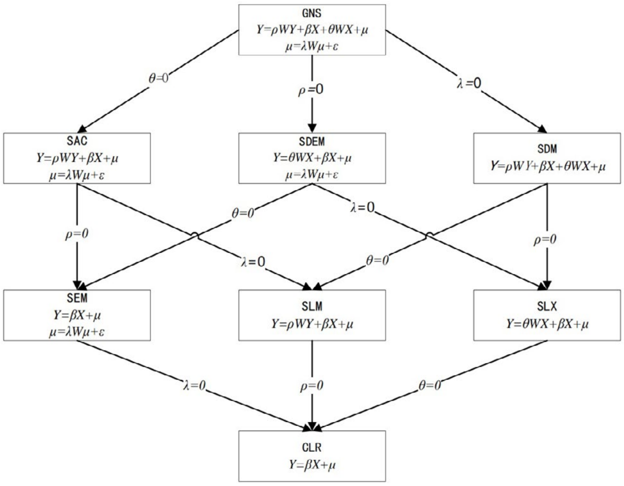

Figure 1. Relationship between spatial section data models

The functional setup of the model is evident from the interaction between the models, as the nesting of spatial lag variables and the superposition of constraints are involved. However, it is important to emphasize that these seven spatial cross-section data models essentially represent seven distinct function models, which results in considerable variations in estimation techniques and a lack of linear superposition in their ability to fit data and describe the data creation process.

2.2. Standard of mass appraisal model of real estate

The method for generating real estate price data is analyzed and restored as part of the construction of the mass evaluation of real estate model. Therefore, the more effectively a model can explain the process of generating the data, the more accurate its evaluation will be. However, in reality, the appraisal outcomes are often skewed due to hidden influences. The IAAO Real Estate Mass Appraisal Criteria provide a ratio test procedure and application standard to identify and address such deviations. Three distinct inspection techniques are available to ensure the accuracy of the appraisal.

2.2.1. Median Ratio (MR)

A number set composed of the ratio of the evaluation value of the test data to the actual value is arranged in ascending or descending order, and the median value is taken as the output result. The formula is as follows.

$$M=\begin{cases} R_{\frac{n+1}{2}}, & n=odds \\ \frac{1}{2}\left\{R_{\frac{n}{2}}+R_{\frac{n+1}{2}}\right\}, & n=even \end{cases}, \quad (4)$$

where: M – median ratio; R – ratio of the evaluated value to the actual value after ranking; n – number of test samples.

According to Standard on Ratio Studies (IAAO, 2013), the acceptable range for MR is 0.90 to 1.10.

2.2.2. Coefficient of Dispersion (COD)

A relative statistic that measures the degree of dispersion of the ratio between the evaluation value and the actual value. The formula is as follows.

$$COD=\frac{\sum\left|AR_i-M\right|}{nM}\times 100, \quad (5)$$

where: AR_i – ratio of the ith sample.

The meanings of other symbols are the same as above.

According to Standard on Ratio Studies (IAAO, 2013), the acceptable range for COD is 5 to 15.

2.2.3. Price Related Difference (PRD)

An indicator to measure the progressiveness or regression of the evaluation value of the test sample. The formula is as follows.

$$PRD=\frac{\sum AR_i/n}{\sum A_i/\sum S_i}, \quad (6)$$

where PRD – price correlation difference; A_i – estimated value of the ith sample; S_i – actual value of the ith sample.

The meanings of other symbols are the same as above.

According to Standard on Ratio Studies (IAAO, 2013), the acceptable range for PRD is 0.98 to 1.03.

2.3. Steps of mass appraisal of ordinary residential real estate

Step 1: Data Collection. Three types of data are required in this step: transaction price data, property data, and geographic information data.

Step 2: Selection and Quantification of Variables. The selection of micro-characteristic factors should consider the characteristics and spatial dependence effects, as well as the actual utility and exogenous nature of variables.

Step 3: Construction of Spatial Weight Matrix. The general expression of the spatial weight matrix is as follows:

$$W=\begin{Bmatrix} 0 & W_{12} & \cdots & W_{1j} \\ W_{21} & 0 & \cdots & W_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ W_{i1} & W_{i2} & \cdots & 0 \end{Bmatrix}, \quad (7)$$

where: 0 indicates that there is no spatial connection between space unit and itself; W_{ij} – the spatial relation between space units and space units, the assignment of W_{ij} depends on the selected setting and the distribution of space units.

Step 4: Spatial Dependency Diagnosis. The global Moran's I test and local Moran's I test are well-researched methods used to diagnose spatial dependency.

Step 5: Setting of Spatial Cross-Section Data Model. Determine the optimal functional form and perform statistical tests.

Step 6: Model Estimation and Testing.

