

The research on factors influencing housing prices-take Beijing as an example

Zining Zhu

Hengshui Tai Hua Middle School, Hengshui, 053000, China

20150435525@mail.sdufe.edu.cn

Abstract. Housing price is not only related to the economic situation of individuals and families, but also has a profound impact on the whole social economy. For many families, their home is their biggest asset. A rise in house prices can increase household net assets. A fall in house prices could lead to a decline in assets. This paper aims to study the influencing factors of housing prices. The data in this paper take Beijing from May to July 2014 as an example. First of all, the article mentions some papers about housing prices, then the author uses scatter plot and linear regression to fix the model. In the results, it can be seen that sqft-living, sqft-above and view have a relatively high correlation with housing prices. Finally, the specific correlation data table is listed. The final conclusion gives the results of the research in this paper and the authors of the articles mentioned. Housing price is not only related to the economic situation of individuals and families, but also has a profound impact on the whole social economy.

Keywords: Housing prices, multiple linear regression, influencing factors.

1. Introduction

The house price, that is, the real estate price, refers to the transaction price of the house and its attached land use right in a certain time and place. Housing price is a direct reflection of the supply and demand relationship in the real estate market and is affected by many factors, including economic conditions, population growth, land policy, financial environment, market expectations, etc. In China, housing prices are particularly concerned because housing is closely related to the quality of life of residents and the real estate market has a significant impact on the national economy [1]. Five explanatory variables were selected from the two levels of demand and supply, and empirical analysis was carried out through nested models. The results show that the main driving force of housing price increase comes from residents' consumption level, land purchase cost and housing sales area. Finally, the paper puts forward the corresponding countermeasures and suggestions [2].

This paper studies the effects of four real estate regulation policies, including purchase restriction, loan restriction, price restriction and sale restriction, on the prices of commercial housing in 29 major cities in China from February 2016 to May 2018. The results show that the purchase restriction and loan restriction policies have a good regulatory effect on first-tier cities and second-tier cities, while the price restriction and sale restriction policies have a certain effect on first-tier cities, but have no obvious effect on second-tier cities. The study suggests that the policy of limiting purchases should be adhered to and targeted according to the situation of different cities [3]. Based on the historical transaction data of second-hand housing in Beijing from January 2017 to January 2018, the housing price of Beijing is

modeled and analyzed from 15 dimensions. The predictors were screened by multiple linear regression, stepwise selection and Lasso regression [4].

With China's rapid economic growth and steady improvement of urbanization level, China's housing price has maintained a momentum of rapid rise. Clarifying the main factors affecting the real estate price in China will undoubtedly play a fundamental role in stabilizing the housing price and formulating regulatory policies [5]. This paper discusses the measurement methods of different levels of housing price bubbles in China and the driving factors behind them, and also discusses the implementation effects of real estate regulation policies [6]. The autoregressive distributed lag (ARDL) model is used to analyze the impact of real estate policy regulation on housing prices [7]. Based on 15 years' data from 2005 to 2019, Stata software was used to analyze the influencing factors of China's real estate prices, and the influence of factors such as residents' consumption level, land purchase cost and housing sales area on housing prices was discussed [8]. This paper studies the path selection of Chinese real estate enterprises in the process of internationalization. The paper may have analyzed China Real estate enterprises in the process of internationalization in the political, economic, social and cultural differences and other complex investment environment, as well as the competitive situation. This paper aims to build a theoretical model and verify the path selection model through case study, so as to help enterprises achieve stable and powerful development in the process of internationalization [9, 10]. This paper summarizes the existing foreign research on the influencing factors of housing price, and summarizes the literature on the influencing factors of housing price from the two aspects of research methods and research perspectives. The paper may point out that the research of foreign scholars mainly starts from five perspectives: supply and demand, individual factors of consumers, housing characteristics, government policies and spillover effect, and with the emergence of price bubbles in the real estate market, more and more attention is paid to the impact of individual factors of consumers on housing prices. The research methods are mainly empirical, including linear and nonlinear, spatial measurement methods.

