Journal of Innovation and Development
ISSN: 2958-4752 | Vol. 15, No. 3, 2026

An Empirical Study on the Influencing Factors of Second-Hand Housing Prices Based on the OLS-WLS Model

Shuohan Jiang, Linjie Chang, Guangchun Lu

School of Economics and Management, North China University of Science and Technology, Tangshan, Hebei, China

Abstract: This study investigates the second-hand housing market in Beijing, focusing on building area and school district status as the core variables, while controlling for housing features such as renovation level, heating type, and elevator availability. The effects of these factors on housing prices are quantified by using a multiple linear regression model. After conducting normality and heteroscedasticity tests on the residuals, and applying weighted least squares (WLS) correction, the results indicate that building area and school district status have significant positive impacts on housing prices. Elevator availability increases prices, centralized heating has a negative effect, and high-end renovation is not statistically significant. The adjusted R^2 of the model is 0.687, indicating a good overall fit. The findings of this study provide practical guidance for buyers and offer a reference for government housing market regulation.

Keywords: Second-Hand Housing Prices; Multiple Linear Regression; WLS Correction; Influencing Factors; Beijing.

1. Introduction

With the continuous rise in commercial housing prices in China and the limited purchasing power of residents [1], public attention has increasingly shifted toward the secondary housing market. Transactions of second-hand houses have become an increasingly important segment of the real estate market. The prices of second-hand housing not only influence buyers' purchasing intentions and residential rights, but also hold significant implications for government housing market regulation and for residents making rational housing decisions [2].

The prices of second-hand housing are influenced by a variety of factors. In recent years, with the continuous development of the real estate market, second-hand housing prices have been affected by location, housing attributes, supporting facilities, geographic factors, policies related to talent settlement, and the overall level of urban development [3][4]. Previous studies have employed multiple linear regression and hedonic pricing models to reveal the mechanisms through which factors such as building area, orientation, and floor level impact second-hand housing prices [5][6].

In recent years, machine learning methods have been widely applied in housing price prediction, with models such as Random Forests, XGBoost, and neural networks demonstrating notable improvements in predictive accuracy [7][8]. While these models are capable of capturing nonlinear relationships in large datasets, they offer limited interpretability in economic terms, making it difficult to analyze the marginal effects of core variables. Therefore, traditional regression analysis remains indispensable for elucidating the mechanisms driving housing prices and quantifying the key influencing factors.

Based on the above context, this study focuses on the second-hand housing market in Beijing, selecting building area and school district status as the core explanatory variables, while incorporating housing features such as renovation level, heating type, and elevator availability as control variables. A multiple linear regression model [9] is constructed to analyze the effects of these features on the listed total price of second-hand houses and to quantify their impact. To ensure the robustness of the conclusions, normality and heteroscedasticity tests of the residuals were conducted, and weighted least squares (WLS) correction was applied. The findings of this study provide rational guidance for homebuyers and offer a reference for the government to optimize housing market regulation and housing security policies.

2. Data and Variables

2.1. Data Sources

The data were obtained from www.macrodatas.cn, and include transaction prices of second-hand houses in Beijing. The sample period spans from January 2020 to June 2024, covering the period following the outbreak of COVID-19 in 2020. The sample size is substantial, and the distribution of housing prices is stable.

The original dataset was relatively large. To ensure the representativeness of the analysis, the data were first grouped by year, and records with missing core information were excluded. Approximately 1,000 observations were then randomly selected from each year, resulting in a total of 5,006 valid samples for the empirical analysis. Meanwhile, categorical variables such as school district status, renovation level, heating type, and elevator availability were converted into dummy variables, and the statistical definitions and formats were standardized to maintain data consistency and comparability.

2.2. Variables

2.2.1. Dependent Variable

In this study, the listed total price of second-hand houses (Price) is selected as the dependent variable in the model. Due to the limited availability of actual transaction prices and the presence of substantial missing data, and given that the trends of listed prices and final transaction prices are largely

consistent with a strong correlation and substitutability, the listed price is used as a proxy for the final transaction price in this analysis.