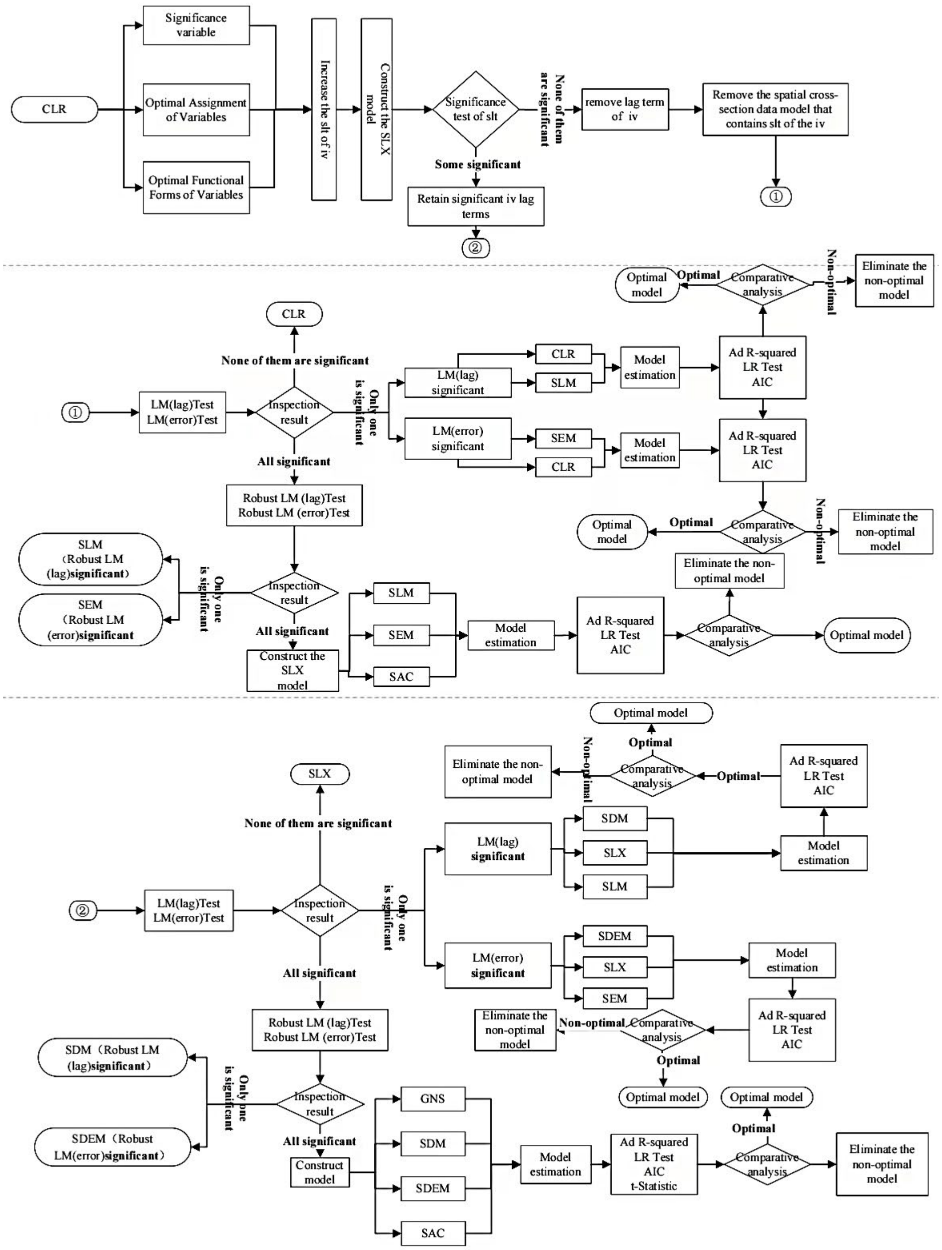

Figure 2. Selection path of spatial section data model (slt = Spatial lag term; iv = independent variable)

2.4. The construction path of the optimal model

Constructing a robust model for mass appraisal of residential real estate necessitates grounded spatial correlation analysis of the research subject. However, existing theoretical research on the spatial lag effect in real estate lacks the necessary foundation to support the presence of spatial lag effects among independent variables, dependent variables, and error terms. Moreover, the scarcity of comprehensive research in this field makes establishing the premise of specific spatial effects challenging, given the limited data validation. Additionally, employing a singular assessment method, such as micro games, along with feedback effects in ordinary residential real estate pricing data generation, makes presumptions of an absence of spatial lag effects untenable. In light of these considerations, this study presents a method to construct an optimal model tailored to specific data, guided by the relationships between models and validated through statistical tests, as depicted in Figure 2.

3. Data selection and quantification standard

3.1. Overview of the appraisal area

Yinchuan, the capital of Ningxia Hui Autonomous Region, has a land area of 9025.38 square kilometers and consists of "three districts, two counties, and one city" (Xingqing District, Jinfeng District, Xixia District, Yongning County, Helan County, Lingwu City). Xixia District, which covers a total area of 1129.3 square meters, is located in the western part of Yinchuan. It is bordered by the Baotou-Lanzhou Railway to the east, the central axis of Helan

Mountain to the west, Yongning County to the south, and Helan County to the north. With a permanent population of about 0.45 million, it is the largest municipal area in Yinchuan City.

3.2. Data collection and preprocessing

From April 1 to April 7, 2022, this study gathered data on transaction prices (in Yuan/m^2) and property attributes of over 2500 ordinary residential real estate properties. Real estate appraisal is dynamic, and for the purposes of this study, the appraisal window is limited to one week, with the resulting appraisal deemed valid for one month. The data were primarily sourced from lianjia.com, with additional information from anjuke.com and 58.com. Initially, more than 2500 data points were collected. Employing a benchmark real estate methodology, property samples were chosen per district, considering property status. After undergoing sample data pre-processing, the study retained a final set of 429 benchmark real estate instances. Sample processing rules are as follows:

(1) Data collected from property listing websites, including listing prices, building characteristics, property details, and location attributes, are organized into sample collections based on residential neighborhoods.

(2) Samples within each collection are filtered to exclude properties listed for over 12 months without any price changes during that period.

(3) Real estate listings marked as "urgent sale" are removed from all collections.

(4) Data points representing significantly high or low prices within each collection are eliminated.

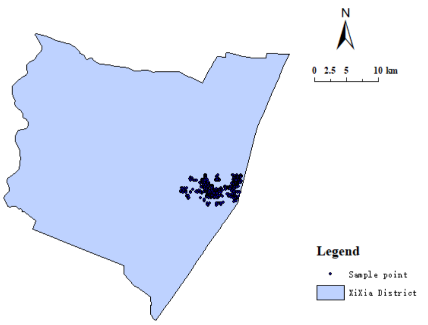

Figure 3. Map of administrative and sample points in Xixia District

(5) Efforts are made to select property samples within each community that encompass various architectural characteristics, focusing on factors like floor levels (high, medium, low) and different levels of interior finishes (luxury, high-end, standard, unfinished), aiming to ensure that the research adequately represents a diverse range of individual properties within residential communities.

