Proceedings of CONF-MPCS 2025 Symposium: Leveraging EVs and Machine Learning for Sustainable Energy Demand Management
DOI: 10.54254/2753-8818/2025.GL24068

Research on the Factors Influencing Housing Prices

Suqi Chen

Superintendent Xuecheng Careers International Curriculum Center (Huang Guang Campus), Shanghai, China
suki071019@gmail.com

Abstract: This paper aims to investigate and demonstrate the factors affecting housing prices. The method used to analyze 1,000 samples for determining these factors is multiple linear regression. After preliminary analysis, some less relevant factors were excluded, such as built year, and finally, seven key influencing factors were selected for the study. The research employs various intuitive charts and clear tables to present the results. A linear relationship between housing prices and the following factors was established: square footage, number of bedrooms, neighborhood quality, number of bathrooms, parking space size, house age, and lot size. This allows for a clear understanding of how much each factor contributes to price changes, as well as the individual impact of each variable. Behind the influence of each of these factors, this paper further analyzes other possible practical factors. In conclusion, this study provides useful reference value for real estate investment and home-buying budget planning.

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

1. Introduction

In recent years, housing prices have fluctuated significantly, especially in major cities, attracting widespread attention. High housing prices not only affect residents' quality of life but also pose challenges to economic policies and urban development. Intensified population mobility has led to a continuous rise in housing prices, which remain persistently high [1]. In Beijing, housing price growth has far outpaced wage increases, resulting in widening price disparities [2]. After 2016, housing prices have risen to approximately 3.35 times those in 2000, though the difficulty of purchasing has slightly decreased [3]. However, people's pressure to buy homes has increased, yet they cannot clearly and specifically identify the important factors influencing housing prices. Understanding the factors that affect housing prices is crucial. Therefore, this article aims to help people make rational decisions when buying a home based on potential factors that may lead to changes in housing prices.

Housing prices are a critical indicator of the real estate market, influenced by a multitude of factors. Predictions of housing prices are also a popular argument among academics. Research on housing prices can be approached from multiple perspectives, such as regulatory policies, location, and analysis of specific city characteristics [4]. Chen pointed out that combining statistical analysis methods like the Panel Vector Autoregression (PVAR) model and impulse response allows for a more intuitive examination of the dynamic effects and mechanisms between variables, providing an in-depth analysis of how policy, supply, and demand influence housing prices. This method effectively maintains the robustness of the research results [5]. Zhang employed econometric methods-

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2025.GL24068

specifically, a panel data model-to study housing price issues and focal contradictions in megacities. During the process, he utilized analyses such as stationarity tests, cointegration tests, F-tests, Hausman tests, fixed effects models, and error correction models. The use of monthly panel data from seven megacities lends greater credibility to the analysis results [6]. Xu chose a mixed regression approach to analyze the problem. He conducted specific comparisons of the three assumptions of the dependent variable in traditional linear models-normality, independence, and homoscedasticity-and ultimately derived precise relationships between influencing factors and housing prices [2]. Zhu et al. also opted for multiple linear regression models, integrating various influencing factors into a comprehensive mathematical equation, clearly illustrating the impact magnitude of each factor on housing prices. They conducted multiple data tests and extensive mathematical formulations, using the least squares method to estimate parameters and derive unknown regression coefficients, thereby solidifying the research findings [7]. Yu similarly employed multiple linear regression models and ordinary least squares methods for analysis. By explicitly controlling for other factors that simultaneously affect the dependent variable, he focused on the impact of individual independent variables on the dependent variable, analyzing multiple factors such as listing time, price, housing type, and area [8]. For some internal factors, Lv et al. pointed out that housing area data is of great significance for analyzing and predicting the development of China's real estate market [9]. All have an impact on housing prices. Zheng claimed that the degree of renovation conditions of the house itself is an important factor in price differentiation [10]. This study focuses on eight variables-housing area, building age, construction year, and others-to investigate their impact on housing prices. The aim is to identify an appropriate model to analyze the correlations between these independent variables and housing prices.