2.2.2. Core Independent Variable

This study selects building area (Area) and school district status (School_District) as the core explanatory variables. As we all know, there is a significant disparity in educational resources across different administrative districts in Beijing. The education quality in Haidian, Xicheng, Dongcheng, and Chaoyang districts is markedly higher than in other districts, resulting in a prominent school district premium. Accordingly, this study classifies school district quality based on the property's administrative district to measure the impact of high-quality educational resources on second-hand housing prices.

2.2.3. Control Variables

To eliminate the potential interference of intrinsic residential attributes on the regression results and to avoid bias in estimating the core variables, this study includes renovation level (Renovation), heating type (Heating), and elevator availability (Elevator) as control variables, as detailed in Table 1.

High-end renovation affects the living quality of a property, while centralized heating and elevator availability are closely related to residential comfort and can also influence the listed price of second-hand houses to a certain extent. By controlling for these variables, the effects of building area and school district status on housing prices can be more accurately identified.

Table 1. Variable Definitions and Coding

Variable Type	Variable Name	Variable Symbol	Description
Dependent Variable	Total Listed Price	*Price*	Total listed price of the property in the market (10,000 CNY)
Core Explanatory Variable	Building Area	*Area*	Registered building area of the property (m²)
Core Explanatory Variable	School District Status	*School_District*	Dummy variable: 1 = Haidian/Xicheng/Dongcheng/Chaoyang; 0 = other districts
Control Variable	Renovation Level	*Renovation*	Dummy variable: 1 = High-end renovation; 0 = Simple/Unfinished
Control Variable	Heating Type	*Heating*	Dummy variable: 1 = Centralized heating; 0 = Individual heating
Control Variable	Elevator Availability	*Elevator*	Dummy variable: 1 = Equipped with elevator; 0 = No elevator

3. Methodology and Model

3.1. Model Construction

This study employs the ordinary least squares (OLS) regression model to analyze second-hand housing prices [10]. The objective is to quantify the effects of core factors, such as building area and school district status, on housing prices.

The regression model is specified as follows:

$$Price_i = \beta_0 + \beta_1 Area_i + \beta_2 School_i + \beta_3 Renovation_i + \beta_4 Heating_i + \beta_5 Elevator_i + \varepsilon_i$$

Where: $Price_i$ is the dependent variable, representing the listed total price of the second-hand house (10,000 CNY), reflecting the market price level. $Area_i$ is the core explanatory variable for building area (m²), capturing the effect of property size on price. $School_i$ is the core explanatory dummy variable for school district status, where "1" indicates properties located in Haidian, Dongcheng, Xicheng, or Chaoyang districts, and "0" indicates non-school district properties, used to quantify the school district premium. $Renovation_i$、$Heating_i$ and $Elevator_i$ are control variables, indicating whether the property has high-end renovation, centralized heating, and elevator availability, respectively, to capture the influence of internal housing facilities on price. β_0、β_1、β_2、β_3、β_4 and β_5 denotes the regression coefficient of each variable. ε_i is the error term, capturing other unexplained factors.

3.2. Model Testing Methods

To ensure the validity and reliability of the regression results, this study conducts tests on the key assumptions of the OLS regression model after estimation. The main aspects examined include the following.

3.2.1. Residual Normality Test

Residual normality is a key assumption of OLS regression. In this study, the distribution of residuals was examined by plotting a histogram. If the histogram exhibits an approximately bell-shaped distribution, the residuals are considered to follow a normal distribution, thereby satisfying the model assumption.

3.2.2. Heteroscedasticity Test

Heteroscedasticity can lead to biased standard errors in OLS estimation, affecting the assessment of statistical significance. In this study, heteroscedasticity was examined using a residual scatter plot [11], with standardized predicted values on the horizontal axis and standardized residuals on the vertical axis. If the residuals are randomly distributed around zero without any systematic pattern, no significant

heteroscedasticity is assumed. In cases of mild heteroscedasticity, the model is corrected using weighted least squares (WLS) to ensure robustness.

