

# Research on Influencing Factors of Second-Hand Housing Price in Haidian District, Beijing

Kunling Han

School of Mathematics and Big Data, Dezhou University, Shandong, China, 253023

han\_dzu@163.com

**Abstract.** The housing issue is an economic issue as well as an issue of people's livelihood, especially in megacities like Beijing, where housing prices in the second-hand housing market fluctuate greatly, so its influencing factors have become the focus of attention. Studying the influencing factors of housing prices in the second-hand housing market can provide reference for the government to formulate appropriate real estate regulation policies and predict the development trend of housing prices in the future. This article collects nearly 3,000 groups of public second-hand housing individual transaction data from the Lianjia website in the domestic second-hand housing market. The time span of the data is from June 2020 to June 2021. After data cleaning and descriptive statistical analysis of the acquired data, we set the total price of the house as the explained variable, and the number of rooms, the area of the house, the decoration, the floor, and the unit price of the house are used as the explanatory variables. We establish a multiple linear regression model, then test and optimize it. The results are as follows: in the second-hand housing market of Haidian District, if the second-hand house has a large construction area, a large number of rooms, a high unit price and has been renovated, the total price of the house will be high.

**Keywords:** Second-Hand Housing Price, Influencing Factors, Multiple Linear Regression.

## 1. Introduction

Haidian District is developed in Beijing. It has gathered many famous universities in Beijing. Zhongguancun has become synonymous with China's high-tech industry and plays a very important role in promoting the development of Beijing in all aspects. The continuous introduction of talents will indirectly lead to the continuous increase of housing demand, and the increase of housing demand will also lead to the rise of housing prices. Second-hand housing, as a powerful supplement to the three major housing guarantees in our country, plays a multifaceted role. Therefore, it is of great practical significance to analyze the second-hand housing market in Haidian District.

Nam et al. [1] used Cronbach's  $\alpha$  test and factor analysis to determine the influencing factors and their impact rates of commercial housing prices in Bac Ninh city. In addition, foreign scholars have made attempts in the field of housing price evaluation by using machine learning methods such as random forests, neural networks, and decision trees [2], as well as GIS and fuzzy mathematics, which have greatly improved the accuracy of the housing price prediction model.

In recent years, Chinese scholars have applied a variety of statistical methods to evaluate housing prices, the most typical method is to establish a characteristic price model for regression analysis [3-5]. Jin Tu et al. [6] used the second-hand housing market transaction data in Chengdu to analyze the important factors affecting urban housing prices through the difference-in-difference method, and compared the two cities of Chengdu and Chongqing to find more effective regulatory policies. Weihua Wang et al. [7] established a VAR model to study the dynamic changes in the growth rate of new and second-hand housing prices in Shenzhen, and added dummy variables and average down payment ratios as exogenous variables to investigate the fluctuation of housing prices. In particular, the research on second-hand housing prices in Beijing, the capital, is also in full swing. Most of the research focuses on the impact of education quality on second-hand housing prices, that is, the premium of housing in school districts [8-10]. In addition, machine learning methods such as random forest regression, Bagging, Boosting, and bat algorithms are also widely used in the prediction of second-hand housing prices in Beijing [11-12].

We find that although the linear regression model will produce large errors when encountering nonlinear problems, it has a simple structure and strong explanatory power. This paper collects the sales data of the second-hand housing market in Haidian District through the Lianjia website, analyzes the current situation of the second-hand housing market and establishes a multiple linear regression model. We study the influence of housing decoration, construction area, orientation and other related factors on the housing price of second-hand housing in Haidian District. The purpose of this research is to obtain a more accurate evaluation model of second-hand housing prices in Haidian District, find the significant influencing factors, and get the impact direction of each influencing factor and the specific numerical performance of the impact of each factor on housing prices. We hope to provide more accurate information to consumers who are interested in the second-hand housing market, so that they can have a more accurate and clear understanding of the value of the second-hand house price, so as to provide a certain reference for buyers and sellers.