To supplement this data, satellite images were obtained from the National Geographic Information Service Platform and imported into ArcMap10.6, where the point element attributes were displayed as longitude and latitude coordinates and outputted as a vector map (as shown on the right in Figure 3).

The distribution of various public facilities in the assessment area plays a crucial role in quantifying neighborhood and location characteristics of the evaluated residential communities. Simultaneously, the arrangement of residential communities influences the construction of the spatial weighting matrix. Hence, a clear understanding of public facility distribution and residential communities forms the foundation for applying the spatial cross-section data model to real estate bulk assessment.

Figure 4 provides an overview of various geographical features in the study area, as follows: The hollow stars signify key locations, with brown, pink and orange representing the Yinchuan Development Zone, Yinchuan CBD, and the Yinchuan WonderVerse Business District, respectively. Red stars indicate the precise positions of elementary schools within Xixia District, while green stars represent parks and amusement gardens. Yellow stars designate areas that host integrated markets, hypermarkets, and influential shopping centers. Blue stars pinpoint the locations of universities and vocational colleges. Purple stars highlight the positions of hospitals. Yellow stars are placed at the 115 residential communities under investigation. Dark blue dots mark the positions of all 115 residential communities included in the study.

3.3. Features and quantification methods

The selection and quantification of characteristic variables of the evaluation sample and their descriptions are shown in Tables 2 and 3.

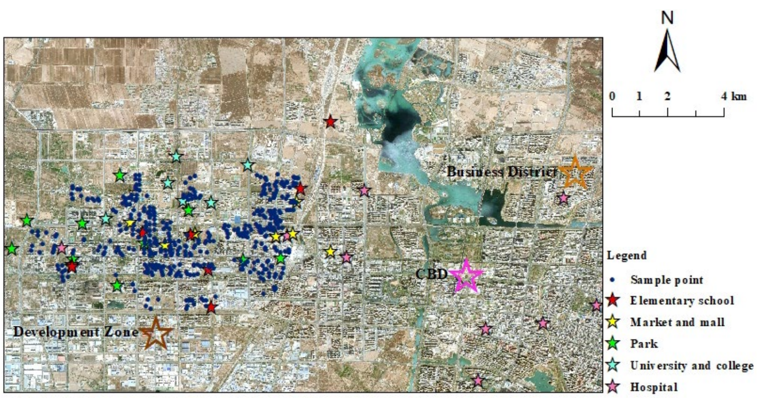

Figure 4. Variable location map

Table 2. Selection and quantification standard of characteristic variables

Feature item	Symbol	Quantization standard	Assignment method	Expected impact
Area	*Ar*	Building area	Actual value	+
Decoration	*D*	Blank 2, simple 4, medium 6, hardcover 8, luxury 10	Virtual assignment	+
Floor	*F*	High-rise residence: low 1, middle 3, high 2	Virtual assignment	unknown
		Low rise house: low 3, medium 2, high 1		
Elevator/househols	*EH*	Number of elevators on the floor/number of single properties on the floor	Actual value	+
Age	*Ag*	Year the house was built – 2022.4.1	Actual value	–
Greening rate	*GR*	Green area/residential area	Actual value	+
Parking space ratio	*PSR*	Total number of households/parking spaces	Actual value	+
Public activity area for residents	*PAA*	Number of areas	Actual value	+
Property fee	*PF*	Monthly charge per square meter	Actual value	unknown
Orientation	*O*	Reasonable = 1, unreasonable = 0	Actual situation	+
CBD	*CBD*	Geographical distance	Anti-geographical distance	+
Shopping mall	*SM*	Service capability: municipal = 3, district level = 2, street level = 1	Service capacity/ geographical distance	+
		Geographical distance		
Main road	*MR*	Geographical distance	Anti-geographical distance	unknown
Bus	*B*	Bus route number	Quantity/distance	+
		The geographical distance to the bus stop		
Medical treatment	*MT*	Qualifications: Grade III Level Special = 10, Grade III Level A = 9, Grade III Level B = 8, Grade III Level C = 7, Grade II Level A = 6, Grade II Level B = 5, Grade II Level C = 4	Qualifications/ geographical distance	+
		Geographical distance		
Education	*E*	Qualifications: high quality = 3, good = 2, ordinary = 1	Qualifications/ geographical distance	+
		Geographical distance		
Developer brand	*DB*	High quality = 3, good = 2, average = 1	Virtual assignment	+
Development zone	*DZ*	Rating: national = 3, provincial = 2, municipal = 1	Rating/geographical distance	+
		Geographical distance		
Park	*P*	Geographical distance	Anti-geographical distance	+

Table 3. The description of variable values

Feature item	Value range	Maximum value	Minimum value	Mean value
Area	40–200	163	44	95
Decoration	2–10	8	2	5
Floor	1–4	4	1	2
Elevator/households	0–1	1	0	0.25
Age	1–35	32	2	15.5
Greening rate	15–50%	45%	15%	28%
Parking space ratio	0–1	0.8	0.1	0.4
Property fee	0–1	1	0.25	0.65
Orientation	0–1	1	0	0.8
CBD	–	0.204	0.0833	1.01
Shopping mall	–	18.75	0.53	2.34
Main road	–	14.29	0.76	3.40
Bus	–	173.91	2.08	30.31
Medical treatment	–	90	2.73	12.05
Education	–	1320	4.57	169
Developer brand	1–6	6	1	3.5
Park	–	20	0.53	2.34

4. Results and discussion

4.1. The spatial dependency diagnosis of the evaluation area

The spatial weight matrix using the K-adjacency matrix is constructed using GeoDa1.20. The number of neighbors is set to 16 based on the sample's actual situation, location similarity, neighborhood characteristics, and practical experience of the market method in single-case assessment.