In summary, this article will use the multiple linear regression model to study the impact of these 8 factors on China’s housing prices. At the same time, this article will refer to the research methods in the above-mentioned literature to make a relatively objective analysis of the data.

2. Methods

2.1. Data source

The dataset used in this paper is fetched from the Kaggle website (House Price Regression Dataset). This dataset is designed for beginners to practice regression problems, particularly in the context of predicting house prices. It contains 1000 rows, with each row representing a house and various attributes that influence its price.

2.2. Variable selection

The original dataset has a very large amount of data, and additionally, many variables contain missing values, such as property rights, subway access, and district information, as well as numerous incorrect entries in the building structure data. At the same time, a significant portion of the construction year data is missing, so this literature chose to remove these variables. Finally, random sampling was applied to obtain 400 observations. The dataset includes eight independent variables and one dependent variable. The seven variables are: floor area, number of rooms, number of bathrooms, neighborhood quality, house age, garage size, and garden size. The dependent variable is house price. A detailed description of this dataset is shown in Table 1:

DOI: 10.54254/2753-8818/2025.GL24068

Table 1: List of variables

Variable	Logogram	Meaning
Square_Footage	x_1	Total housing area
Num_Bedrooms	x_2	The bedroom's number
Num_Bathrooms	x_3	The bathroom's number
Neighborhood_Quality	x_4	The neighborhood's quality
Age-Built	x_5	The age of the house
Garage_Size	x_6	The size of garage
Lot_Size	x_7	The size of lot
Housing Price	Y	Total Housing prices

2.3. Method introduction

The paper uses a multiple linear regression model that is a statistical method used to analyze the relationship between one dependent variable and several independent variables. Its main advantage is that it can measure how each factor influences the outcome while controlling for other variables. For example, in housing price prediction, it can estimate the impact of factors like house size, location, and age simultaneously. Additionally, this model provides clear numerical results, such as coefficients and p-values, making it easy to interpret. Although it requires assumptions like linearity and no multicollinearity, it remains widely used due to its simplicity and effectiveness in revealing key trends.

3. Results and discussion

3.1. Descriptive analysis

Figure 1 presents a scatter plot illustrating the relationship between house price (dependent variable) and square footage (independent variable). The plot reveals a clear upward trend, indicating that properties with larger floor areas generally command higher prices. This suggests a strong positive correlation between the two variables, as the data points cluster closely around an imaginary diagonal line. Notably, there are very few outliers, reinforcing the significance of property size as a key determinant of housing value. The steepness of the trend line further demonstrates that each additional square foot corresponds to a substantial increase in price, highlighting the market's emphasis on spacious living areas.

