

Factors influencing housing prices: A case study of Shanghai

Xinning Fang

Liaoning Experimental Middle School, Shenyang, Liaoning, 110841, China

XinningFang@outlook.com

Abstract. Housing is a major consumer good for residents, and housing prices affect people's quality of life and happiness. This paper analyzes the factors influencing housing prices based on the housing price dataset for Shanghai in 2023-2024. Comparison is made via multiple linear regression model to discover the difference between with and without considering interaction terms, assessing the importance and accuracy of the two models' results, and optimizing the model. According to the analysis, living area, decoration type, and building usage are the factors most positively correlated with housing prices. The number of bathrooms, living rooms, and bedrooms also shows a significant positive correlation with housing prices. Additionally, building type and the construction age of the house are negatively correlated with housing prices, and this correlation is very significant. Finally, the housing area weakens the positive impact of the number of bathrooms and bedrooms on housing prices.

Keywords: Housing prices, multiple linear regression, influence factors, Shanghai.

1. Introduction

Housing prices are crucial to people's life and have always been an important indicator and public concern. Since 2004, housing prices nationwide have been continuously rising: housing price in China experienced unprecedented growth from 2002 to 2010, by nearly 1.5 times and seems to be stabilizing recently [1, 2]. The daily consumption of residents in both China's urban and rural regions were affected by the high private housing prices in the real estate sector. However, the public is not particularly aware of the influencing factors [3]. This paper analyzes the various potential factors that may affect the housing prices in Shanghai, in attempt to help residents with tools to assess and predict the trend of housing prices in Shanghai, and possibly other cities in China.

The real estate sector is a sophisticated regime: a myriad of factors may contribute to the mobile housing prices. Housing price assessment and prediction is also a subject of research academia. Lü claims that area of the houses is significant for analyzing the development of the Chinese real estate market [4]. Yan et al. also used housing price dataset in Beijing, combining the hedonic models, the lasso regression, and the random forests in their analysis, and optimizing their results [5]. They studied such variables as the number of living rooms, bedrooms, and bathrooms on housing prices. The combination of multiple tools increased the accuracy and reliability of their analysis. As for second-hand houses, Zhen et al. suggested that the availability of elevators is an important factor for prices, as shown in their study of Chengdu [6]. Multiple linear regression, decision trees, and extreme gradient boosting models are used to fit the price prediction curve of influencing factors. The XGBoost algorithm stands out, when compared with other models, as the most accurate, widely applicable, and reliable in

data prediction, while reducing overfitting. However, it might slightly lack precision when analyzing results. Domestic scholars believed that the structure and the age of the buildings are among the influencing factors for housing prices [6, 7]. It is claimed by Zheng that the renovation of the houses is a crucial factor leading to differences of their prices [8]. Multiple linear regression and semi-log models are employed in the study, and ordinary least squares (OLS) regression is used to estimate parameters and verify the model's accuracy [9]. However, the paper used a small data set of 219 samples. While studying housing mortgage prices, Fan et al. discovered that property rights significantly affect housing prices [10]. Wang et al. Focuses their study on external factors, using a global regression model to analyze the significant spatial distribution pattern of housing prices, taking their adjacency to subway stations into consideration [11].

In this research, nine variables are assessed for their impact on housing prices and model to study the correlation between these factors and housing prices, including housing area, building type, decoration type, building orientation, housing purpose, building type, number of bathrooms, number of living rooms, and number of bedrooms, using multiple linear regression model, to discover the influencing factor for housing prices in Shanghai.

2. Methodology

2.1. Data source

The dataset used in this study is collected from various websites showing Shanghai housing prices data and spans from 2023 to 2024. The dataset contains 395,001 entries, and 48,234 of these were selected as the samples for this study.

2.2. Variable selection

The original dataset is huge and contains many variables with missing values, such as construction time, building type, and building structure. These variables were removed in this study. Ultimately, random sampling was used to obtain the data. The data includes nine variables (housing area, building type, decoration type, building orientation, housing purpose, number of bathrooms, number of living rooms, and number of bedrooms) and one dependent variable (the housing price). The dataset used is described in Table 1:

Table 1. Variables description.

Variables	Symbols	Meaning
price	Y	House price
sqft	x_1	House area
Building type	x_2	Assigned value of 1 for panel buildings, 0 otherwise
Decoration type	x_3	Assigned value of 1 for luxury decoration, 0 otherwise
Building orientation	x_4	Assigned value of 1 for south-facing, 0 otherwise
Building usage	x_5	Assigned value of 1 for residential use, 0 otherwise
Housing life span	x_6	Housing life span
Restrooms	x_7	Number of restrooms
Living rooms	x_8	Number of living rooms
bedrooms	x_9	Number of bedrooms

2.3. Method introduction

The multiple linear regression model is employed to compare results with and without considering interaction terms, aiming to compare the importance and accuracy of the two models' results. Ultimately, the optimization of the models is conducted.