## 2. Descriptive statistics

### 2.1. Data processing

This article employs the Scrapy framework in Python to gather and organize publicly available information on second-hand houses in Haidian District from the Lianjia website. We randomly selected the top 100 pages, each containing 30 listings, resulting in a total of 3000 data points.

After data acquisition, preprocessing is necessary to prepare for subsequent modeling and analysis. (1) Remove Unusable Variables: Although information such as attention levels and release times can have a certain impact on the displayed second-hand housing prices on the platform, these variables are difficult to be quantified and incorporated into the model construction. Therefore, community names, attention levels, and release times should be removed. (2) Missing Value Handling: Due to the large variation in values of each variable in each group of data obtained from web crawling, and some variables having missing values, accurate analysis and reliable conclusions can only be obtained by first handling the missing values in the raw data obtained from the website. The types with fewer values in each category of variables were removed, resulting in 2722 valid data points. The time span of these valid data points is from June 2020 to June 2021.

### 2.2. Indicator selection

During the process of selecting indicators, we comprehensively considered the reliability and availability of data and combined them with the characteristics of Haidian District itself to determine eight residential characteristic factors. The selected residential characteristics include: housing type, housing area, orientation, decoration, floor, building structure, total housing price, and unit housing price.

### 2.3. Data description

#### 2.3.1 Overall market data analysis

Table 1 presents a descriptive statistical analysis of the total and unit price distribution in the second-hand housing market in Haidian district. The results indicate that the average total price of houses in this market is 10.1282 million yuan, with significant variance and range among different groups, displaying a severe right-skewed distribution. On the other hand, the average unit price is 103,877.06 yuan per square meter, with large standard deviation and variance, also showing a severe skewed distribution. The total price is influenced by multiple factors, while the fluctuation of unit price is most likely related to the location of the property. Generally speaking, properties located in more developed areas tend to have higher unit prices.

**Table 1.** Descriptive statistics of total price and unit price

|             | Minimum | Maximum | Mean      | Std. Deviation | Variance    | Skewness | Kurtosis |
|-------------|---------|---------|-----------|----------------|-------------|----------|----------|
| Total price | 275     | 5800    | 1012.82   | 559.009        | 312491.292  | 2.120    | 7.630    |
| Unit price  | 29994   | 179960  | 103877.06 | 26794.988      | 717971393.1 | 0.475    | 0.094    |

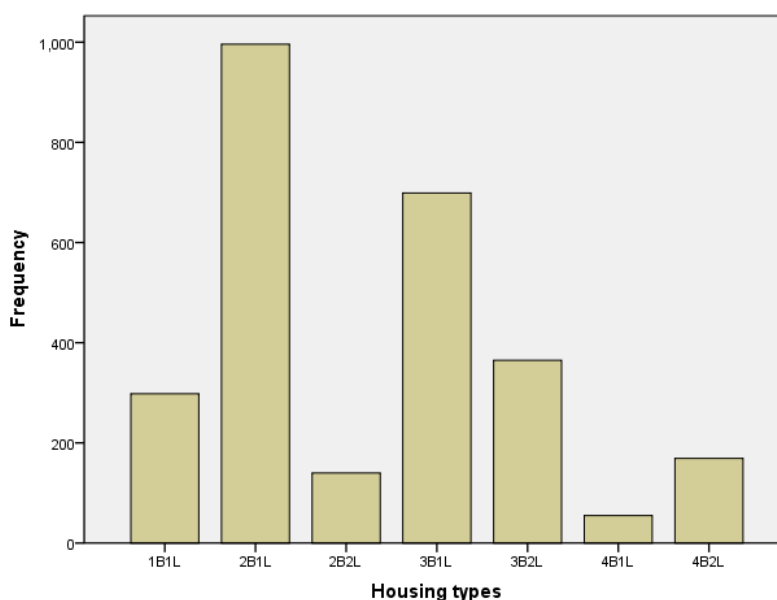
**2.3.2 Analysis of the various elemental factors of the variables**

