Dean&Francis
ISSN 2959-6157

The Research on Factors Influencing Housing Price in Shanghai

Ziyi She[1, *]

[1]Oxford International College, Oxford, OX1 3QR, The United Kingdom
*Corresponding author: sheziyiiii@outlook.com

Abstract:

This article aims to identify the factors that may influence housing prices in Shanghai. A multiple linear regression model is used to analyse significant factors based on a sample of 300 from Shanghai in 2021. Under certain assumptions, eleven factors were found to be correlated with housing prices. To assess the data's effectiveness, this study compares the Variance Inflation Factor (VIF) and tolerance levels of these variables. It also considers the incident relations using a scatter diagram and screens for abnormal data through box plots to illustrate the data distribution situation. The findings suggest that loft floor heating and the number of bathrooms may affect housing rental prices; however, the number of bedrooms, living and dining rooms, WIFI, outdoor space area, square meters, use type, and balcony did not show a linear relationship. Overall, fluctuations in housing rental prices in Shanghai can be attributed to the varying influence of these factors.

Keywords: Housing price; multiple linear regression; scatter diagram; box plot.

1. Introduction

With the improvement of life quality, housing price has always been a quite crucial issue for people. National real estate prices have increased hugely from 421.77 billion yuan in 2001 to 10064.6 billion yuan in 2022. The real estate industry experienced rapid development, leading to significant growth in housing prices [1]. Since 2013, the trend for housing prices in various cities has shown an obvious polarisation and brought several negative impacts, such as the risk of housing prices' fluctuation and the monetary policy [2]. The change in housing prices has influenced people's daily lives and daily consumption. However, the specific influencing factors are not familiar to most people. It has essential meaning for people to explore the reasons that can affect the housing price. Therefore, this essay will analyse various kinds of realistic and potential factors to help people understand the housing price according to the research on the housing price all over Shanghai.

The real estate market is made up of distinct parts of factors, including data originating from research on housing price mechanisms, such as housing price and other variables, mainly coming from the department related to the land and management, as well as other management departments [3]. For some realistic factors, Xu used multi-scale geographical weighted regression (MGWR) and Ordinary Least Squares (OLS) regression by stepwise comparison to the model of OLS, Geographically Weighted regression (GWR) and MGWR, and revealed the performance

ISSN 2959-6157

of housing price [4]. The area of houses, the age of houses and the number of bedrooms is examined. The result proved that a single factor in different cities has different influences, and different influencing factors also led to the different housing prices. This method can make the comparison clearer, and the results are more reliable. Tang et. al stated that a city's traffic and basic infrastructure are the main factors that can affect the housing price [5]. They use Point of interest (POI)-based analysis, Exploratory Spatial Data Analysis (ES-DA) technique, and GWS for the study. Compared to the previous model, this model has a larger number of samples, which can give more exact data, predicts the housing price based on the geographical locations, and uses spatial visualisation display.

The researchers Geoffrey and Mark Andrew got the result that the increase of population, rate of interest and the level of income influence housing prices [6]. Deng showed the impact on housing prices of the basic educational resources using a hedonic model [7]. This model can provide a non-linear relationship, and the variable can be changed according to the object. For the potential reasons, Li used Pearson's correlation coefficient to choose the factors that affect housing price and then used the grey model and neural network model to find that Gross Domestic Product (GDP), per capita disposable income and permanent resident population are the influencing factors of housing prices [8]. Li set up an error correction model (ECM) model about the annual data in housing price and land price from 2006 to 2019 in Fuzhou, thinking that the investment in development and completed area have an influence on housing price [9]. Luo et al. used a Panel Data model to analyse the effects of Consumer Price Index (CPI) and GDP. Different cities and different lengths of data can be used to examine this model [10].

This paper mainly focuses on the eight factors (the number of bedrooms, area, age of houses, basic infrastructure, the increase of population, rate of interest, basic educational resources, GDP) to study the impact of these factors affecting housing price and the selection of suitable mathematical models to analyse the relationships among these factors and housing price.

In summary, the investigation of housing prices has garnered considerable attention from numerous scholars. This article primarily employs multiple linear regression models to examine the impact of various factors on Shanghai's housing market prices.