4.1.1. Global spatial autocorrelation test

The global autocorrelation test was conducted using the single variable global Moran's I test tool in GeoDa1.20, and the results of the significance test obtained through 999 permutations are shown in Figure 5. It can be observed that the global Moran's I index is approximately 0.4858 under the 0.001 significance level test, which provides strong evidence for the need of spatial econometric analysis.

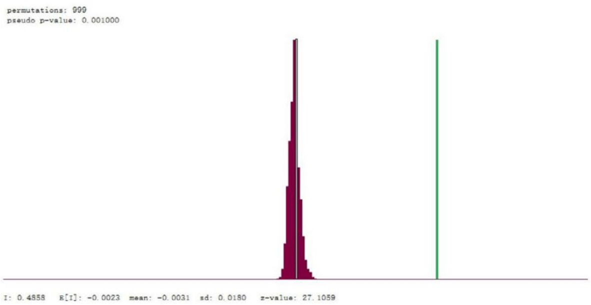

Figure 5. Global autocorrelation test results

4.1.2. Local spatial autocorrelation

The local autocorrelation test was conducted using the univariate local Moran's I test tool in GeoDa1.20, and the resulting Moran's I scatter plot is displayed in Figure 6. The scatter plot reveals the existence of both high-high and low-low spatial adjacency effects within the sample group.

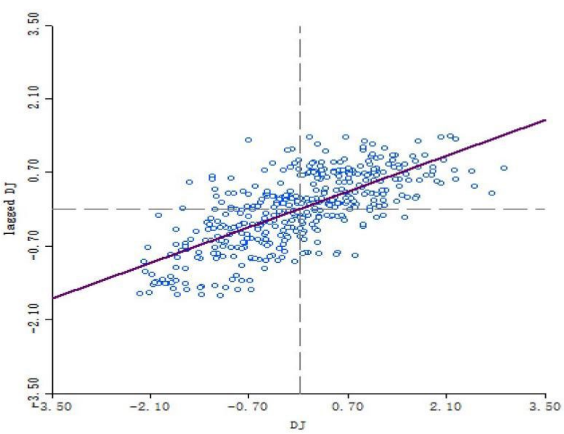

Figure 6. Local Moran's I scatter diagram

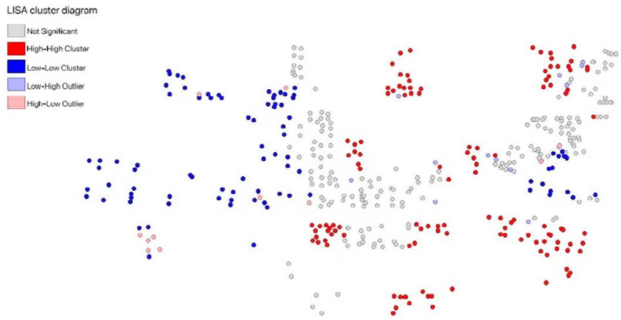

Figure 7. LISA cluster diagram

LISA cluster diagram is shown in Figure 7, and the distribution of four spatial clustering phenomena can be intuitively observed.

4.2. Setting and testing of spatial section data model

Before constructing the spatial cross-section data model, it's imperative to subject the characteristic variables and variable function form to relevant statistical tests for validation. These tests encompass evaluating the goodness-of-fit for linear, semi-logarithmic, and full-logarithmic models, examining multicollinearity and heteroscedasticity, assessing the significance of variables, and computing AIC and log likelihood. Table 4 shows the results of these tests reveal: (1) all three forms pass the F-statistical test, signifying substantial explanatory variable significance; (2) only the full logarithmic model exhibits significant multicollinearity and heteroscedasticity; and (3) the semi-logarithmic form outperforms the linear model in terms of AIC, log likelihood, and adjusted R-squared. Consequently, the optimal selection is the semi-logarithmic function form. Specific variables and their associated assignment methods under this form are meticulously outlined in Table 5.

Table 4. Inspection and selection results

Test	Semi log	Full log	Linear
Log likelihood	375.449	362.16	−3359.46
AIC	−728.898	−702.32	6740.92
Multicollinearity condition number	28.48	96.72	28.48
Adjusted R-squared	0.8263	0.8152	0.8113
Breusch-Pagan test	0.05727	0.00647	0.06454
F-statistic	0.00000	0.00012	0.00000
Koenker-Bassett test	0.09693	0.00207	0.05673
Number of significant variables	10	11	10

Note: The output results of the Breusch-Pagan test, Koenker-Bassett test and F-statistic in the table are the adjoint probabilities of the test results.

Table 5. Significant variables and assignment methods

Variable	t-Statistic	Probability	Assignment method
Ar	11.82180	0.00000	Actual value
D	8.00813	0.00000	Actual value
F	4.94734	0.00000	Virtual assignment
EH	4.64670	0.00000	Actual value
Ag	−15.09490	0.00000	Actual value
CBD	10.10060	0.00000	Anti-geographical distance
SM	5.46333	0.00000	Service capacity/ geographical distance
MR	−4.24358	0.00003	Anti-geographical distance
P	3.17022	0.00164	Anti-geographical distance
B	4.59475	0.00001	Quantity/distance

4.3. Comparative analysis of model estimates

4.3.1. Summary of model estimation and test results

The results of LM test and Robust LM test of residuals based on the regression results of classical linear models are shown in Table 6.