(1) Distribution chart of housing types

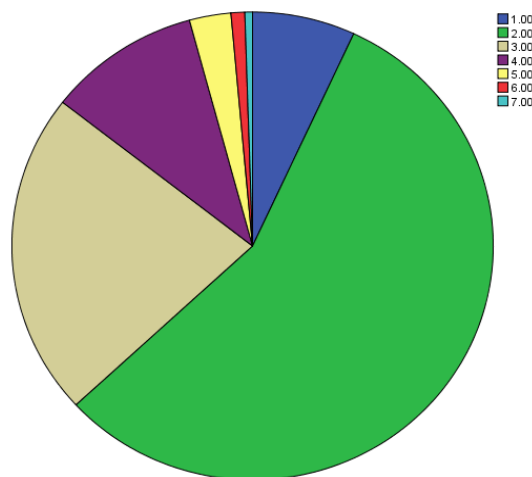
Figure 1 displays a distribution chart of residential unit types. Upon examining the seven most prevalent unit types, it is evident that the 2-bedroom, 1-living room layout has the highest number of transactions for second-hand properties. Following closely are the 3-bedroom, 1-living room and 3-bedroom, 2-living room layouts. Together, these three types account for a total of 2484 units, representing 75.7% of the total number. This indicates that when selecting second-hand properties in Haitian District, most people tend to consider the 2-bedroom, 1-living room layout due to the large influx of non-native residents in Beijing and its high housing prices compared to other cities in China. Therefore, people tend to opt for smaller unit types with lower prices when renting properties. The aforementioned layout is reasonably priced and affordable for most individuals seeking rental accommodations.

(2) Distribution chart of housing area

Based on the data, the area of houses ranges from 26 square meters to 642 square meters with significant differences and scattered distribution. Therefore, the areas are first classified into groups with width of 50 square meters. The data is divided into seven groups based on the area variable: below 50, [50, 100), [100, 150), [150, 200), [200, 250), [251, 300) and above 300. The group names are represented by numbers 1 to 7. The area distribution chart and the number statistics for each category are shown in Figure 2 and Table 2, respectively. Combining with the pie chart, it is known that the largest number of second-hand houses sold are in the [50, 100) square meter range, while the number of transactions for second-hand houses above 250 square meters is relatively small. In other words, the vast majority of second-hand house buyers choose houses with an area between 50 and 100 square meters.



**Figure 1.** Distribution of housing types



**Figure 2.** Distribution of housing area

**Table 2.** Distribution of housing area

|                  | Frequency | Percent | Valid Percent | Cumulative Percent |
|------------------|-----------|---------|---------------|--------------------|
| Below 50         | 188       | 6.9     | 6.9           | 6.9                |
| [50, 100)        | 1530      | 56.2    | 56.2          | 63.1               |
| [100, 150)       | 610       | 22.4    | 22.4          | 85.5               |
| Valid [150, 200) | 279       | 10.2    | 10.2          | 95.8               |
| [200, 250)       | 76        | 2.8     | 2.8           | 98.6               |
| [250, 300)       | 25        | 0.9     | 0.9           | 99.5               |
| Above 300        | 14        | 0.5     | 0.5           | 100.0              |
| Total            | 2722      | 100.0   | 100.0         |                    |

(3) Distribution chart of house orientations

According to the website data, the second-hand houses sold in Haidai district have different orientations depending on their geographical location, and even within the same range, there are variations in orientations. Due to the complex orientations of houses in Haiden district, only six types are selected for descriptive statistics, namely East, Southeast, West, Southwest, South, and North. The frequency of each orientation is calculated as shown in Figure 3.

The first three orientations are south, southwest, and southeast, with frequencies of 331, 166, and 148 respectively. People tend to choose south-facing houses because they have good ventilation and lighting, and are warm in winter and cool in summer. Houses facing southwest and southeast are not as good as south-facing ones, but still have relatively good lighting compared to those facing north and west. Fewer people choose houses facing west mainly because the lighting time is short and there may be exposure to direct sunlight in the afternoon during summer. The smallest number of people choose houses facing north because they have the worst lighting and are relatively cold in winter.