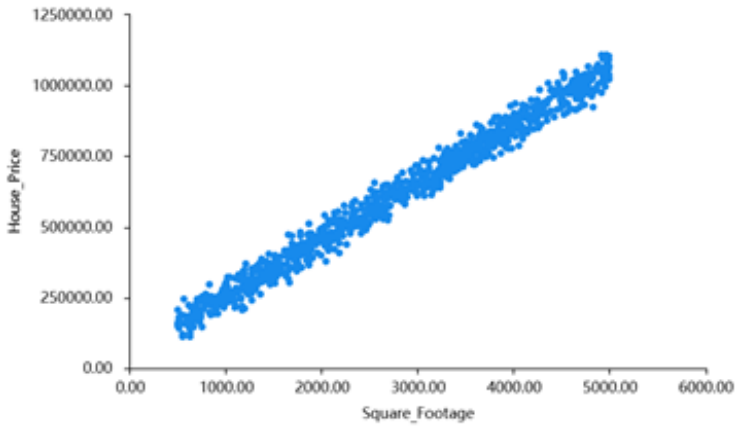

Figure 1: Scatter plot of square footage vs. house price

DOI: 10.54254/2753-8818/2025.GL24068

Figure 2 presents a detailed box plot analysis comparing housing prices across residential properties with one to five bedrooms. The data identifies the most popular property types in the real estate market. Key findings include: Firstly, three-bedroom properties show the highest median price and smallest interquartile range (IQR), indicating strong market consensus on their value. This pricing pattern likely results from balanced demand between small families and property investors. Secondly, two-bedroom units have the lowest median price and consistently narrow IQR, making them the most affordable and stable option. This appeals particularly to first-time buyers and budget-conscious investors. Thirdly, four-bedroom properties demonstrate the widest price range and largest IQR, suggesting Premium pricing for luxury homes, location-based value differences and Varied buyer preferences regarding space usage. The highest-priced four-bedroom properties are typically located in prime areas or feature luxury amenities. This visualization clearly shows how bedroom count serves as an important, though not perfectly linear, predictor of housing value.

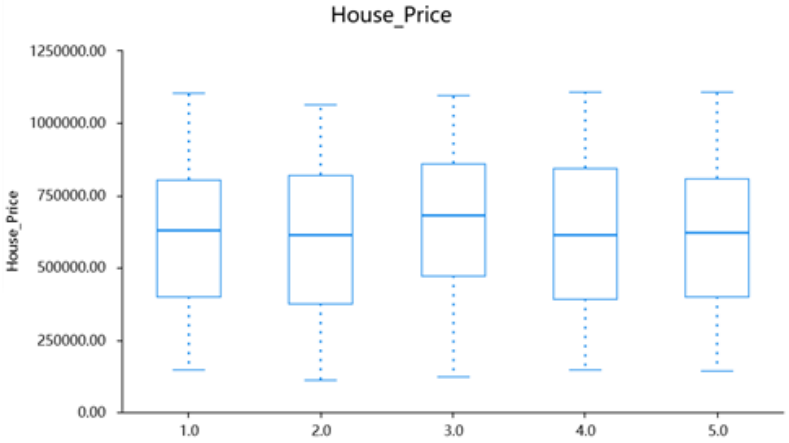

Figure 2: Box chart of number of bedrooms and price

Figure 3 displays a histogram analyzing the relationship between neighborhood quality ratings and housing prices. The data reveals distinct patterns in how residential area classifications affect property values. Key observations from the chart include: Firstly, Premium pricing, which means Properties in grade 10 neighborhoods command the highest housing prices. Secondly, Secondary tier: Medium-quality areas (grade 5) show the second highest price range. Thirdly, Limited variation: Lower quality grades (below 5) have minimal impact on price differences. The histogram clearly illustrates that while superior neighborhood quality correlates with higher property values, the overall impact of community quality on housing prices is not particularly significant, showing only subtle variations.

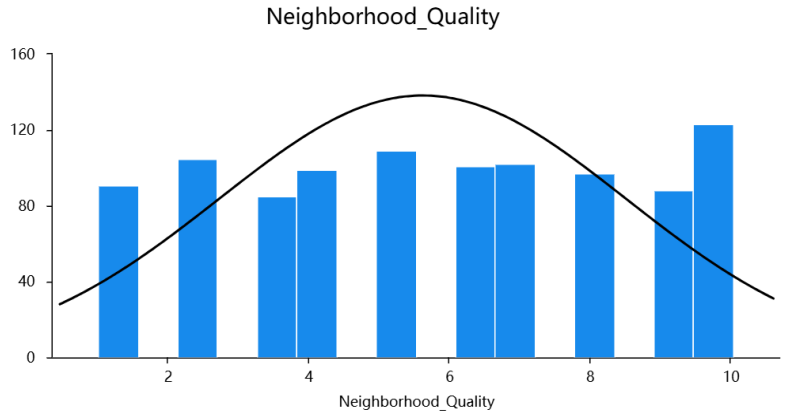

Figure 3: Histogram of neighborhood quality and housing prices

DOI: 10.54254/2753-8818/2025.GL24068

3.2. Correlation analysis

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



Figure 4: Relevance analysis between dependent and independent variables

As shown in Figure 4, the Pearson correlation coefficients reveal the relationships between these factors and housing prices. The study found that square footage and lot size had the strongest positive correlations with price, indicating current buyers prioritize spacious layouts. While the number of bathrooms also showed a positive correlation, its impact was less significant. Besides, factors like the number of bedrooms, neighborhood quality, garage size, and building age exhibited negative correlations. This demonstrates the multidimensional nature of housing preferences, suggesting modern buyers make rational trade-offs across various attributes.