2. Methods

2.1 Data Source

The database utilised in this paper is available from Kaggle website (Shanghai Lane House Rentals 2021). It was collected from Jiazaishanghai by Leslie Cardone. The original dataset is saved in CSV format and contains 2609 groups of data, and this research selected 300 of them as samples.

2.2 Variable Selection

In the original dataset, there was quite a large amount of data, and there were 22 variables, such as latitude and longitude, which were less important to my research. Meanwhile, due to these variables related to some variables that I will mention, and the same data for one variable, this paper chose to remove these variables. Finally, a random sampling is done to get 300 samples. The data contains 11 variables (Distinct, Bedrooms, Living-dining, Bathrooms, Loft, Square meters, Use type, Balcony, WIFI, Outdoor space, Floor-heat) and one dependent variable (Rent). The variable names and explanations of each variable are shown in Table 1.

Table 1. Dependent and independent variables

Variable	Logogram	Explanation
Distinct	X1	Sixteen regions in Shanghai
Bedrooms	X2	The bedroom's number
Living dining	X3	The living dining room's number
WIFI	X4	Whether the WIFI is provided
Outdoor space	X5	The area of outdoor space
Floor-heat	X6	Whether the floor-heat is provided
Square meters	X7	Total housing area
Loft	X8	Whether the house is loft
Bathrooms	X9	The bathroom's number
Use type	X10	Residential (1), multi-use (2)
Balcony	X11	Whether the house has a balcony
Rent	Y	Total housing rent in Shanghai

2.3 Method Introduction

This paper employs a multiple linear regression model to compare housing prices without including these intersection terms. The primary aim is to contrast these two different conditions and assess the accuracy of the results. Ultimately, this will aid in enhancing the model. The multiple linear regression: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \epsilon$.

Multiple linear regression is a linear regression model with two or more variables. It can be used to examine the linear relationship between the dependent variable and various explanatory variables. Its main advantage is that it can control for the other variables and estimate the impact of one explanatory variable. Moreover, the fundamental principle involves calculating the best parameters using Ordinary Least Squares (OLS) by minimising the sum of squared errors between the estimated and actual values.

3. Results and Discussion

3.1 Correlation Analysis

The analysis of Pearson's correlation coefficient gives information about many factors influencing housing prices. As the graph shows:

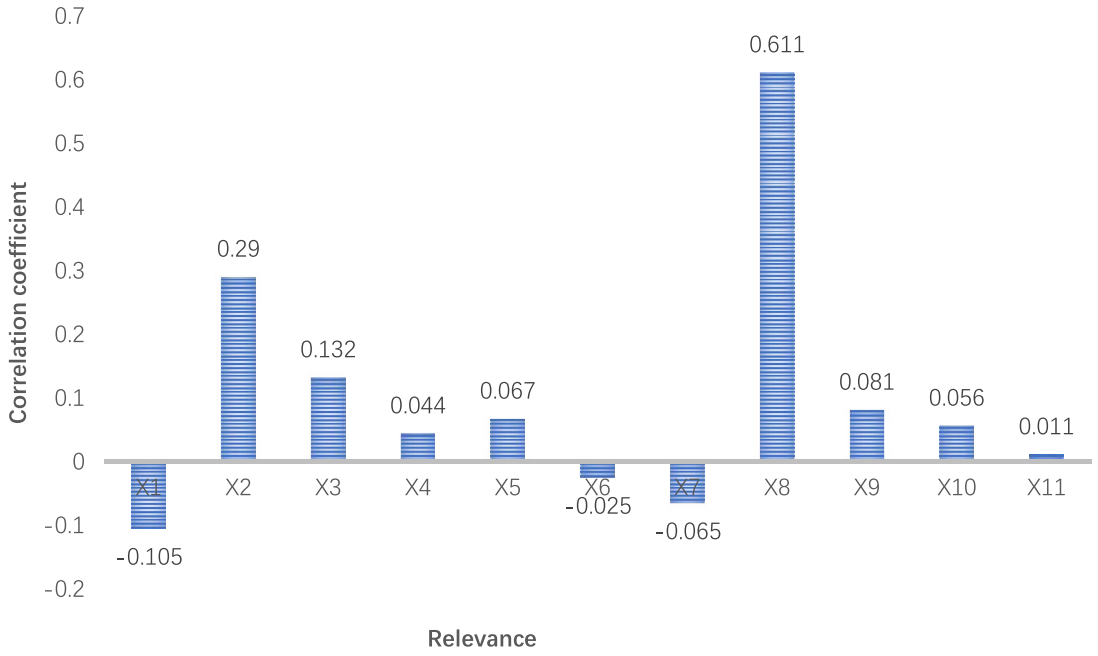

Fig. 1 Relevance Analysis between dependent and independent variables (Picture credit: Original)