4. Empirical Analysis

4.1. Descriptive Statistics of Core Variables

A total of 5006 second-hand housing samples from Beijing were included in this study. To clearly understand the characteristics of each variable and the distribution of the sample, descriptive statistical analyses were first conducted for the dependent variable, core explanatory variables, and control variables. The results are presented in Table 2 and Table 3.

Table 2. Descriptive Statistics of Continuous Sample Variables

Variable	Mean	Max	Min	Standard Deviation
Total House Price (10,000 CNY)	497.92	2950.00	52.00	278.62
Building Area (m²)	79.94	395.00	22.00	30.47

The sample of total house prices has a mean of 4.9792 million CNY, a minimum of 0.52 million CNY, a maximum of 29.50 million CNY, and a standard deviation of 2.7862 million CNY, indicating substantial price dispersion, which is consistent with the pronounced heterogeneity of the Beijing second-hand housing market. The mean building area is 79.94 m², ranging from 22 to 395 m², covering properties from small starter units to large apartments.

Table 3. Sample Distribution of Housing Attribute Dummy Variables

Variable	Category	Sample Size	Percentage (%)
School District Status	High-Quality District	2,589	51.7
	Regular District	2,417	48.3
Renovation Level	High-End Renovation	2,714	54.2
	Simple / Unfinished	2,292	45.8
Heating Type	Centralized Heating	4,267	85.2
	Individual Heating	739	14.8
Elevator Availability	Equipped with Elevator	3,110	62.21
	No Elevator	1,896	37.79

Among the sample, 54.2% of the properties are high-end renovated, while 45.8% are either lightly renovated or unfinished. Centralized heating is used in 85.2% of the properties, and 62.21% of the units are equipped with elevators, whereas 37.79% are not. The distributions of these housing attributes show no extreme skewness, and they closely reflect the actual supply and demand structure of the Beijing second-hand housing market, demonstrating good representativeness.

4.2. Baseline OLS Regression Results

This study first constructs a full regression model including all explanatory variables. Based on statistical significance and economic relevance, insignificant variables such as house type, floor level, and orientation were gradually removed, resulting in a final model retaining the core significant variables. To visually illustrate the effect of building area on housing prices, the following scatter plot is presented:

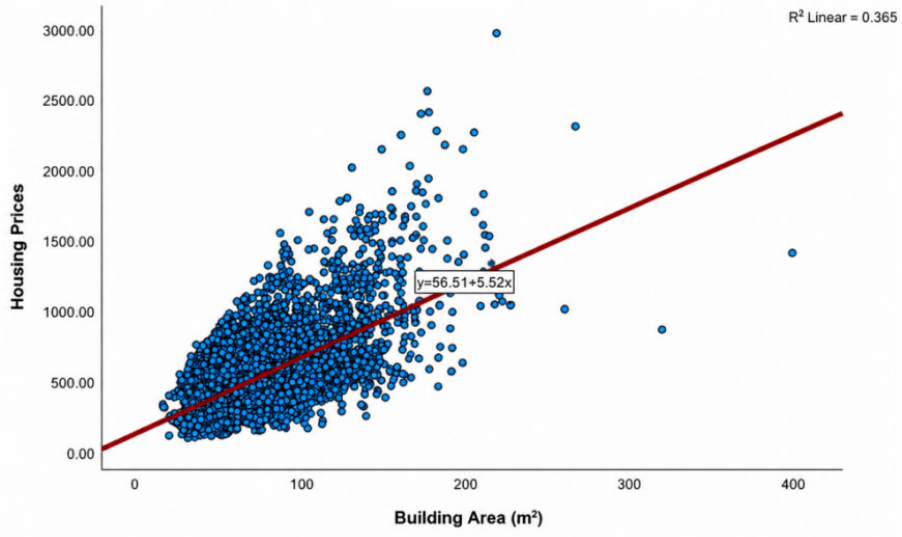

Figure 1. Scatter Plot of Second-Hand Housing Prices and Building Area in Beijing

As shown in Figure 1, the total price of second-hand houses in Beijing exhibits a positive trend with increasing building area, visually reflecting the distribution of housing prices and the overall upward trend, providing a basis for subsequent analysis.