In summary, this article will use the multiple linear regression model to study the impact of these 12 factors on Beijing's housing prices.

2. Methodology

2.1. Data source

This dataset summarizes home prices from May 2, 2014 to July 10, 2014, along with individual variables. It contains 59,800 data in 4,600 places and times. This dataset contains 4600 groups of data, and this research selected 400 of them as samples. The original dataset remained in .csv format.

2.2. Variable selection

The original dataset has a very large amount of data, and there are a lot of nulls for variables such as construction time, building type, and many bad values for building structure. At the same time, due to too many Days on Market (DOM) vacancies, and mixed forms of data for floor, this literature chose to remove these variables. Eventually, a random sampling is done to get 400 observations. The data contains 12 variables (Square, Living Room, Drawing Room, Bathroom, Building Type, Elevator, Construction Time, Building Structure, Renovation Condition, Property Rights, Subway, and District) and one dependent variable (Housing Price). The specific description of this dataset is shown in Table 1:

Table 1. Variable introduction.

Variable	Symbol	Range
bedrooms	X1	1.0-9.0
bathrooms	X2	0-8
sqft-living	X3	370-13540
sqft-lot	X4	638-1074218
floors	X5	1-3.5
waterfront	X6	0-1
view	X7	0-4
condition	X8	1-5
sqft-above	X9	370-9640
sqft-basement	X10	0-4820
yr-built	X11	1900-2014
yr-revenovated	X12	0-2014
house price	Y	0-26590000

2.3. Method introduction

The paper uses a multiple linear regression model to compare the situation with and without considering the interaction terms. This section will mainly aim to compare the significance of the two models and the accuracy of the results. Eventually, it will enable the optimized processing of models.

The multiple linear regression model is a linear regression model with multiple explanatory variables. It is used to explain the linear relationship between the explained variable and multiple other explanatory variables. Moreover, its basic principle is to estimate a set of parameters by ordinary least squares (OLS) so that the sum of squares of the residuals between the dependent variables and independent variables is minimized.

3. Results and discussion

3.1. Correlation analysis

The analysis in this paper shows that there are many factors Influencing Housing Prices. As Table 2 shows:

Table 2. Correlation results of house price and other factors.

Variable	Correlation
bedrooms	0.437**
bathrooms	0.288**
sqft-living	0.556**
sqft-lot	0.011
floors	0.276**
waterfront	0.000
view	0.459**
condition	0.267*
sqft-above	0.469**
sqft-basement	0.168

Table 2. (continued)

yr-built	-0.292**
yr-revenovated	0.010
house price	Y

From Table 2, it can be seen that this data set lists 12 factors related to housing prices (variables X1 to X12), and the third column in the table gives the range range of variables from May to July 2014. Therefore, linear regression equation can be used to depict the linear regression image to see the correlation between 12 variables and housing prices.

Linear regression equation is a mathematical model used in statistics to estimate the linear relationship between independent variables (explanatory variables) and dependent variables (response variables). In the context of house price analysis, linear regression equations can be used to predict changes in house prices (dependent variables) based on a series of influencing factors (independent variables).

3.2. Model results

The most commonly used correlation coefficient is the Pearson correlation coefficient, which is usually expressed as r . Its values range from -1 to 1, where -1 means completely negative correlation, 1 means completely positive correlation, and 0 means no linear correlation. The closer the absolute value of the correlation coefficient is to 1, the stronger the correlation is. The closer the absolute value is to 0, the weaker the correlation. Usually, $r > 0.7$ or $r < -0.7$ is considered a strong correlation, $0.3 < |r| < 0.7$ is considered a moderate correlation, and $0.3 < |r|$ is considered a weak correlation. Scatter plots can be used to visually observe the distribution pattern of data points. If the data points show a clear linear trend on the graph, indicating a strong correlation. If the distribution of data points is more random, the correlation is weak. Regression analysis can be performed to obtain the regression coefficient and the goodness of fit (R^2) of the model. The closer R^2 is to 1. The more variation, the model explains, the stronger the correlation. A low R^2 value indicates less variation and weak correlation in model interpretation. The results show that sqft-living, sqft-above, bedrooms, bathrooms and view have strong positive correlation (Table 3).