From Figure 1, it can be seen that the Pearson correlation coefficient between these factors and housing rental prices. The analysed data showed whether the house is a loft, and the number of bedrooms and living dining rooms are factors that can most positively influence housing prices. There is a significant positive correlation between standardised residuals and housing prices, too. Additionally, whether the Wi-Fi is provided in the house, the size of the outdoor space, the number of bathrooms, the type of use for the house, and whether there is a balcony are positive correlation factors, which are not as obvious as the previous three factors. However, the distinct, whether the house has floor heat, and the number of square meters is negatively correlated with prices in Shanghai. From all the above, the reasons that can reflect the housing prices are multi-dimensional. Thus, people can know that people usually choose to rent a house from different angles nowadays to find their dream house.

3.2 Scatter Diagram

The relationship between X and Y can be directly illustrated using a scatter diagram. As the graph shows below:

ISSN 2959-6157

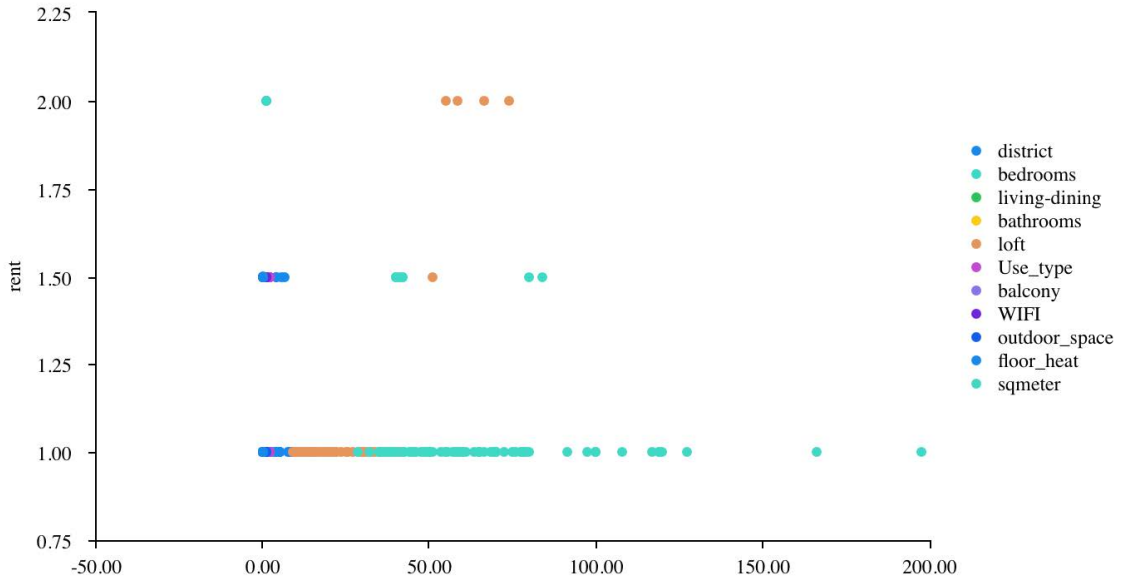

Fig. 2 Incident relations between dependent and independent variables (Picture credit: Original)

From Figure 2, it can be seen that the incident relationship between these variables and housing rental prices is shown using a scatter diagram. The diagram gives the information that the linear trend for the number of bedrooms, the number of living dinings, the use type of the house, whether the house has WIFI and whether the house is a loft is rising. However, there is no linear trend for the factor of whether the house has floor heat, and only two points can be seen to determine the relationship. It is also noticeable that there is a decreasing trend in the distinct of the house, the area of outdoor space, the number of square meters, the number of bathrooms and whether the house has a balcony.