Table 6. LM inspection and robust LM inspection results

Test	Value	Probability
LM (lag)	81.7809	0.0000
LM (error)	138.9328	0.0000
Robust LM (lag)	30.0195	0.0000
Robust LM (error)	87.1714	0.0000

The results of R-squared, Log-likelihood, AIC and LR test of each model are shown in Table 7. In the first test of LR test (LR_1), the constraint models corresponding to SLX, SLM, and SEM is CLR, the constraint models corresponding to SDM, SDEM, SAC, and GNS are SLX, SLX, SLM, and SEM respectively, and the constraint models corresponding to SDM, SDEM, and SAC in the second test (LR_2) are SLM, SEM, and SEM respectively.

The regression results of each model are shown in Table 8.

Table 7. Summary of model test results

Variable	CLR	SLX	SLM	SEM	SDM	SDEM	SAC	GNS
R^2	0.8263	0.8389	0.8728	0.9033	0.9037	0.9041	0.9069	0.8909
Log-likelihood	375.449	386.575	431,617	467.324	473.693	469.429	467.5302	469.869
AIC	−728.898	−745.149	−839.235	−912.648	−913.386	−906.858	−910.164	−902.126
LR_1		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1032
LR_2					0.0002	0.0932	0.0740	

Table 8. Summary of model estimates

Variable	CLR	SLX	SLM	SEM	SDM	SDEM	SAC	GNS
	t(P)	t(P)	t(P)	t(P)	t(P)	t(P)	t(P)	t(P)
Ar	11.82	10.21	12.5	13.06	13.05	12.87	12.91	12.84
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
D	8	7.73	8.63	9.19	8.88	6.76	9.11	5.93
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
F	4.94	5.32	5.35	5.93	5.93	5.42	5.91	5.38
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
EH	4.64	4	4.29	5.45	5.22	5.43	5.42	5.46
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ag	−15.09	−11.75	−12.79	−16.25	−15.74	−15.85	−15.97	−15.87
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
B	4.59	4.87	3.78	2.6	2.99	2.51	2.66	2.29
	(0.000)	(0.000)	(0.000)	(−0.009)	(−0.003)	(−0.012)	(−0.008)	(−0.022)
MR	−4.24	−3.88	−2.97	−2.88	−2.68	−2.97	−2.86	−2.88
	(0.000)	(0.000)	(−0.003)	(−0.004)	(−0.007)	(−0.003)	(−0.004)	(−0.004)
SM	5.46	5	3.58	0.7	1.94	0.56	0.81	0.17
	(0.000)	(0.000)	(0.000)	(−0.485)	(−0.052)	(−0.576)	(−0.417)	(−0.864)

End of Table 8

Variable	CLR	SLX	SLM	SEM	SDM	SDEM	SAC	GNS
	t(P)	t(P)	t(P)	t(P)	t(P)	t(P)	t(P)	t(P)
CBD	10.1	9.43	5.35	3.08	3.06	3.04	3.15	2.35
	(0.000)	(0.000)	(0.000)	(–0.002)	(–0.002)	(–0.002)	(–0.002)	(–0.019)
P	3.17	3.12	3.38	2.9	2.61	2.79	2.93	2.73
	(0.002)	(0.002)	(–0.001)	(–0.004)	(–0.009)	(–0.005)	(–0.003)	(–0.006)
W_Ar		1.57			–3.32	1.48		1.83
		(0.116)			(–0.001)	(–0.14)		(–0.067)
W_EH		–0.18			–2.32	0.47		0.75
		(0.861)			(–0.021)	(–0.639)		(–0.451)
W_Ag		–0.09			6.07	0.94		–0.24
		(0.930)			(0.000)	(–0.350)		(–0.813)
W_F		2.42			–0.52	1.2		1.43
		(0.016)			(–0.606)	(–0.231)		(–0.153)
W_D		3.83			–1.41	–0.66		–0.6
		(0.000)			(–0.159)	(–0.507)		(–0.546)
ρ			10.96		16.22		0.65	–1.13
			(0.000)		(0.000)		(–0.513)	(–0.259)
λ				18.81		17.64	11.56	12.04
				(0.000)		(0.000)	(0.000)	(0.000)

Note: where: *W* – spatial lag term of the corresponding variable; ρ – dependent variable spatial regression coefficient; λ – error term spatial regression coefficient.

4.3.2. Comparative analysis of appraisal models

The construction path selected based on the optimal model of ordinary residential real estate (Figure 2) and the test results (Tables 6 and 7) is shown in Figure 8.

Step 1: Based on Tables 7 and 8, the significance of the spatial lag variable in SLX is observed, indicating that SLX outperforms CLR and cannot be downgraded. Thus, including the spatial lag effect of the dependent variable in the model setting is necessary.

Step 2: Table 6 shows that the dependent variable and the error term have passed the LM test and the Robust LM test, and the spatial lag effect of the error term is more robust than that of the dependent variable (SEM is better than SLM).

Step 3: Table 8 reveals that SDM is superior in terms of R-squared, log-likelihood, and AIC test results, and it cannot be reduced to a single variable model (SLM, SLX). Thus, the model can effectively identify the spatial lag effect of independent and dependent variables. Table 9