The multiple linear regression model falls into the category of linear regression models. The multiple explanatory variables used in the analysis enables it to illustrate the linear relationship between the

explanatory variables and the dependent variable [12]. The model estimates multiple parameters through Ordinary Least Squares (OLS), thereby minimizing the sum of squared residuals between the dependent and independent variables.

3. Results and discussion

3.1. Correlation analysis

The analysis in this paper shows that there are many factors influencing housing prices. As shown in the Table 2:

Table 2. Correlation Analysis between Dependent and Independent Variables

	lnsqr	buildtype	decoration	buildingor	buildingus	housinglife
lnprice	1					
lnsqr	0.102***	1				
buildingtype	0.023***	-0.008*	1			
decoration	0.179***	0.173***	0.00400	1		
buildingor	-0.010**	0.092***	0.281***	-0.010**	1	
buildingus	0.283***	0.078***	0.229***	0.032***	0.200***	1
housinglife	0.178***	-0.535***	0.046***	-0.158***	-0.051***	0.136***
restrooms	0.202***	0.648***	0.027***	0.143***	0.018***	-0.020***
livingrooms	0.114***	0.720***	0.046***	0.198***	0.090***	0.103***
bedrooms	0.070***	0.810***	0.048***	0.130***	0.076***	0.049***

The data reveals that living area, decoration type, and building usage are the factors most positively correlated with housing prices. Therefore, it seems that people are particularly interested in the living area, decoration type, and building usage of a house. The number of bathrooms, living rooms, and bedrooms also shows a significant positive correlation with housing prices. Additionally, building type and the age of the house are negatively correlated with housing prices, and this correlation is very significant. In summary, the factors influencing housing prices are comprehensive. Nowadays, people desire a perfect house from various perspectives. After analyzing the Pearson correlation matrix of the factors, a multiple regression analysis is conducted.

3.2. Linear model results

The results of the linear regression are shown in Table 3. The regression coefficients of the multiple linear regression equation model indicate that the area of the house negatively affects its price, while the building type positively affects the price. Houses that are in slab-type apartment buildings are more expensive than those that are not. The decoration type also has a positive impact on housing prices, with luxury-decorated houses being more expensive than non-luxury-decorated ones. The orientation of the house has a significant positive impact on the price as well, with south-facing houses being more expensive than others. The regression coefficient for housing usage is the largest, indicating that residential-use houses are more expensive than non-residential-use houses. The construction age of the house negatively affects the price, but the impact is relatively small. The number of bathrooms, living rooms, and bedrooms positively influences housing prices.

Table 3. Regression Coefficients Table

	FE	RE
	lnprice	lnprice
lnsqrt	-0.060*** (-9.58)	-0.029*** (-4.72)
buildingtype	0.063*** (12.93)	0.046*** (9.63)
decorationtype	0.037*** (23.31)	0.042*** (25.96)
buildingorientation	0.026*** (7.11)	0.018*** (4.92)
buildingusage	1.128 (120.23)	1.030 (115.42)
housinglifespan	-0.001*** (-2.63)	0.002*** (8.22)
restrooms	0.042*** (15.31)	0.057*** (21.01)
livingrooms	0.032*** (15.44)	0.034*** (16.15)
bedrooms	0.015*** (6.75)	0.004* (1.81)
cons	9.761 (398.99)	9.732 (393.99)

3.3. Linear regression with interaction terms

Interactions between some independent variables might also affect housing prices; these interaction terms are known as interaction effects. In fact, the number of bedrooms and bathrooms may be somewhat related to the size of the house, meaning the effect of the number of bedrooms and bathrooms on the house price depends on the housing area. Interaction terms between the number of bedrooms and housing area and between the number of bathrooms and housing area are added to the regression equation. The regression results are shown in Tables 4 and 5.

The interaction term between the number of bedrooms and the housing area (x1x5) is significantly negative, while the coefficient for the number of bedrooms is significantly positive. This indicates that the housing area significantly weakens the positive effect of the number of bedrooms on housing prices. When the housing area is larger, the impact of the number of bedrooms on the price may become secondary, as shown in Table 4.

The interaction term between the number of living rooms and the housing area (x1x8) is significantly positive, while the coefficient for the number of living rooms is significantly positive. This suggests that the housing area significantly weakens the positive effect of the number of living rooms on housing prices. When the housing area is larger, the impact of the number of living rooms on the price diminishes. These results are shown in Table 5.