3.3 The Box Plot

The box plots can show the distribution of data in the dataset to explore whether there is abnormal data inside it. As the graph shows below:

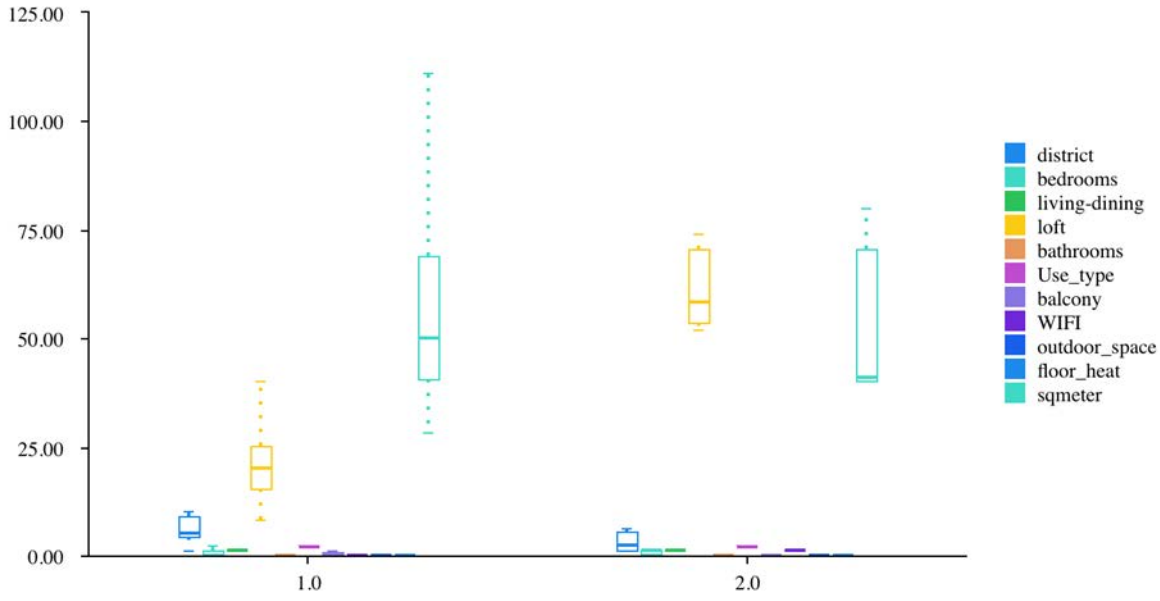

Fig. 3 The box plot between dependent and independent variables (Picture credit: Original)

From Figure 3, the distribution of data can be seen. To find the abnormal data, the exact value of the first quar-

tile (Q1), the third quartile (Q3) and the interquartile range (IQR) should be calculated, which are the 25%, 75% and 75%- 25% of the data. The maximum value is Q3+1.5IQR, and the minimum value is Q1-1.5IQR. The following data demonstrated that there is no abnormal data in the distinct, the number of bedrooms, and whether the house has a balcony and floor heat. However, whether the house is a loft, the number of bedrooms, the use type of the house, whether the house has WIFI, the area of outdoor space and the number of square meters of the house have abnormal data. Thus, these abnormal data should be excluded. After a series of analysis, people know that the housing rental prices should be considered from different angles and multiple regression analysis was conducted.

3.4 Multiple Linear Regression

The multiple linear regression:

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \epsilon \quad (1)$$

In the above formula, β_0 is a constant term, and ϵ is a residual term.

Table 2. Regression coefficient table

	B	S.E.	Beta	T	VIF	Tolerance
Constant	0.519	0.140	-	3.709	-	-
1	0.002	0.006	0.016	0.341	1.093	0.915
2	-0.012	0.031	-0.020	-0.375	1.424	0.702
3	0.076	0.078	0.046	0.971	1.067	0.937
4	-0.026	0.037	-0.034	-0.709	1.112	0.899
5	0.022	0.063	0.017	0.344	1.112	0.900
6	-0.412	0.172	-0.111	-2.397	1.035	0.967
7	-0.000	0.001	-0.040	-0.861	1.049	0.953
8	0.014	0.001	0.636	11.928	1.370	0.730
9	0.140	0.065	0.101	2.163	1.056	0.947
10	0.063	0.049	0.059	1.272	1.023	0.978
11	-0.024	0.035	-0.032	-0.675	1.053	0.949