(4) Distribution chart of floor

Lianjia, based on the location of each property and its position within the entire building, divided the floors into three segments: low floors, mid floors, and high floors. The frequency of these three segments was statistically counted and a bar figure was drawn. Figure 4 shows that there were 864, 964, and 894 housing units in the high, mid, and low floor segments respectively. The distribution of the segments was relatively even, with the mid floor segment being more popular. This could be due to the fact that the elevators in the mid to high floors are more advanced, the scenery is more beautiful, and the lighting is brighter, which provides a better quality of life. On the other hand, lower floors may have some disadvantages such as relatively humid environment, depressive atmosphere, and poor privacy protection. Therefore, second-hand properties on middle floors are highly sought after by homebuyers in Haidian District.

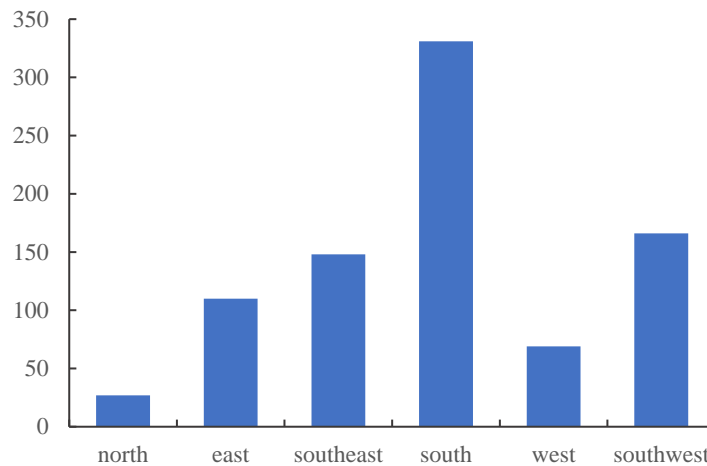


Figure 3. Distribution of house orientations

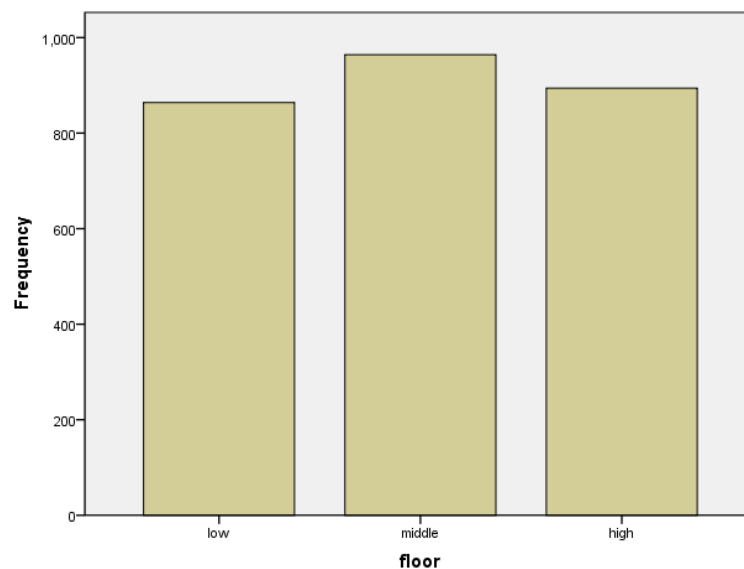


Figure 4. Distribution of floor

### 3. Model construction and analysis

#### 3.1. Variable selection and measurement

In the real estate market, the price of second-hand houses is influenced by many factors. This article collects and organizes all data categories and references literature, taking the total house price in each data set as the dependent variable and other variables as independent variables, including the number of rooms, house area, decoration, floor level, and unit price.

There are many types of housing layouts on the website, but the distribution of each type of layout is uneven, and the range is large. After performing frequency statistics on all types of layouts and eliminating layouts with a frequency of less than or equal to 30, 7 types of layouts were finally retained, namely: 1 bedroom and 1 living room, 2 bedrooms and 1 living room, 2 bedrooms and 2 living rooms, 3 bedrooms and 1 living room, 3 bedrooms and 2 living rooms, 4 bedrooms and 1 living room, and 4 bedrooms and 2 living rooms.