Variance inflation factor (VIF) diagnostics of the regression model indicate that there is no serious multicollinearity among the variables, suggesting that the model is reliable. The baseline OLS regression results are presented as follows:

Table 4. OLS Regression Results

Variable	B (Unstandardized Coefficient)	Std. Error	Beta (Standardized Coefficient)	t-value	p-value	VIF
Constant	172.580	10.876	-	15.869	0.000	-
Building Area	5.924	0.086	0.648	68.734	0.000	1.114
School District Status	267.830	5.011	0.480	53.449	0.000	1.013
High-End Renovation	17.927	5.072	0.032	3.535	0.000	1.032
Centralized Heating	40.429	7.142	0.051	5.661	0.000	1.037
Elevator Availability	22.849	5.369	0.040	4.256	0.000	1.096

As shown in Table 4, the overall model fit is satisfactory, with an adjusted R^2 of 0.601 and an F-statistic of 1,508.744, indicating significance at the 1% level. This suggests that the explanatory variables effectively account for variations in the listed total price of second-hand houses. The variance inflation factor (VIF) values for all variables are below 1.2, indicating no serious multicollinearity.

Regarding the core explanatory variables, the regression results show that an increase of 1 m² in building area corresponds to an approximate increase of 5.92 thousand CNY in house price. Properties located in school districts exhibit a premium of 267.83 thousand CNY compared with non-school district properties, both statistically significant at the 1% level. Among the control variables, high-end renovation, centralized heating, and elevator availability also have significant positive effects on housing prices.

Based on the regression analysis results, a box plot is drawn to visually illustrate the impact of school district status on housing prices, as shown below:

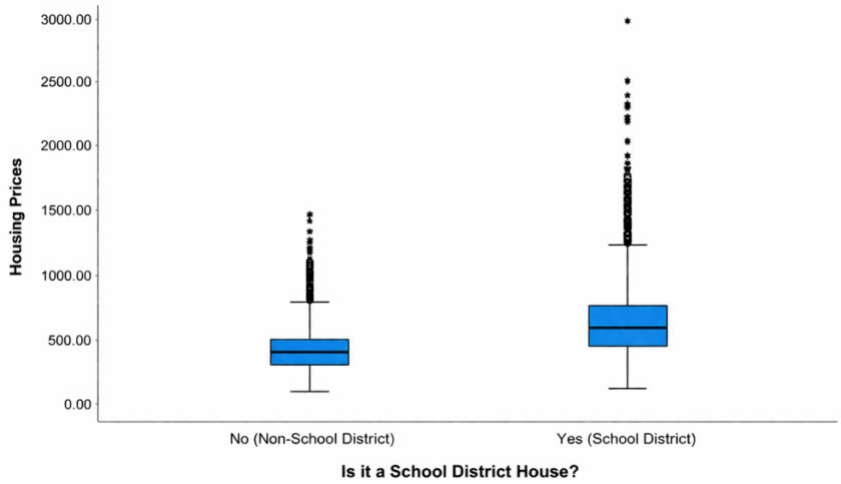

Figure 2. Box Plot of Housing Prices for School District and Non-School District Properties

As shown in Figure 2, the prices of school district properties are significantly higher than those of non-school district properties, with a higher median and a wider range of price distribution, reflecting the school district premium.

4.3. Model Applicability Test

4.3.1. Residual Normality Test

The above figures and regression results indicate that both building area and school district status have significant positive effects on housing prices. To ensure the robustness of the regression results, the key assumptions of the OLS regression model were subsequently tested. First, the normality of the residuals was examined through statistical analysis. The results of the residuals are presented in the table below:

Table 5. Descriptive Statistics of Model Residuals

Indicator	Min	Max	Mean	Standard Deviation
Predicted Value	6.79	2,497.83	497.92	216.07
Residual	-1,115.83	1,477.11	0.00	175.91
Standardized Predicted Value	-2.27	9.26	0.00	1.00
Standardized Residual	-6.34	8.39	0.00	1.00

As shown in Table 5, the residuals have a mean of 0 and a standard deviation of 175.91, while the standardized residuals have a mean of 0 and a standard deviation of 1, indicating that the overall residuals satisfy the basic characteristics of zero mean and unit variance. The range of standardized residuals is from -6.34 to 8.39, with most values concentrated within the ±2 interval, and no extreme outliers are observed.

