

Investigation of Influential Factors of Housing Price

Shengjie Zhou^{1,a,*}

¹UC San Diego, San Diego, 92122, USA

a. Shz055@ucsd.edu

*corresponding author

Abstract: The fluctuation of housing prices and its influencing factors are important research topics in economics. Under the current background of rapid globalization and urbanization, understanding the driving factors of housing prices is not only an important contribution to academic research but also has important guiding significance for policymakers and individual home buyers. This paper explores the key factors affecting housing prices, focusing on the impact of variables such as living space, land space, housing landscape, housing conditions, housing quality, and housing construction year on housing prices. This study uses multiple regression analysis to analyze the US housing price data set. The results show that living space, housing landscape, housing quality, and housing construction year have statistically significant effects on housing prices at a significance level of 1%. This study provides an effective reference to real estate developers, policymakers, and individual home buyers, which can help them understand market demand more deeply, make decisions that are more in line with market trends, and promote the dynamic and functional development of the real estate market.

Keywords: Real Estate, Housing Price, Predictive Analytics, Market Insights.

1. Introduction

Housing prices are one of the topics that people cannot do without and pay special attention to in their daily lives. People will mention the high or low housing prices in daily conversations, or require an independent residence when getting married. Housing prices are also an important measure of the economy, which can be seen from the bubble in the Japanese real estate market. From personal experience, when housing prices are high, individuals are under great pressure. People's wages are not enough to support them to buy a house. When housing prices are low, they are often in a period of poor economic conditions. People also do not have enough funds to buy a house. In addition, owning a house is related to personal happiness. Owning a house can bring a sense of belonging. Excessively high housing prices will reduce an individual's disposable income, which is also related to personal happiness. What is even more striking is that housing prices have undergone great changes during the epidemic. According to the analysis of Bernstein et al., the additional influencing factors brought by the epidemic will disappear with the end of the epidemic, but its impact will remain for a long time [1]. Therefore, housing prices affect the living conditions of most people in society. In order to better understand the profound impact of housing prices on society, it is particularly important to delve into the driving factors of housing prices. The research topic of this paper explores the main factors affecting housing prices. The specific question of the research is which factors impose an

influential impact on housing prices. Relying on the Housing dataset, this paper uses quantitative research methods to determine these factors in the Housing dataset. A better understanding of housing prices can help predict future market conditions, thereby providing relevant planning and investment strategies for governments, businesses, and individuals.

2. Literature Review

There are different aspects of research on housing prices. Some researchers study housing prices to reveal the health of the entire property market and the impact of its price changes on the economy. Cerutti et al. studied the relationship between housing finance characteristics and real estate booms [2]. The authors used econometric models to analyze data to reveal the correlation and causal relationship between financial variables and the market. They found the importance of strict housing finance management and supervision because they play an important role in maintaining economic stability and preventing overdevelopment of the housing market and thus triggering financial crises.

Some researchers explore elements of housing prices, such as income, employment rate, and population changes. Some researchers focus on the impact of human behavior and psychology and the impact of people's consumption behavior on housing prices. Kruniawan et al. investigated the factors that influence the home purchase decision of millennials in Indonesia, and the results emphasized the importance of affordability, location, developer reputation, property characteristics, etc. [3] Similarly, San Ong explored the various factors that affect Malaysian housing prices, aiming to use these factors to help stakeholders make informed decisions and formulate corresponding strategies to stabilize Malaysia's real estate market [4]. In the author's opinion, economic conditions, population trends, supply relations, government policies and market sentiment are closely related to price fluctuations in the real estate market.

Some researchers also use housing price research to measure the effectiveness of policies and regulations, which helps to provide effective suggestions for policymaking. Li et al. studied how the housing purchase restriction policies influence the housing prices in the real estate market [5]. Through the analysis of actual data in Beijing suburbs, they found that housing purchase restriction policies can effectively reduce speculative buying and selling activities. This shows that government policies are conducive to helping people in need obtain housing.