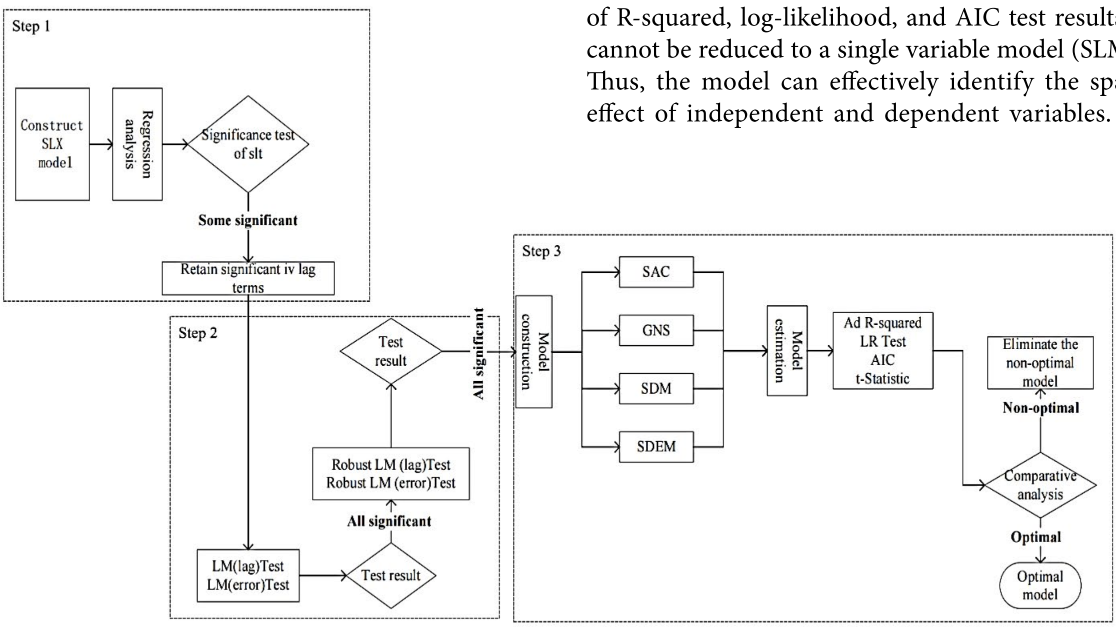

Figure 8. Model building path (slt = Spatial lag term; iv = independent variable)

shows that the two spatial lag variables of SDM are significant, while only the spatial lag of the error term is significant in SAC, SDEM, and GNS. Therefore, if insignificant variables are removed from the model, the three models will degenerate into SEM models. SDM is capable of accurately describing the impact of neighboring real estate prices and characteristics on real estate prices during data generation as there is no significant variable competition between the dependent variable spatial lag term and the independent variable spatial lag term. However, variable competition exists between the error term spatial lag term and the dependent variable and independent variable spatial lag term, resulting in the regression analysis of SAC, SDEM, and GNS lacking spatial lag essence. Additionally, the over-parameterization of GNS significantly reduces its overall analytical ability.

4.3.3. Optimization of model based on mass appraisal of real estate

It is not difficult to see from the comprehensive comparative analysis that the comparison between SDM model and SAC model, SDEM model, and GNS model can be essentially understood as a comparative analysis between SDM model and SEM model.

From the perspective of mass real estate appraisal, the functional model should exhibit the following characteristics: (1) The estimated results of the model should be close to the actual market value; (2) The variable setting should have good interpretability; (3) The model should have practical utility.

While the SEM model's spatial lag term of the error can enhance the model estimation's accuracy, this variable is essentially a "black box." The insignificant influence factors, hidden influence factors, and exogenous errors in the "black box" are difficult to observe, and researchers and evaluators cannot explain the economic significance of this variable. Consequently, the evaluation results of this model will be questioned by both parties, and its guiding significance for practical application will be lost.

In contrast, the estimation results of SDM demonstrate that the regression coefficients of the independent variables, the independent variable spatial regression coefficients, and the dependent variable spatial regression coefficients are significant, and the estimated coefficients of each variable's influence direction are consistent with expectations. Hence, it can better elucidate the economic significance of each variable than SEM. Additionally, even without considering the theoretical significance of model setting, the SDM test results for Log Likelihood, AIC, and R-squared are marginally better than those of SEM.

To conclude, SDM has the best evaluation ability. The optimal functional form for mass real estate appraisal based on ordinary residential real estate data in the study area is Equation (8), and the model estimation coefficients are shown in Table 9.

$$
\begin{aligned}
&Ln(Y) = \rho WY + \beta_1 \cdot Ar \cdot \theta_1 + \beta_2 \cdot D + W \cdot D \cdot \theta_2 + \\
&\beta_3 \cdot F + W \cdot F \cdot \theta_3 + \beta_4 \cdot EH + W \cdot EH \cdot \theta_4 + \\
&\beta_5 \cdot Ag + W \cdot Ag \cdot \theta_5 + \beta_6 \cdot B + W \cdot B \cdot \theta_6 + \\
&\beta_7 \cdot MR + W \cdot MR \cdot \theta_7 + \beta_8 \cdot SM + W \cdot SM \cdot \theta_8 + \\
&\beta_9 \cdot CBD + W \cdot CBD \cdot \theta_9 + \beta_{10} \cdot P + W \cdot P \cdot \theta_{10} + \varepsilon,
\end{aligned}
$$

where: Y – the vector formed by the single trading price of each sample, Yuan/m^2; β_i – linear regression coefficient; θ_i – spatial autoregressive coefficient of independent variable; *Ar*, *D*, *Ag*, *F*, *EH*, *B*, *MR*, *SM*, *CBD*, *P* – vectors composed of corresponding eigenvalues of each sample; W – K adjacency matrix; ε – error term; ρ – spatial regression coefficient of dependent variable.