According to different layouts, the actual number of rooms can be calculated. The house area and unit price refer to the actual values provided in the website's housing sources. The degree of decoration is divided into four categories: roughcast, fine decoration, simple decoration, and other. The floor level is divided into three categories: low, middle, and high. In addition, to improve the universality and reliability of the conclusions, only conventional transaction data in the second-hand

housing market of the city are studied, so housing data with total prices exceeding 30 million are directly deleted from the original data. Finally, a model is constructed using 2698 sets of data.

**3.2. Model construction**

The variable view coding for the explanatory variables, with the explained variable y being the total price of the house, is shown in Table 3.

**Table 3.** Variable view encoding

| Variable             | Quantitative indicator                               | Code  |
|----------------------|--|---|
| the number of rooms  | number of actual rooms, unit: pieces                 | X <sub>1</sub>  |
| house area           | actual area size, unit: square meters                | X <sub>2</sub>  |
| decoration           | fine decoration                                      | 1=yes, 0=no   |
|                      | simple decoration                                    | 1=yes, 0=no   |
|                      | roughcast  | 1=yes, 0=no   |
|                      | other  | when the values of X <sub>31</sub> , X <sub>32</sub> , and X <sub>33</sub> are all equal to 0 |
| floor                | 1=low, 2=middle, 3=high                              | X <sub>4</sub>  |
| unit price of house  | actual unit price value, unit: yuan per square meter | X <sub>5</sub>  |
| total price of house | actual total price value, unit: ten thousand yuan    | Y   |

A multiple linear regression analysis was conducted on the second-hand housing market data in Haidian District, and the regression equation was derived based on Table 4, followed by a significance test of the regression coefficients. It can be seen that the absolute value of the T-test statistic for the number of rooms, house area, whether it is fine decorated, and house unit price are relatively large. Under the condition that  $\alpha$  is 0.05, the corresponding p-values for the explanatory variables are all less than  $\alpha$ , indicating that the regression coefficients for the explanatory variables are significant. At the same time, the tolerance is close to 1 and the variance inflation factor (VIF) is close to 1, indicating that there is no serious multicollinearity. The regression equation is as follows:

$$\hat{Y} = -1029.187 + 27.117X_1 + 9.278X_2 + 30.728X_{31} + 0.01X_5 \tag{1}$$

**Table 4.** List of variables for model fitting

| Model           | Unstandardized Coefficients | Std. Error | Standardized Coefficients | t       | Sig.  | Collinearity Statistics |       |
|-----------------|-----------------------------|------------|---------------------------|---------|-------|-------------------------|-------|
|                 |                             |            |                           |         |       | Tolerance               | VIF   |
| Constant        | -1029.187                   | 16.954     |                           | -60.705 | 0.000 |                         |       |
| X <sub>1</sub>  | 27.117                      | 3.520      | 0.057                     | 7.704   | 0.000 | 0.431                   | 2.320 |
| X <sub>2</sub>  | 9.278                       | 0.081      | 0.873                     | 114.957 | 0.000 | 0.413                   | 2.421 |
| X <sub>31</sub> | 30.728                      | 8.688      | 0.031                     | 3.537   | 0.000 | 0.306                   | 3.271 |
| X <sub>32</sub> | 15.563                      | 8.880      | 0.016                     | 1.753   | 0.080 | 0.305                   | 3.282 |
| X <sub>33</sub> | -35.498                     | 23.800     | -0.008                    | -1.492  | 0.136 | 0.897                   | 1.115 |
| X <sub>4</sub>  | -0.473                      | 3.004      | -0.001                    | -0.158  | 0.875 | 0.992                   | 1.008 |
| X <sub>5</sub>  | 0.010                       | 0.000      | 0.518                     | 102.718 | 0.000 | 0.938                   | 1.066 |

**3.3. Statistical test of regression equations**

After establishing the multiple linear regression equation, a series of tests should be conducted to validate the effectiveness of the model. This article mainly includes four types of tests: goodness-of-fit test for the regression equation, significance test for the regression equation, significance test for the regression coefficients, and heteroscedasticity test.