Some researchers have kept up with the times and focused on understanding the impact of technological progress and environmental changes on housing prices. Wittowsky et al. found that convenient public transportation and facilities have a positive impact on housing prices [6]. Besides, the authors mentioned that adding new facilities to old residential buildings is associated with rising housing prices. Similarly, Yang et al.'s results highlight the close relationship between the presence of subways and surrounding housing prices. In addition, they suggest that sustainable urban development can benefit from urban planning and transportation policies [7].

Other researchers explore overall trends in the housing market, and they usually do this by comparing housing market conditions in different countries and regions. In the study, the authors considered the economic conditions, cultural characteristics, and policies and regulations of different countries and regions to summarize the general characteristics of the real estate market. Duca et al.'s study used econometric models to analyze and compare the housing price cycles of multiple countries [8]. They believe that policies and regulations are conducive to preventing the negative impact of extreme fluctuations, thereby achieving stability in the real estate market. Fischer et al.'s article reveals the effect of monetary policy in different regions of the United States [8]. Their research results show that the US government should consider the differences in different regions when formulating policies and strive to make the policies have the desired effect. The research on the real estate market covers economics, society, human behavior, policies, technology, and cross-regional aspects, providing a comprehensive reference for people to understand the real estate market.

Although the current research on housing prices has covered many aspects, there are still some key gaps in the research on the real estate market. First, housing price data mostly comes from large cities, and there is a blank in the research on housing prices in rural areas. The real estate market in rural areas is very different from that in cities in many aspects, such as economic development level, population density, infrastructure, etc. Understanding the real estate market in rural areas is crucial to help reach a clear understanding of the real estate market in the whole country.

Second, housing prices may be affected by some non-traditional factors, such as climate change, fire, and advanced technology. At present, the impact of these non-traditional factors is not well understood. Climate change may affect the safety and comfort of the living environment, while fire will directly affect the safety and market reputation of housing. With the advancement of technology, the application of emerging technologies such as smart homes and green buildings may also have an important impact on housing prices. Additionally, the long-term impact of COVID-19 also needs more in-depth research. The complexity of these factors requires more research to reveal the specific impact.

Third, the population structure of many countries has changed greatly. For example, many countries have entered the stage of aging population. People know little about the impact of population aging on housing prices, which is also a topic that needs to be explored. This involves the housing needs and changes of the elderly. Understanding the actual situation in this regard can provide policymakers with more effective information and a more scientific plan to deal with market challenges that may arise in the future. Research in this area will help real estate stakeholders have a deeper understanding of this market and make more informed decisions, which will also help shape a more stable and fairer real estate market. For example, real estate developers can adjust the focus of market strategies and project designs according to the specific needs of different regions and different groups of people. Policymakers can also use the research results to formulate more targeted policies to promote the healthy operation and development of the real estate market.

3. Methodology

This study mainly relies on the Housing dataset. This dataset has 19 variables, among which the variable price is the dependent variable. The detailed description of the variables is shown in Figure 1. The Housing dataset has the variables that are concerned in traditional research, such as the number of rooms, floor space, building age, region, etc., and also includes some non-traditional or relatively subjective factors, such as whether it overlooks the waterfront, the condition of the house, and the building quality rating. Therefore, exploring the variables in this dataset helps individuals to comprehensively evaluate the price of a house.

Variables	Description
price	price of each home sold
bedrooms	Number of bedrooms
bathrooms	Number of bathrooms, where .5 accounts for a room with a toilet but no shower, .75 is a bathroom that contains one sink, one toilet and a shower or a bath
sqft_living	Square footage of the apartments interior living space
sqft_lot	Square footage of the land space
floors	Number of floors
waterfront	0 = not overlooking the waterfront, 1 = overlooking the waterfront
view	An index from 0 to 4 of how good the view of the property was: 0 = No view, 1 = Fair, 2 = Average, 3 = Good, 5 = Excellent
condition	An index from 1 to 5 on the condition of the apartment: 1 = Poor Worn out, 2 = Fair-Badly worn, 3 = Average, 4 = Good, 5 = Very Good
grade	An index from 1 to 13, where 1-3 falls short of building construction and design, 7 has an average level of construction and design, and 11-13 have a high quality level of construction and design.
sqft_above	The square footage of the interior housing space that is above ground level
sqft_basement	The square footage of the interior housing space that is below ground level
yr_built	The year the house was initially built
yr_renovated	The year of the house's last renovation
zipcode	What zipcode area the house is in
lat	Latitude
long	Longitude
sqft_living15	The square footage of interior housing living space for the nearest 15 neighbors
sqft_lot15	The square footage of the land lots of the nearest 15 neighbors