Table 9. SDM coefficient estimation

Variable	Coefficient	Std. Error	t-Statistic	Probability
CONSTANT	2.4459	0.3477	7.04	0.000
Ar	0.0029	0.0002	13.05	0.000
D	0.0219	0.0025	8.88	0.000
F	0.0285	0.0048	5.93	0.000
EH	0.0990	0.0190	5.22	0.000
Ag	–0.0157	0.0010	–15.74	0.000
B	0.0005	0.0002	2.99	0.003
MR	–0.0037	0.0014	–2.68	0.007
SM	0.0035	0.0018	1.98	0.048
CBD	0.4269	0.1393	3.06	0.002
P	0.0048	0.0018	2.61	0.009
W_Ar	–0.0016	0.0005	–3.32	0.001
W_EH	–0.0758	0.0327	–2.32	0.021
W_Ag	0.0109	0.0018	6.07	0.000
W_F	–0.0058	0.0113	–0.52	0.606
W_D	–0.0080	0.0057	–1.41	0.159
ρ	0.6902	0.0425	16.22	0.000

Table 9 shows that for every 1 m^2 increase in the area of the surrounding properties compared to the current property (*W_Ar*), the house price of the current property decreases by 0.0016 Yuan/m^2. Similarly, for every unit increase in the elevator-to-household ratio of the surrounding properties compared to the current property, the house price decreases by 0.0758 Yuan/m^2. Conversely, for each additional year in the age of the surrounding properties compared to this property, the price of the property increases by 0.0109 Yuan/m^2. Additionally, when the decoration level of the surrounding properties increases by 1 unit compared to this property, the house price of this property decreases by 0.0080 Yuan/m^2. The coefficient of *Ar* is 0.0029, indicating that, in addition to the influence of spatial relations, the housing price increases by 0.0029 Yuan/m^2 for every square meter increase in property area.

The coefficient ρ is 0.6902, meaning a 1 Yuan/m^2 change in surrounding residential housing price corresponds to a 0.6902 Yuan/m^2 change in real estate price.

Furthermore, while the CBD coefficients appear substantial, their comparison with coefficients from other spatial econometric models highlights the SDM's reasonable estimations. This variable has also successfully cleared the multicollinearity test, and the resulting model outcomes align with the IAAO standard ratio. As such, the model's applicability is well-supported and can be confidently acknowledged.

4.4. Appraisal ability measure

To further assess the practical usability of the optimal model, a single real estate transaction price was randomly selected from the 115 residential districts not included in the training data for testing. The model's appraisal performance was then measured against the mass appraisal standard outlined in section 3.2. As depicted in Table 10, the SDM produced three measurement outcomes that align with international standards and exhibit minimal Coefficient of Dispersion (COD) values. Moreover, the Mean Ratio (MR) and Price-Related Differential (PRD) values are closely approximating 1. These outcomes underscore the excellent appraisal capabilities of the SDM within the study area, thus fulfilling international standards.

Table 10. Summary of measurement results

Indicator	MR	COD	PRD
International standard	0.9-1.1	5-15	0.98-1.03
Test result	1.0293	8.4840	1.0097

Conclusions

The study aims to build and compare the explanatory and appraisal abilities of different spatial cross-section data models using a set of data. Based on the empirical analysis and relevant tests, the following conclusions can be drawn:

Firstly, the significant high-high and low-low aggregation phenomena in the flat transaction price of ordinary residential real estate prove the necessity of spatial econometric analysis and highlight the bias in the traditional hedonic price model's assumption of sample independence.

Secondly, the significant spatial regression coefficients of independent and dependent variables in the SLX and SDM regression results explain the influence of neighboring real estate characteristics and prices in generating real estate price data.

Thirdly, the comparative analysis and optimization of the models show that SDM has excellent appraisal ability to describe the spatial spillover effect in the data generation process. The measurement results conform to international standards, indicating that SDM is superior to other spatial cross-section data models in terms of appraisal accuracy and applicability to explain the data generation process.

In conclusion, the spatial Durbin model based on double fixed effect has good applicability in the field of mass appraisal of real estate, and it can effectively capture the spatial dependence in real estate price data. These findings can provide a reference for the practical application of mass appraisal of real estate.

Currently, the model's scope is restricted to general residential real estate, excluding apartments and commercial properties. Nevertheless, its usability is constrained by existing limitations such as non-transparent property transactions and the absence of standardized criteria for property characteristics. The following contents can be further explored in the follow-up research. First of all, with regard to characteristic variables, it is necessary to select characteristic variables according to the actual situation of each kind of city and to standardize the way of assigning values, that is, to establish a standard assignment system for mass appraisal; the second is, in terms of the spatial weighting matrix, when selecting "neighboring" properties, in addition to determining the "neighborhood range" based on the empirical data of house buyers and sellers, a comprehensive survey of the entire study area should be conducted in order to determine a universal "neighborhood range" and to verify the impact of different "neighborhood ranges" on the accuracy of the assessment based on actual data.

Funding

This study was supported by the Humanities and Social Sciences of the Ministry of Education (22YJA630121), Liaoning Revitalization Talents Program (XLYC2203004), Liaoning Federation of Social Sciences (2024lslybkt-043), Social Science Foundation of Liaoning Province (L20BJY010), and Basic Scientific Research Project of Colleges and Universities in Liaoning Province (lnms202140).

Ethical statement

I testify on behalf of all co-authors that our article submitted to the International Journal of Strategic Property Management: This article does not contain any studies with human participants or animals performed by any of the authors; This manuscript is original and has not been published and will not be submitted elsewhere for publication; No data have been fabricated or manipulated (including images) to support the conclusions; All the authors listed have approved the manuscript that is enclosed and they have no conflict of interest.

References

Anderson, S. T., & West, S. E. (2006). Open space, residential property values, and spatial context. *Regional Science and Urban Economics*, *36*(6), 773–789. https://doi.org/10.1016/j.regsciurbeco.2006.03.007

Antipov, E. A., & Pokryshevskaya, E. B. (2012). Mass appraisal of residential apartments: an application of random forest for valuation and a CART-based approach for model diagnostics. *Expert Systems with Applications*, *39*(2), 1772–1778. https://doi.org/10.1016/j.eswa.2011.08.077

Blettner, R. A. (1969). Mass appraisals via multiple regression analysis. *The Appraisal Journal*, *37*(4), 513–521. https://doi.org/10.1111/j.1467-6494.1969.tb01761.x

Brankovic, S. (2013). Real estate mass appraisal in the real estate cadastre and GIS environment. *Geodetski List*, *67*(2), 119–134.