3.3. Model results

Following this correlation analysis, a multiple linear regression was conducted. The general mathematical model is expressed as:

$$E(Y) = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon \quad (1)$$

In the above formula: β_p is the parameter coefficients, and ϵ is a residual term.

Table 2: Regression coefficient table

	B	S.E.	Beta	T	significance	VIF
Constant	735.360	1622.667	-	0.453	0.651	-
X1	199.756	0.249	0.989	803.789	0.000	1.013
X2	10170.955	219.157	0.057	46.409	0.000	1.018
X3	8244.535	378.698	0.027	21.771	0.000	1.004
X4	80.617	107.618	0.001	0.749	0.454	1.004
X5	-991.468	15.065	-0.081	-65.812	0.000	1.005
X6	5158.064	383.350	0.017	13.455	0.000	1.016
X7	14919.323	240.583	0.076	62.013	0.000	1.014

DOI: 10.54254/2753-8818/2025.GL24068

Table 2 presents the regression coefficients of the multiple linear regression model. All six independent variables, X1, X2, X3, X5, X6 and X7 show p-values of the T-test below 0.003, indicating statistically significant effects on the dependent variable (house price). Based on these results, the predictive regression equation can be formulated as:

$$Y = 735.36 + 199.756X_1 + \cdots + 14919.323X_7 \quad (2)$$

4. Conclusion

This study selected 1000 data samples for analysis, covering seven variables. The method is clear and comprehensive, using Pearson correlation coefficients to measure each variable's relationship with housing prices. During the analysis phase, a multiple linear regression model was applied to identify potential connections between these variables and housing prices.

To better understand the impact of individual factors, the study analyzed their effects after deriving the linear regression equation. Various graphs were used to present the results in a clear and accessible way. The findings show that housing prices are influenced by floor area, number of bedrooms, number of bathrooms, house age, garage size, and lot size-with floor area having the strongest effect.

This research helps buyers make more informed budget decisions when purchasing property. However, there are some limitations: the dataset used was simulated, whereas real-world conditions involve more uncertainties, which may introduce bias. To improve accuracy, future studies could incorporate real-world surveys for more detailed and precise conclusions.

References

[1] *Fan, J.N. (2021) Research on the influencing factors of housing prices in Population outflow cities. Dissertation defense paper of the School of Management and Economics of Tianjin University.*

[2] *Xu, Y. (2021) Analysis of Factors Influencing Commercial Housing Prices in Beijing. Economic Management Digest, 14.*

[3] *Gao, Q.D. (2021) Analysis of the characteristics of China's real estate market and factors affecting house prices. An industry-focused journal for Habitat, 25.*

[4] *Tu, J., Jiang, W.C. and Leng, Z.X. (2021) Differentiated Study of Factors Affecting Urban House Price in China (Analysis of Big Data in the Chengdu Primary Housing Market). Journal of Price and Theory Practice, 10.*

[5] *Chen, M.D. (2022) Study on influencing factors and regional differences of housing prices in China. Degree thesis at Liaoning University, 4.*

[6] *Zhang, Y. (2022) Analysis of the causes of the impact of housing prices in the largest cities in China. University of Foreign Economics and Trade Master's Degree Essay, 5.*

[7] *Zhu, Y.R., Qian, Y.H. and Guo, H.P. (2022) Statistical inference of the determinants of house price. Department of Statistics and Management, School of Mathematics and Statistics, Hunan Normal University, 5.*

[8] *Yu, Y.R. (2019) A study of basic factors influencing housing rental prices based on multiple linear regression (using Shanghai housing rentals market as an example). Shanghai University of Finance and Economics, 29.*

[9] *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.*

[10] *Zheng, Y.F. (2007) Research on the spatial difference of housing prices in different urban areas of Hangzhou. Economic Forum, 20, 32-34.*