To visually illustrate the distribution of the residuals, a histogram of the residuals was plotted as follows:

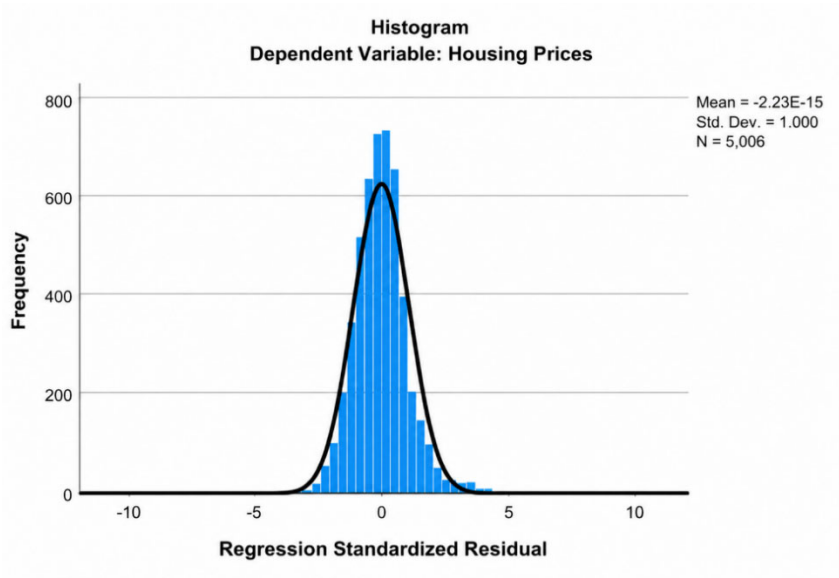

Figure 3. Histogram of Model Residuals

As shown in Figure 3, the residuals exhibit an approximately bell-shaped distribution, with most standardized residuals concentrated within the ±2 range, consistent with the pattern of a normal distribution. Combined with the results in Table 5 and Figure 1, it can be concluded that the residuals are approximately normally distributed, satisfying the normality assumption of the OLS regression.

4.3.2. Heteroscedasticity Test

To examine heteroscedasticity in the initial OLS regression model, a scatter plot was drawn with standardized predicted values on the horizontal axis and standardized residuals on the vertical axis, as shown below:

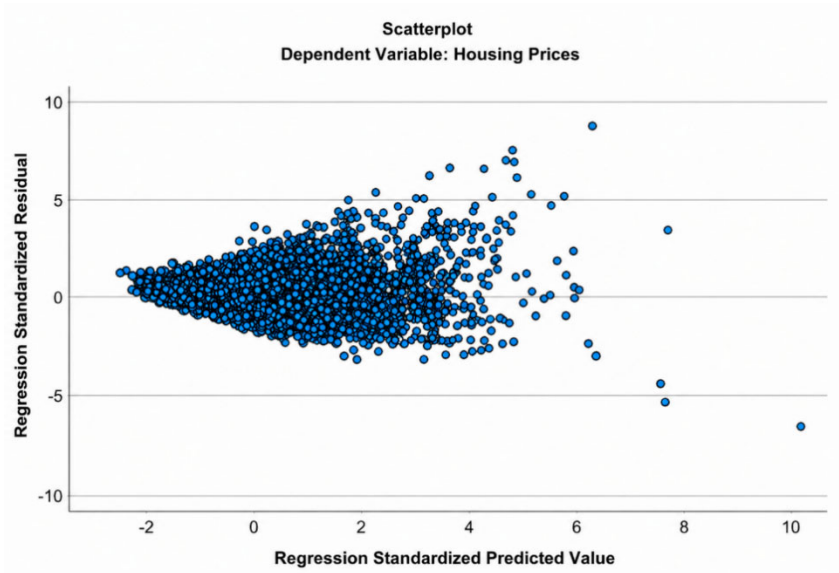

Figure 4. Scatter Plot of Standardized Residuals from Regression

As shown in Figure 4, the residuals slightly spread as the standardized predicted values increase, forming a mild funnel shape, indicating the presence of heteroscedasticity in the initial OLS regression. Heteroscedasticity may lead to underestimated standard errors of the regression coefficients, affecting the assessment of statistical significance. Therefore, the model was corrected using weighted least squares (WLS) to ensure robustness.