(1) Goodness-of-fit test for the regression equation

Based on Table 5, a goodness-of-fit test can be conducted for the equation. The adjusted coefficient of determination is 0.936, which means that 93.6% of the total variance in the total price of the second-hand housing market can be explained by the regression equation, and the part that cannot be explained by the model is very small. Because this number is very close to 1, it can be determined

that a high degree of fit has been obtained. The DW test value is 1.926, and the closer it is to 2, the greater the possibility that there is no serial correlation in the residuals.

**Table 5.** Regression equation test result

| Model | R     | R-square | Adjusted R-Square | Std. Error of the Estimate | Durbin-Watson |
|-------|-------|----------|-------------------|----------------------------|---------------|
| 1     | 0.967 | 0.936    | 0.936             | 124.761                    | 1.926         |

(2) Significance test for the regression equation

According to Table 6, the total sum of squares for the explained variable is 653078735.253. As for the residuals, the sum of squares is 41870697.585, and the mean square is 15565.315. The sum of squares for regression is 611208037.668, and the variance is 87315433.953. In the significance test for the regression equation, the value of the F-test statistic is very large, and as for the probability p-value, it can be approximated to be close to 0. Under the condition that the significance level  $\alpha$  is 0.05, the null hypothesis "there is no significant difference between each partial regression coefficient and zero" in the significance test for the regression equation is rejected, and there is sufficient reason to believe that the regression coefficients of this multiple linear regression equation are not all zero. Therefore, it can be concluded that the linear relationship between the total price of second-hand housing in Haidian District and the explanatory variables is significant, and a linear model can be established between the explanatory variables and the explained variable.

**Table 6.** ANOVA

| Model      | Sum of Squares | df   | Mean Square  | F        | Sig. |
|------------|----------------|------|--------------|----------|------|
| Regression | 611208037.668  | 7    | 87315433.953 | 5609.616 | .000 |
| Residual   | 41870697.585   | 2690 | 15565.315    |          |      |
| Total      | 653078735.253  | 2697 |              |          |      |

(3) significance test for the regression coefficients

According to Table 4, under the condition that the significance level  $\alpha$  is 0.05, the probability p-values for the coefficients of the number of rooms, house area, whether it is fine decorated, floor level, and house unit price are all less than  $\alpha$ , indicating that the null hypothesis can be rejected and there is a significant difference between the regression coefficients and zero. Therefore, the linear relationship between the explained variable y and the above explanatory variables is significant, and these variables should be retained in the regression equation.

(4) heteroscedasticity test

To test whether the assumption of homoscedasticity for the random error term is satisfied, that is, to determine whether there is a significant correlation between the absolute value of the unstandardized residual term and the unstandardized predicted value, the test results are shown in Table 7. The Spearman rank correlation coefficient between the predicted value and the residual is 0.263, and the P-value is less than the significance level, indicating that there is heteroscedasticity. By plotting the scatter plot of the unstandardized predicted value and the unstandardized residual term (see Figure 5), it can also be seen clearly that these points are not randomly distributed, but there is some kind of correlation.

**Table 7.** Heteroscedasticity test

|                |                                | Unstandardized Predicted Value | ABSE    |
|----------------|--------------------------------|--------------------------------|---------|
| Spearman's rho | Unstandardized Predicted Value | Correlation                    | 1.000   |
|                |                                | Coefficient                    | 0.263** |
|                |                                | Sig.(2-tailed)                 | 0.000   |
|                | ABSE                           | N                              | 2698    |
|                |                                | Correlation                    | 0.263** |
|                |                                | Coefficient                    | 1.000   |
|                |                                | Sig.(2-tailed)                 | 0.000   |
|                |                                | N                              | 2698    |