Carbone, R., & Longini, R. L. (1977). A feedback model for automated real estate assessment. *Management Science*, *24*(3), 241–248. https://doi.org/10.1287/mnsc.24.3.241

Chen, L., Wei, Y., & Yang, X. (2020). An integrated machine learning approach for real estate appraisal: a case study of Shanghai, China. *Sustainability*, *12*(4), 1564.

D'amato, M. (2004). A comparison between MRA and rough set theory for mass appraisal. A case in Bari. *International Journal of Strategic Property Management*, *8*(4), 205–217. https://doi.org/10.3846/1648715X.2004.9637518

Dimopoulos, T., Tyralis, H., Bakas, N. P., & Hadjimitsis, D. (2018). Accuracy measurement of random forests and linear regression for mass appraisal models that estimate the prices of residential apartments in Nicosia, Cyprus. *Advances in Geosciences*, *45*, 377–382. https://doi.org/10.5194/adgeo-45-377-2018

Doszyń, M. (2020). Algorithm of real estate mass appraisal with inequality restricted least squares (IRLS) estimation. *Journal of European Real Estate Research*, *13*(2), 161–179. https://doi.org/10.1108/JERER-11-2019-0040

Dubin, R., Pace, R. K., & Thibodeau, T. G. (1999). Spatial autoregression techniques for real estate data. *Journal of Real Estate Literature*, *7*(1), 79–96. https://doi.org/10.1080/10835547.1999.12090079

Elhorst, J. P. (2010). Applied spatial econometrics: raising the bar. *Spatial Economic Analysis*, *5*(1), 9–28. https://doi.org/10.1080/17421770903541772

Feng, Y., Zhang, H., & Liu, Y. (2019). Spatial-temporal autoregressive model for real estate mass appraisal. *Mathematical Problems in Engineering*, *2019*, 4267532. https://doi.org/10.1155/2019/4267532

Fletcher, M., Gallimore, P., & Mangan, J. (2000). Heteroscedasticity in hedonic house price models. *Journal of Property Research*, *17*(2), 93–108. https://doi.org/10.1080/095999100367930

Gloudemans, R. J. (2002). Comparison of three residential regression models: additive, multiplicative, and nonlinear. *Assessment Journal*, *9*(4), 25–36.

Hermans, L. D., Bidanset, P. E., Davis, P. T., & McCord, M. J. (2022). Using property-level ratio studies for the incorporation of validation models in single-family residential real estate assessment. *Journal of Property Tax Assessment & Administration*, *19*(1), 83–102.

Hermans, L. D., McCord, M. J., Davis, P. T., & Bidanset, P. E. (2023). An exploratory approach to composite modelling for real estate assessment and accuracy. *Journal of Property Tax Assessment & Administration*, *20*(1). https://researchexchange.iaao.org/jptaa/vol20/iss1/4

International Association of Assessing Officers. (2012). *Standard on mass appraisal of real property*. https://www.iaao.org/media/standards/StandardOnMassAppraisal.pdf

International Association of Assessing Officers. (2013). *Standard on ratio studies*. https://www.iaao.org/media/standards/Standard_on_Ratio_Studies.pdf

Kanji, G. K. (1975). Analysis of the relative importance of some factors affecting house prices in a local market in 1970. *International Journal of Mathematical Education in Science and Technology*, *6*(3), 277–282. https://doi.org/10.1080/0020739750060303

Lo, D., Chau, K. W., Wong, S. K., McCord, M., & Haran, M. (2022). Factors affecting spatial autocorrelation in residential property prices. *Land*, *11*(6), 931. https://doi.org/10.3390/land11060931

McCluskey, W., Deddis, W., Mannis, A., McBurney, D., & Borst, R. (1997). Interactive application of computer assisted mass appraisal and geographic information systems. *Journal of Property Valuation and Investment*, *15*(5), 448–465. https://doi.org/10.1108/14635789710189227

McCord, M., Lo, D., Davis, P., McCord, J., Hermans, L., & Bidanset, P. (2022). Applying the geostatistical eigenvector spatial filter approach into regularized regression for improving prediction accuracy for mass appraisal. *Applied Sciences*, *12*(20), 10660. https://doi.org/10.3390/app122010660

Nellis, J. G., & Longbottom, J. A. (1981). An empirical analysis of the determination of house prices in the United Kingdom. *Urban Studies*, *18*(1), 9–21. https://doi.org/10.1080/00420988120080021

Osland, L. (2010). An application of spatial econometrics in relation to hedonic house price modeling. *Journal of Real Estate Research*, *32*(3), 289–320. https://doi.org/10.1080/10835547.2010.12091282

Pace, R. K. (1995). Parametric, semiparametric, and nonparametric estimation of characteristic values within mass assessment and hedonic pricing models. *The Journal of Real Estate Finance and Economics*, *11*(3), 195–217. https://doi.org/10.1007/BF01099108

Song, H. (2021). Research on the impact of COVID-19 on China's real estate industry. In *2021 6th International Conference on Social Sciences and Economic Development* (ICSSED 2021) (pp. 401–406). Atlantis Press. https://doi.org/10.2991/assehr.k.210407.078

Wilhelmsson, M. (2002). Spatial models in real estate economics. *Housing, Theory and Society*, *19*(2), 92–101. https://doi.org/10.1080/140360902760385646

Williams, R. M. (1955). The relationship of housing prices and building costs in Los Angeles, 1900–1953. *Journal of the American Statistical Association*, *50*(270), 370–376. https://doi.org/10.1080/01621459.1955.10501271

Yasnitsky, L. N., Yasnitsky, V. L., & Alekseev, A. O. (2021). The complex neural network model for mass appraisal and scenario forecasting of the urban real estate market value that adapts itself to space and time. *Complexity*, *2021*, 5392170. https://doi.org/10.1155/2021/5392170