Table 3. Linear regression analysis results.

	Nonnormalized		Standardization		t	p	Collinearity	
	B	SE	Beta				VIF	tolerance
constant	152884.712	70742.007	-		2.161	0.033*	-	-
sqft_living	188.900	55.813	0.485		3.384	0.001**	3.289	0.304
view	135825.021	35625.542	0.318		3.813	0.000**	1.117	0.895
sqft_above	-13.238	56.825	-0.033		-0.233	0.816	3.283	0.305
R^2					0.400			
adjust R^2					0.382			
F					F (3,96)=21.368,p=0.000			
D-W					1.874			

Y: price

* $p < 0.05$ ** $p < 0.01$

4. Conclusion

This paper uses a variety of methods, aiming to find out several factors with relatively high correlation with housing prices and several factors with relatively low correlation. First, this paper searched the data

of factors related to housing prices from May 2014 to July 2014. Statistical data to find out the range of each factor, in the use of linear regression equations and other mathematical methods to find out the high correlation with housing prices. The results show that there are three factors that have a high correlation with house prices: sqft-living, sqft-above and view. The more mature the surrounding supporting facilities, the more significantly it can push up housing prices. Here, the maturity of the support may be reflected by the number of POI types, that is, the richness of the business format is more positive than the simple number of POI. Housing development investment, housing construction area and housing completed area are significant factors affecting the level of housing prices. The housing price is positively correlated with the investment amount of housing development, but negatively correlated with the construction area and completed area of housing.

References

- [1] Wu Z K, Tang W G and Wu B 2007 Using the Priority Factor Method to Analyze the Impact of House Price Factors on Buyers' Orientation. *Journal of Tianjin University of Commerce*, 27(3).
- [2] Hu Q 2013 Analysis of housing price factors based on the SVAR model. *Times Finance*, 2017.
- [3] Yang Dianxue, Zhang Zhimin. An empirical study on incorporating housing price factors into China's CPI. *Statistics & Information Forum*, 28(3).
- [4] Lv C Y, Liu Y X and Wang L D 2022 Analysis and Forecast of Influencing Factors on House Prices Based on Machine Learning. *Proceedings of 3rd 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 *Proceedings of the 2020 4th High-Performance Computing and Cluster Technologies Conference & 2020 3rd International Conference on Big Data and Artificial Intelligence*. Association for Computing Machinery, New York, NY, USA, 64-71.
- [6] Peng Z, Huang Q and Han Y C 2019 Model Research on Forecast of Second-Hand House Price in Chengdu Based on XGboost Algorithm. *2019 IEEE 11th International Conference on Advanced Infocom Technology (ICAIT)*. IEEE.
- [7] Pan J, et al. 2023 Analysis and prediction of second-hand housing prices in Qingdao based on integrated algorithms. *Advances in Applied Mathematics*, 12(4), 1671-1682.
- [8] Wang X J 2019 Research on the impact of second-hand housing prices in Chongqing. *Journal of Langfang Normal University (Natural Science Edition)*, 19(3).
- [9] Zheng Y F 2007 Research on the spatial difference of housing prices in different urban areas of Hangzhou. *Economic Forum*, 20, 32-34.
- [10] Wang N, et al. 2018 The Heterogeneity of the Impact of Major Transportation Facilities on Residential Prices under Urban Crossing Rivers: A Case Study of Binjiang New City in Nanchang City. *Urban Studies*, 10, 123-130.