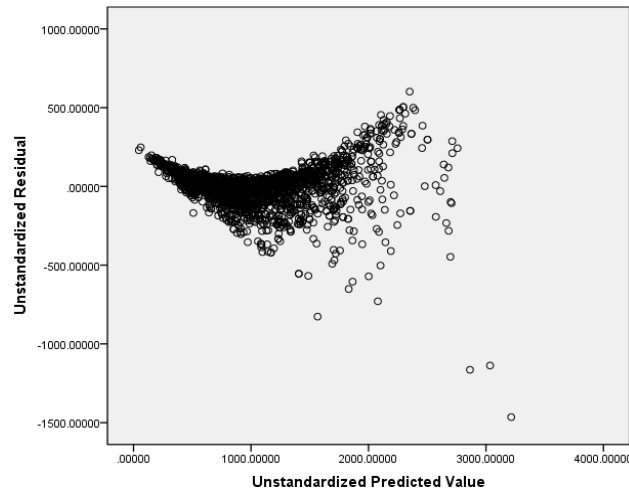


Figure 5. Residual plot

### 3.4. Model modification

To eliminate the impact of heteroscedasticity on the regression equation, this article uses the Weighted Least Squares method for model estimation. That is, different weights are assigned based on the degree of data variability during model fitting. Larger weights are assigned to data with smaller variability and more accurate measurements, while smaller weights are assigned to data with larger variability and less stable measurements. This ensures that the sum of squares of the residuals of the weighted regression line is minimized, and the fitted model has better predictive value. The weight used here is the reciprocal of the absolute value of the residual. The goodness-of-fit and regression coefficients of the corrected model are shown in Table 8 and Table 9, respectively.

It can be observed that after using the Weighted Least Squares method to correct the model, the goodness-of-fit of the model has improved, and the explanatory power of the independent variables is stronger compared to the ordinary least squares method. Except for X4, all variables have passed the significance test at the 0.05 level, indicating that the regression coefficients are significant. At the same time, the tolerance is close to 1, and the variance inflation factor (VIF) is less than 10, indicating that there is no serious multicollinearity.

In summary, the final form of the model is determined as:

$$\hat{Y} = -992.206 + 21.335X_1 + 9.406X_2 + 23.442X_{31} + 9.401X_{32} - 28.342X_{33} + 0.009X_5 \quad (2)$$

Where Y is the total price of the second-hand housing market in Haidian District, and  $X_1$ ,  $X_2$ ,  $X_{31}$ ,  $X_{32}$ ,  $X_{33}$ ,  $X_5$  are the number of rooms, house area, whether it is fine decorated, whether it is simple decorated, whether it is bare-shell, and house unit price, respectively.

Table 8. Regression equation test result

| Model | R     | R-square | Adjusted R-Square | Std. Error of the Estimate | Durbin-Watson |
|-------|-------|----------|-------------------|----------------------------|---------------|
| 1     | 0.996 | 0.993    | 0.993             | 8.792                      | 1.980         |

Table 9. List of variables for model fitting

| Model           | Unstandardized Coefficients | Std. Error | Standardized Coefficients | t        | Sig.  | Collinearity Statistics |       |
|-----------------|-----------------------------|------------|---------------------------|----------|-------|-------------------------|-------|
|                 |                             |            |                           |          |       | Tolerance               | VIF   |
| Constant        | -992.206                    | 5.824      |                           | -170.379 | 0.000 |                         |       |
| X <sub>1</sub>  | 21.335                      | 1.011      | 0.055                     | 21.103   | 0.000 | 0.392                   | 2.552 |
| X <sub>2</sub>  | 9.406                       | 0.028      | 0.893                     | 339.091  | 0.000 | 0.383                   | 2.614 |
| X <sub>31</sub> | 23.442                      | 2.182      | 0.040                     | 10.742   | 0.000 | 0.189                   | 5.304 |
| X <sub>32</sub> | 9.401                       | 2.278      | 0.015                     | 4.127    | 0.000 | 0.190                   | 5.251 |
| X <sub>33</sub> | -28.342                     | 13.798     | -0.003                    | -2.054   | 0.040 | 0.976                   | 1.025 |
| X <sub>4</sub>  | -1.100                      | 0.598      | -0.003                    | -1.841   | 0.066 | 0.850                   | 1.176 |
| X <sub>5</sub>  | 0.009                       | 0.000      | 0.558                     | 289.771  | 0.000 | 0.716                   | 1.396 |

## 4. Conclusion

This article selected five categories of indicators provided by the original dataset as independent variables, and the total housing price as the dependent variable, conducting a multiple linear regression analysis on these variables. A multiple linear regression model was obtained, and a series of significance tests were conducted on the model. The independent variables that entered the model were determined through the tests.