Figure 1: Variable Description Table

This article uses multiple regression to explore the factors that influence housing prices. First, a basic understanding of the dependent variable price is required, including the use of histograms to reflect the overall distribution of variables and the use of five numbers to understand the specific characteristics of variables. Second, a simple understanding of the relationship between the dependent variable and the related dependent variables is required. This relationship can be displayed using scatter plots. With this understanding, which independent variables may have a significant impact on price can be preliminarily identified. By observing the scatter plot, it can be recognized whether the relationship between the two variables is linear or nonlinear. If there is a strong correlation between the two variables, it can provide a basis for subsequent multiple regression analysis. Finally, multiple regression is applied to find influential variables. In multiple regression, all selected independent variables are included in the model, and the degree and direction of their impact on the dependent variable are determined by evaluating the significance and regression coefficient of each independent variable. Multiple regression can not only determine the significant independent variables, but also quantify the magnitude of these effects, thereby providing individuals with a more powerful analytical tool.

4. Result and Discussion

Firstly, it is important to have a basic understanding of the dependent variable, Price. Figure 2 shows the distribution of housing prices in the dataset. The distribution is right-skewed. Most prices are gathered in the lower level, but few are high, pulling the tail of the distribution to the right.

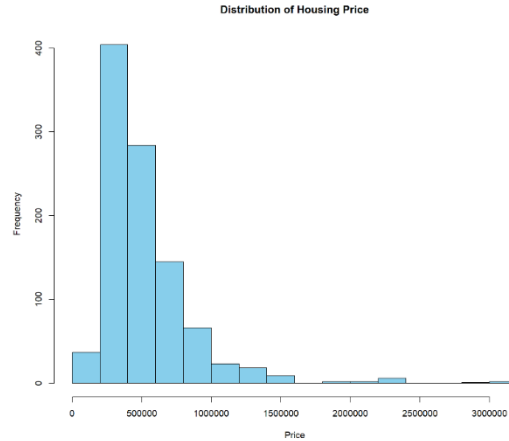


Figure 2: Distribution of Housing Price

Table 1 displays the descriptive statistics of housing prices. The mean is 520646, which is greater than the median of 43500, and it is consistent with the right-skewed distribution. The range, which means that the difference between the most expensive house and the cheapest house, is 2995000. Although there are some outliers in the dependent variable, not deleting the outliers aims to get a model that can reflect different situations.

Table 1: – Descriptive Statistics of Housing Price

Mean	SD	Median	Min	Max	Range	Skew	Kurtosis
520646.5	339579.4	435000	80000	3075000	2995000	2.97	13.84

Secondly, it is necessary to explore the relationship between dependent variable and independent variables. From the dataset, Living Space (sqft_living), Land Space (sqft_lot), Housing View (view), Housing Condition (condition), Housing Quality (grade), and Housing Built Year (yr_built) are selected as independent variables. Among them, Living Space, Land Space, and Housing Built years can be regarded as numeric variables, and others are categorical variables. With the help of R programming, the relationships between Price and other numeric variables are presented in Figure 3.