Table 4. Regression Results for Interaction Term Between Number of Bedrooms and Housing Area

	FE	RE
	lnprice	lnprice
lnsqrt	-0.009 (-1.42)	0.010 (1.47)

Table 4. (continued).

x1x9	-0.001*** (-21.23)	-0.000*** (-15.99)
buildingtype	0.067*** (13.67)	0.049*** (10.27)
decorationtype	0.037*** (23.19)	0.042*** (25.86)
buildingorientation	0.024*** (6.53)	0.017*** (4.51)
buildingusage	1.087 (114.10)	0.999 (109.60)
housinglifespan	-0.001** (-2.14)	0.002*** (8.65)
restrooms	0.069*** (23.13)	0.078*** (25.96)
livingrooms	0.077*** (26.04)	0.067*** (22.76)
bedrooms	0.017*** (7.92)	0.005** (2.49)
cons	9.555 (364.86)	9.574 (360.83)

Table 5. Regression Results for Interaction Term Between Number of Living Rooms and Housing

	FE	RE
	lnprice	lnprice
lnsqr	-0.020*** (-3.19)	0.003 (0.47)
x1x8	-0.000*** (-23.06)	-0.000*** (-18.05)
buildingtype	0.065*** (13.43)	0.048*** (10.12)
decorationtype	0.037*** (23.16)	0.042*** (25.84)
buildingorientation	0.025*** (6.96)	0.018*** (4.82)
buildingusage	1.092 (115.53)	1.000 (110.55)
housinglifespan	-0.001** (-2.54)	0.002*** (8.27)
restrooms	0.075*** (24.58)	0.084*** (27.16)
livingrooms	0.031*** (15.06)	0.033*** (15.76)
bedrooms	0.046*** (17.97)	0.027*** (10.80)

Table 5. (continued).

cons	9.588 (377.06)	9.594 (372.33)
------	-------------------	-------------------

4. Conclusion

Ownership of an apartment or house in a major city is a lifelong pursuit for many Chinese urban residents. The soaring housing prices in megacities like Shanghai and Beijing make it difficult for many to afford. This paper examines over 48,000 data sets of Shanghai's real estate transactions from 2023-2024 to understand the factors driving these rising house prices.

This paper uses a multiple linear regression model to compare results with and without considering interaction terms. 9 variables were selected for the analysis. Results reveals shows that Living area, decoration type, and building usage are the factors most positively correlated with housing prices. Therefore, it seems that people are particularly interested in the living area, decoration type, and building usage of a house. The number of bathrooms, living rooms, and bedrooms also shows a significant positive correlation with housing prices. Additionally, building type and the construction age of the house are negatively correlated with housing prices, and this correlation is very significant. Finally, the housing area weakens the positive impact of the number of bathrooms and bedrooms on housing prices.

References

- [1] Wu Z K, et al. 2007 Analyzing the impact of housing price factors on homebuyer orientation using the priority factors method. *Journal of Tianjin University of Commerce*, 27(3).
- [2] Hu Q 2017 Analysis of housing price factors based on SVAR model. *Time Finance*.
- [3] Yang D F and Zhang Z M 2013 Empirical research on incorporating housing price factors into China's CPI. *Statistics and Information Forum*, 28(3).
- [4] Lü C Y, et al. 2022 Analysis and prediction of housing price influencing factors based on machine learning. *Proceedings of the Third International Symposium on Information Science and Engineering Technology*, 117-121.
- [5] Yan Z Y and Zong L 2020 Spatial prediction of housing prices in Beijing using machine learning algorithms. In the *Proceedings of the 2020 Fourth Conference on High-Performance Computing and Cluster Technology and the Third International Conference on Big Data and Artificial Intelligence (HPCCT & BDAI '20)*. ACM, New York, NY, USA, 64-71.
- [6] Peng Z, Huang Q and Han Y C 2019 Research on the price prediction model of second-hand housing in Chengdu based on XGBoost algorithm: *Proceedings of the 11th International Conference on Advanced Information Technology (ICAIT)*. IEEE.
- [7] Pan J, et al. 2023 Analysis and prediction of second-hand housing prices in Qingdao based on ensemble algorithms. *Advances in Applied Mathematics*, 12(4), 1671-1682.
- [8] Wang X J 2019 Study on the impact of second-hand housing prices in Chongqing. *Journal of Normal University (Natural Science Edition)*, 19(3).
- [9] Zheng Y F 2007 Study on spatial differences of housing prices in different districts of Hangzhou. *Economic Forum*, 20, 32-34.
- [10] Fan G Z, et al. 2022 Housing property rights, collateral, and entrepreneurship: Evidence from China. *Journal of Banking and Finance*.
- [11] Wang N, et al. 2018 Heterogeneity of the impact of major transportation facilities crossing rivers on housing prices: A case study of Nanchang Riverside New Town. *Urban Research Press*, 10, 123-130.
- [12] Yang C G and Li H B 2019 Population migration, changes in housing supply and demand, and regional economic development: An economic analysis of the current "battles for talents" in domestic cities. *Theoretical Research*, 3, 93-98.