Through the analysis of the model, the following conclusions can be drawn:

(1) The total price of second-hand housing in Haidian District is influenced by many factors. Among them, the number of rooms, house area, house unit price, and whether it is decorated have a greater impact. Second-hand houses with good layout, more rooms, larger area, with decoration and higher unit price, have a higher total price.

(2) Under the same conditions, appropriately renovating second-hand houses can increase their price. People prefer second-hand houses that have been renovated, and the impact coefficient of fine decoration on the total housing price is the highest, at 23.442, while the impact coefficient of simple decoration on the total housing price is 9.401. Because in other cases, choosing this type of housing allows buyers to move in directly, saving a lot of time. The impact of bare-shell houses that have not been renovated on the total housing price is negative, with the largest absolute value of the coefficient being -28.342. The reason may be that choosing bare-shell houses will greatly increase the costs that buyers have to pay.

(3) The height of the floor does not have a significant impact on the price of second-hand houses because Haidian District is one of the highly developed economic areas in Beijing, and most residential buildings are high-rise buildings. Therefore, there is no significant difference in the height of the floor. From the previous descriptive statistics, it can also be found that there is no significant difference in the frequency of low, middle, and high floors, indicating that when people choose housing, the height of the floor is not a key factor they consider compared to other factors.

## References

- [1] Pham Phuong Nam, Tran Trong Phuong. Factors affecting commercial housing prices in Vietnam [J]. *International Journal of Housing Markets and Analysis*, 2022, 15(5).
- [2] Koktashev Vladislav, Makeev Vladimir, Peresunko Pavel, Mikhalev Anton, Tynchenko Vadim. Comparison of prices depending on factors in the secondary housing market [J]. *SHS Web of Conferences*, 2021, 116.
- [3] XIA Qiuyue, LU Jie, LIU Chaojie, JIA Ximei. Study on hedonic price of second-hand housing in Zhongyuan district of Zhengzhou city under the background of large data [J]. *Areal Research and Development*, 2020, 39(01): 83-88.
- [4] YANG Dashan, LIU Wei. A study on the influencing factors of second-hand housing transaction prices based on the feature price model—A case study of the second-hand housing market in Shanghai [J]. *The Industrial Study*, 2020(30): 42-45.
- [5] ZHU Jin, XIONG Yang. Research on the factors and development trends of second-hand housing market prices under multi-party mechanisms: based on the transaction data of Chengdu's second-hand housing market [J]. *Public Investment Guide*, 2020(01): 241-242+244.
- [6] TU Jin, JIANG Wanchen, LENG Zhengxing. Differentiation research on influencing factors of urban housing prices in China—Analysis based on big data of second-hand housing market in Chengdu [J]. *Price: Theory and Practice*, 2021(10): 75-78.
- [7] WANG Weihua, YUAN Chiping. Research on the problem of urban housing prices under the background of urban policy—Analysis of the price transmission relationship between new houses and second-hand houses in Shenzhen [J]. *Price: Theory and Practice*, 2019(12): 35-38.
- [8] Xuan Han, Yan Shen, Bo Zhao. Evaluation of quality education premium in housing price—The evidence from Beijing [J]. *Quarterly Journal of Economics*, 2020, 20(05): 257-276.

- [9] SONG Yu, CUI Nana, SHEN Tiyan. Research on the Educational Premium of School District Housing—A Case Study of Beijing [J]. Price: Theory and Practice, 2018(02): 37-40.
- [10] Wei Ha, Huiqin Jin. The Effect of Education Expenditure on Housing Prices in Beijing [J]. Education and Economy, 2018(01): 35-41.
- [11] Wenjing Shi. The price factor analysis of second-hand house based on the method of data mining and LASSO [D]. Beijing University of Technology, 2017.
- [12] Tang Xiaobin, Zhang Rui, Liu Lixin. Research on forecast of second-hand house price in Beijing based on SVR model of bat algorithm [J]. Statistical Research, 2018, 35(11): 71-81.