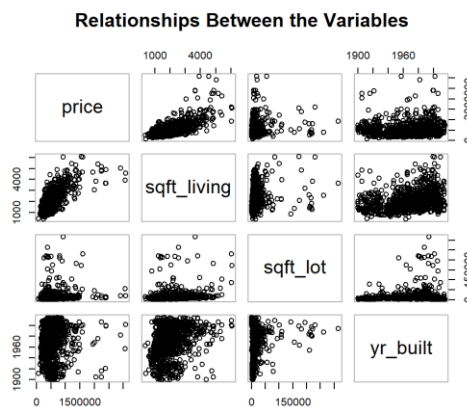


Figure 3: Relationships Between Price and Other Variables

The relationship between Price and Living Space is relatively clear and positive. The relationship between Price and Land Space as well as the relationship between Price and Built Year are not clear

from the scatterplots. Furthermore, correlation is applied to determine the relationships between Price and other interested variables, and the result is shown in Figure 4. The circle size indicates the strength of the correlation. From this figure, the correlations between Price and Living Space, Housing View, and Housing Quality are relatively high.

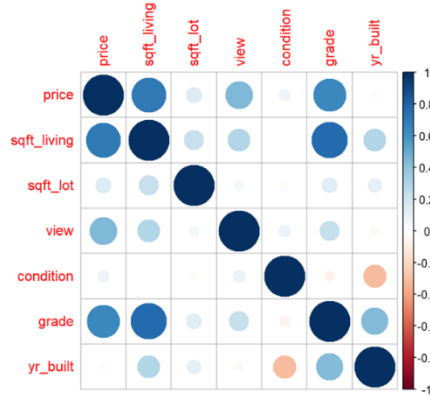


Figure 4: Correlation Between the Variables

Finally, the regression model is used to evaluate the significance of independent variables. Figure 5 shows the best model by using the variables in the Housing dataset. Price is the dependent variable, Living Space (sqft_living), Land Space (sqft_lot), Housing View (view), Housing Condition (condition), Housing Quality (grade), and Housing Built Year (yr_built) are independent variables. Among the six independent variables, four variables, Living Space, Housing View, Housing Quality, and Housing Built Year, are statistically significant at a 1% level. Specifically, Living Space, Housing View, and Housing Quality are positively correlated to Price, while Housing Built Year is negatively related to Price. This means that an older house is more likely to have a lower price. On the contrary, the year a house was built is negatively correlated with its price, which is consistent with the common perception that older houses may be cheaper.

Living space refers to the size of the house, and its coefficient is 158.42, indicating that the price will increase by 158.42 for a one-unit increase in the size of the house. This is because a larger living space can bring a more comfortable living experience. People can get a more comfortable living environment and meet more needs. In contrast, land space did not show significance, indicating that land area is not a significantly important factor. This may reflect that current home buyers pay more attention to living quality rather than the size of the land occupied by the house. Housing landscape and housing quality can be regarded as important representatives of quality of life. Good housing landscape can improve the beauty and comfort of the living environment, which makes home buyers voluntarily pay higher prices. High-quality houses usually mean better building materials and craftsmanship, which makes consumers believe that such houses can have a long service life and higher living comfort. In addition, the coefficient of the variable grade is very high, which means that home buyers will adopt the opinions of third parties when purchasing houses. The higher the rating of the house, the higher the transaction price. This also reflects the attention of new home buyers to non-traditional factors. They are willing to follow the guidance of big data and choose houses that perform well in big data [9].

Dependent variable:	
price	
sqft_living	158.416*** (12.031)
sqft_lot	0.205 (0.231)
view	98,128.190*** (8,973.513)
condition	-4,867.347 (9,982.265)
grade	113,259.900*** (9,451.113)
yr_built	-3,240.113*** (272.363)
Constant	5,704,707.000*** (528,096.800)
observations	999
R2	0.639
Adjusted R2	0.637
Residual Std. Error	204,719.900 (df = 992)
F Statistic	292.326*** (df = 6; 992)
Note: ***p<0.01; **p<0.05; *p<0.1	

Figure 5: Regression Model Output

5. Conclusion

Through the multiple regression analysis of housing prices, people can gain a more comprehensive understanding of the impact of each variable on housing prices and their significance. The results show that living space, housing landscape, housing quality and housing construction year are the key factors affecting housing prices, among which living space, housing landscape and housing quality are positively relevant to housing prices, while housing construction year is negatively correlated. These findings provide effective references for real estate developers, policymakers, and individual buyers. For example, individual buyers can bring the situation of a certain house into the existing multiple regression equation and determine whether to buy the house by comparing the predicted housing price with the actual housing price in the same area.