4.4. Re-estimation Using Weighted Least Squares (WLS)

Considering that heteroscedasticity may result in biased standard errors and affect statistical inference, this study applies weighted least squares (WLS), using the inverse of the squared predicted values as weights to re-estimate the model. The WLS regression results are presented in Table 6.

Table 6. WLS Regression Results

Variable	B (Unstandardized Coefficient)	Std. Error	t-value	p-value	Direction& Significance
Constant	22.010	4.298	5.121	<0.001	-
Building Area	4.422	0.065	67.531	0.000	Positive, significant (1%)
School District Status	221.803	3.985	55.659	0.000	Positive, significant (1%)
High-End Renovation	0.710	2.953	0.240	0.810	Positive, not significant
Centralized Heating	-26.296	3.436	-7.653	0.000	Negative, significant (1%)
Elevator Availability	28.161	3.280	8.586	0.000	Positive, significant (1%)

As shown in Table 6, the model fit improves noticeably, with the adjusted R^2 increasing from 0.601 to 0.687, and an F-statistic of 2,199.926 ($p < 0.001$), indicating that the model fits well and explains approximately 68.7% of the variation in the listed total price of second-hand houses. The significance and direction of the core explanatory variables remain unchanged, while the effect of high-end renovation is no longer significant, suggesting that the heteroscedasticity issue has been effectively addressed and that the model conclusions are robust and reliable.

The core explanatory variables indicate that building area and school district status have significant positive effects on housing prices. Specifically, an increase of 1 m² in building area corresponds to an average increase of approximately 44.2 thousand CNY in the price of second-hand houses, while properties located in school districts exhibit an average premium of about 2.218 million CNY compared with non-school district properties, both significant at the 1% level. Among the control variables, properties equipped with elevators have significantly higher prices than those without, whereas centralized heating has a negative effect on prices. After controlling for the core variables, high-end renovation shows no significant marginal effect ($p = 0.810$), suggesting that homebuyers focus more on core structural attributes, and renovation quality is less reflected in the price.

5. Conclusion

5.1. Analysis of Regression Results

This study focuses on the second-hand housing market in Beijing, selecting five core variables, including building area, school district status, heating type, and elevator availability, to construct a multiple linear regression model. After conducting normality and heteroscedasticity tests of the residuals and applying weighted least squares (WLS) correction, the model results are robust and reliable. The main conclusions are as follows:

Building area has a significant positive effect on housing prices. Specifically, an increase of one square meter corresponds to an average increase of approximately 0.442 million CNY in the total price of second-hand houses, indicating that living space plays an important role in property value.

School district status has a significant positive effect on housing prices. Properties located in school districts exhibit an average premium of approximately 8,000 CNY per square meter, highlighting the substantial influence of high-quality educational resources on property value.

Elevator availability has a significant positive effect on housing prices. Properties equipped with elevators have noticeably higher prices than those without, with an average increase of approximately 0.2816 million CNY, indicating that elevators enhance residential convenience and comfort, particularly for high-rise apartments and elderly households.

Centralized heating has a negative effect on housing prices. Second-hand houses with centralized heating are priced on average approximately 0.263 million CNY lower than those with individual heating. This is mainly because such properties are often located in peripheral areas or older residential complexes, where the baseline prices are relatively low, offsetting the potential positive impact of heating.

High-end renovation has no significant effect on housing prices, with a regression coefficient of 0.710 and $p = 0.810$. This indicates that, after controlling for core variables, homebuyers are relatively insensitive to renovation level, and its value is not strongly reflected in the total price.