Table 2 shows the regression coefficients of the multiple linear regression equation model. The p-value of the T-test for the independent variables floor heat, bathrooms, and loft did not exceed 0.05. Therefore, it is obvious that these three variables have a significant impact on the dependent variable Y. IN addition, from the analysed data, all the value of VIF is less than 5. And tolerance is equal to one over VIF, which can also decide multicollinearity when tolerance is less than 0.2. So, the model does not have the situation of multicollinearity. Based on the analysed data above, the relevant multiple linear regression equation can be formed:

$$E(Y) = 0.519 + 0.002x_1 - 0.012x_2 + \ldots - 0.024x_{11} \quad (2)$$

The multivariate correlation coefficient R obtained from this model is 0.635; then, the coefficient R-squared for fitting the above multiple linear regression is 0.403, and the adjusted R-squared is 0.380. Thus, the model can explain 40.3 percentage of the reasons for the changing of housing prices.

4. Conclusion

This research selected 300 samples in the year 2021 in Shanghai from the dataset, which has 11 variables. The method (Multiple linear regression analysis) is accurate, effective, useful and comprehensive. Since this study explores the various factors that can affect housing prices and calculates the relevant Pearson correlation coefficient for every variable.

In the stage of discussion, this paper utilised multiple linear regression to find the most possible relationship between variables and the housing prices. To figure out the more accurate data, the study also conducted some graphs. The scatter diagram is used to show the relationship between the eleven factors. In addition, the box plot is used to determine that whether the data is normal to use in the study of multiple linear regression. Hence, in this study, the number of bathrooms, whether the house is a loft and whether the house has floor heat can influence the housing prices. Among these factors, only the floor heat is negatively correlated with the housing prices in Shanghai. From all of these, whether the house is a loft is the most

ISSN 2959-6157

important factor.

Through this research, people who tend to get a dream house can get a reference from different sights. Hence, they can estimate the budget to rent a house. However, there are still some disadvantages in the research. For example, the number of samples in this research is quite small, and the data is not the latest version, which means there may be some changes for today's people to raise the rental prices. To improve these, more data and the latest version can be collected to explore the relationship between each factor and the housing rental prices.

References

[1] Que Ziyuan. Research on the Spatial and Temporal Differences of Housing Price Influencing Factors in Nanning, Shenyang Architectural University, 2024.

[2] Chen Xiaoliang, Chen Heng, Wang Zhaorui, Xiao Zhengyan. The Research for Influencing Factors of Housing Price Differentiation Among Ci Academic Journal Electronic Publishing Hpublishing house, 2024, 2.

[3] Xue Bing, Xiao Xiao, Li Jing-zhong, Xie Xiao, Ren Wan-xia, Lu Cheng-peng, JIANG Lu. POI-based analysis on the affecting factors of property prices' spatial distribution in the traditional industrial area, Human Geography, 2019, 4.

[4] Xu Jiaqian, The impact of multi-scale locational factors on housing price in major China cities, Shandong University, 2024.

[5] Tang Hongtao, Liu Yipeng, Wu Zhongcai. The Factors on housing price spatial heterogeneity analysis based on POI, China Academic Journal Electronic Publishing House, 2021, 2.

[6] Geoffrey, Mark Andrew. Modelling regional house price: a review of the literature. The centre for the spatial and real estate department of Economics, The University of Reading, 2005, 31-43.

[7] Deng Tingting. Analysis of the Impact of Basic Education Resource Allocation on Housing Prices, Hunan Normal University, 2018.

[8] Li Xiuzhi, Xu Wenjing, et al. Prediction of house price fluctuation based on grey BP neutral network, Journal of Green Science and Technology, 2022.

[9] Li Yun. Research on the Causal Relationship and Influencing Factors between Commercial Housing Prices and Land Prices in Fuzhou, Fujian agricultural and forestry university, 2019.

[10] Luo Xiaoling, Zhou Linjie, Ma Shichang. Spatial Nonuniformity of factors influencing housing prices and different macro-control policies- an empirical research based on panel data model, East China Economic Management, 2014.