Although this study provides some valid references, there are still some limitations. On the one hand, as a data set to reveal the national real estate market, the sample size of the data set used in this study is only 1,000. The outliers in the dependent variable were not removed, which led to uncertainty as to whether the existence of outliers made some variables not show significance. On the other hand, different regions and markets have uniqueness and corresponding regulations. These non-traditional factors may affect the applicability of big data analysis and the universality of the results.

This study also provides some insights for future research. The credibility of the research can be enhanced by improving data quality. Battisti et al.'s research reveals the great potential of big data in changing real estate risk management [10]. Through predictive analysis, real-time monitoring, and data integration, companies can enhance relevant risk management strategies and provide operational efficiency and management capabilities. Future research can also focus on the role of policies in the property market. Due to the volatility of the market economy and the cyclical nature of economic crises, policies can play a stabilizing role. How to formulate effective policies will become a matter of great concern. A study of 256 cities in China led by Jia et al. showed that changes in public policies have a significant impact on China's urban expansion, emphasizing the importance of constantly evaluating and adjusting relevant policies to adapt to urban dynamics [11].

References

- [1] Bernstein, J., Tedeschi, Ernie., & Robinson, S. *Housing Prices and Inflation*, by The White House. Sep 9, 2021.

<https://www.whitehouse.gov/cea/written-materials/2021/09/09/housing-prices-and-inflation/#:~:text=Rising%20home%20prices%20have%20direct,important%20role%20in%20overall%20inflation.>

- [2] Cerutti, E., Dagher, J., & Dell'Ariccia, G. *Housing finance and real-estate booms: A cross-country perspective*. [J] *Journal of Housing Economics*, vol. 38, 2017, pp.1-13.
- [3] Kurniawan, C., Dewi, L. C., Maulatsih, W., & Gunadi, W. *Factors influencing housing purchase decisions of millennial generation in Indonesia*. [J] *International Journal of Management*, vol. 11, no. 4, 2020, pp.350-365.
- [4] San Ong, T. *Factors affecting the price of housing in Malaysia*. *Journal of Emerging Issues Economics, Finance and Banking (JEIEFB)*, vol. 1, no. 5, 2013, pp. 414-429.
- [5] Li, Y., Zhu, D., Zhao, J., Zheng, X., & Zhang, L. *Effect of the housing purchase restriction policy on the Real Estate Market: Evidence from a typical suburb of Beijing, China*. [J] *Land Use Policy*, vol. 94, no. 104528, 2020.
- [6] Wittowsky, D., Hoekveld, J., Welsch, J., & Steier, M. *Residential housing prices: impact of housing characteristics, accessibility and neighbouring apartments—a case study of Dortmund, Germany*. [J] *Urban, Planning and Transport Research*, vol. 8, no. 1, 2020, pp.44-70.
- [7] Yang, L., Yu, B., Liang, Y., Lu, Y., & Li, W. *Time-varying and non-linear associations between metro ridership and the built environment*. [J] *Tunnelling and Underground Space Technology*, vol. 132, no. 104931, 2023.
- [8] Duca, J. V., Muellbauer, J., & Murphy, A. *What drives house price cycles? International experience and policy issues*. [J] *Journal of Economic Literature*, vol. 59, no.3, 2021, pp.773-864.
- [9] Fischer, M. M., Huber, F., Pfarrhofer, M., & Staufer-Steinnocher, P. *The dynamic impact of monetary policy on regional housing prices in the United States*. [J] *Real Estate Economics*, vol. 4, no.4, 2021, pp/1039-1068.
- [10] Battisti, E., Shams, S. R., Sakka, G., & Miglietta, N. *Big data and risk management in business processes: implications for corporate real estate*. [J] *Business Process Management Journal*, vol. 26, no.5, 2021, pp.1141-1155.
- [11] Jia, M., Liu, Y., Lieske, S. N., & Chen, T. *Public policy change and its impact on urban expansion: An evaluation of 265 cities in China*. [J] *Land Use Policy*, vol. 97, no.104754, 2020.