Overall, the model demonstrates good robustness. The initial OLS model exhibited mild heteroscedasticity; after WLS correction, the adjusted R^2 increased from 0.601 to 0.687. The direction and significance of the core variable coefficients remained stable, with only high-end renovation becoming insignificant, confirming the robustness of the model conclusions.

5.2. Policy Recommendations

Based on the above research findings, relevant recommendations can be proposed from the perspectives of the government, homebuyers, and real estate developers:

For the government, it is recommended to adjust the allocation of educational and infrastructure resources to balance high-quality education, control the school district premium, and promote educational equity; advance elevator retrofitting in older residential complexes to enhance property value and residential convenience; improve urban heating systems to reduce disparities across different locations, ensuring that heating value is reasonably reflected; and strengthen the disclosure of second-hand housing transaction information and market supervision to ensure fairness and transparency in transactions.

For homebuyers, attention should be focused on the core value of the property to enhance rational decision-making. Priority should be given to key factors such as building area, school district status, and elevator availability, while also considering location, property age, and heating conditions. Housing choices should be aligned with actual household needs—for instance, properties with elevators are preferable for households with elderly members, and school district properties should be considered for families with educational requirements.

For homebuyers, attention should be focused on the core value of the property to enhance rational decision-making. Priority should be given to key factors such as building area, school district status, and elevator availability, while also considering location, property age, and heating conditions. Housing choices should be aligned with actual household needs—for instance, properties with elevators are preferable for households with elderly members, and school district properties should be considered for families with educational requirements.

This study analyzes the main factors influencing second-hand housing prices, providing references for the government, homebuyers, and real estate developers, thereby contributing to the stable development of the secondary housing market. The study is based on the Beijing second-hand housing market from 2020 to 2024, and the conclusions primarily apply to this period and regional context. Future research could extend the analysis to other cities or a longer time span to further examine the robustness of the findings.

References

[1] Hu C, Liu Q. Suggestions on the development of China's second-hand housing market[J]. China Real Estate Finance, 2007, (07): 32-36.

[2] Huo Y. Research on forecasting second-hand house prices based on single and combined models[D]. China University of Geosciences (Beijing), 2024. DOI:10.27493/d.cnki.gzdzy.2024.000002.

[3] Wang Y, Feng Y, Han K, et al. Analysis of the Temporal and Spatial Patterns of Residential Prices in Qingdao and Its Driving Factors[J]. Buildings, 2025, 15(2): 195.

[4] Tang Q L, Hu W X. Study on the impact of rail transit on housing price along the line based on characteristic price model–Take Changsha Rail Transit Line 1 as an example[J]. Chinese Journal of Railway Science and Engineering, 2021, 19(2): 570-578.

[5] Wen H. Hedonic pricing of urban residential housing: Theoretical analysis and empirical study[D]. Zhejiang University, 2004.

[6] Yang J, Shao X, Peng Z, et al. Study on the factors affecting second-hand house prices in Beijing[J]. Real Estate World, 2023, (04): 25-28.

[7] Peng Z, Huang Q, Han Y. Model research on forecast of second-hand house price in Chengdu based on XGboost algorithm[C]//2019 ieee 11th international conference on advanced infocomm technology (icait). IEEE, 2019: 168-172.

[8] Li H, Wei J, Lu Y. Prediction and analysis of second-hand house prices in Shenzhen based on random forest[J]. Modern Information Technology, 2021, 5(15): 100-104. DOI:10.19850/j.cnki.2096-4706.2021.15.026.

[9] Wang H, Meng J. Predictive modeling method based on multiple linear regression[J]. Journal of Beijing University of Aeronautics and Astronautics, 2007, (04): 500-504. DOI:10.13700/j.bh.1001-5965.2007.04.028.

[10] Gui J, Li S. Empirical analysis of factors affecting real estate prices in Sichuan Province[J]. Quality and Market, 2022, (17): 4-7.

[11] Bai X. Methods and review of heteroscedasticity tests[J]. Journal of Dongbei University of Finance and Economics, 2002, (06): 26-29. DOI:10.19653/j.cnki.dbcjdxxb.2002.06.002.