

The Research on Factors Influencing Housing Prices-Taking California for Example

Zihan Lu*

Beijing New Oriental Foreign Language School at Yangzhou, Taizhou, 225300, China

*Corresponding author: baochangdi@ldy.edu.rs

Abstract. This article is about something on impacting factors of house prices. In this research, multiple linear regression is used as a method. Regression method is a scientific tool that provides insights into the dynamics of how a dependent X factor responds to changes in one or more independent variables. This research finds that there are many factors and Indies do influence on the housing prices, including GDP, CPI, Income, Urban population, Delinquency rate, etc. Most of those factors or Indies have positive correlation with housing price. And through further analysis selection and analysis, research find that CPI and GDP has positive relation vs housing price and its correlation degree is most highly in those factors. And then GDP and CPI working as independent variables in multi linear regression come to a model. For this model, T test, R-squared test, F test and practical test have been done to proof this model is reasonable. And this model can be used in practical housing price analysis.

Keywords: House price; interaction; linear regression way.

1. Introduction

As an important economic and social issue, housing prices have always attracted much attention. Whether in rapidly developing cities or relatively stable rural areas, the fluctuation of housing prices has a deep influencing on the society to most homes; a house may be not only a place to live, but also a fixed property and an important way to accumulate wealth. Therefore, understanding and analyzing the factors affecting housing prices is not only crucial for academic research, but also has practical significance for government policymakers, real estate developers and home buyers though some policy factors are not quantifiable [1].

In recent days, house prices in China fluctuated significantly, which has triggered extensive discussions and research. One thing, rising or falling housing prices may reflect economic changes and urbanization, but on the other hand, they may also exacerbate social inequality, especially when low income families find it difficult to afford housing. Therefore, this paper has to admit that people's quality of life will change with changes in housing prices [2].

This study will use linear regression analysis to explore the factors that influence housing prices. Line regression is commonly used a tool which could effectively reveal the degree of influence of multiple independent variables on the dependent variable of housing prices. Through the linear regression model, the contribution of each factor to housing prices can be quantified. This method is also very helpful for interpreting data [3].

First, economic factors are one of the important variables that affect housing prices. There is a strong hierarchical relationship between a country's economic development and infrastructure [4]. Generally speaking, economic growth tends to drive up housing prices because economic growth is usually accompanied by an increase in residents' disposable income, which enhances their purchasing power and, in turn, increases need for house. In addition, interest rates are also a key economic factor. Reducing rates of interest mean reducing borrowing costs, allowing more people to afford mortgage loans, driving up housing prices. However, rising interest rates may suppress housing demand and lead to lower housing prices. However, judging the impact of housing prices only by rising and falling interest rates is still inaccurate. It would be more accurate to compare the interest rate elasticity of housing supply and the interest rate elasticity of housing demand [5].

Secondly, demographic factors also have apparent impact on house price fluctuations. Not only does the growth of the registered population have a big influencing on house prices, but floating population also is every important in increasing house prices [6]. Areas with a large net inflow of population will see relatively large fluctuations in housing prices [7]. Take urbanization as an example. Rapid urbanization is usually accompanied by large-scale population migration, especially in developing countries. The increase in urban population often leads to a rapid expansion of housing need, thus driving house prices up.

In addition to economic and demographic factors, that factor of government policies on housing prices cannot be ignored. Although this impact has a lag, the government can influence housing prices through a variety of policy tools, such as land supply policy, tax policy, mortgage loan policy, etc. For example, a shortage of supply of land may cause a rapid increase in price of house, while tax incentives may promote the development of the real estate industry and indirectly affect housing prices [8, 9]. In addition, real estate market regulatory policies, such as limitation on home purchase qualifications and loan interest rate caps, will also have an important impact on the market. Government regulatory measures usually play a key role when housing prices rise or fall too quickly.

In summary, the fluctuation of housing prices is the result of the combined effect of multiple factors. Economic, demographic, and policy factors all play an important role. This study will analyze these influencing factors through linear regression, reveal the degree of influence of each factor, and provide a basis for the formulation of relevant policy. In the following research, this paper will deeply analyze how these factors interact with each other and have a specific impact on price of house.

2. Methods

2.1. Data Source

The data set shown in the paper is abstracted from the web of Kaggle (House Price in some region). It was from Jan. 2011 to Dec. 2022 monthly. And the data include GDP, CPI, Income, Urban population, Delinquency rate, Housing subsidies, Building permits, Interest rate, Mortgage rate, Total house, Construction unit and House for sale or sold. And all those data set are raw data in CSV format.

2.2. Variable Selection

According to desk research, in economic factors, GDP, CPI and Income are chosen to do housing price research as independent variable. In demographic factors, this paper chooses urban population. And as to government policies, this paper chooses housing subsidies, as the following Table 1 shows.

Table 1. Key influencing factors and independent variable in housing price

Influencing factor	Indies
Economic factor	GDP, CPI, Income
Demographic factor	Urban population, Delinquency rate
Government policies	Housing subsidies, Building permits, Interest rate, Mortgage rate
Others	Total house, Construction unit, House for sale or sold

Generally, those factors have the same trend with housing price. That's to say, the house price will be up or down accordingly when those factors show increase or decrease. Here housing price, this paper uses home price index to express. More analysis and proofed will be done in the next chapter to find the actual highly influencing factors from those factors [10].

2.3. Method Introduction

In this research, linear regression analysis method is used. Regression way is a tool that provides depth into the dynamics of how a dependent variable responds to changes in one or more independent variables. It quantifies the influence of these variables, enabling a mathematical assessment of their impact. At its core, a regression analysis model hinges on the principle of minimizing the sum of

squares, a sophisticated mathematical technique for gauging the spread of data points. The objective is to achieve low possibility of those sum data, thereby crafting a figure which best approximates the data.

In the realm of statistics, there is a distinction between simple and multiple linear regressions. Simple linear regression delineates the relationship between a single dependent variable and one independent variable through a linear equation. Conversely, multiple linear regressions extend this concept to encompass more explanatory factors in predicting the one dependent X. Formula is like this,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \quad (1)$$

In this equation, X_i is independent factor. Y is a dependent one. β_i coefficient. And β_0 is intercept and ε is some term. Specific step is as follows,

First, correlation test is to be done. In this step, those low correlation degree factors will be removed. And regression analysis will be seen in variable and independent variable.

Second, if the variable X and independent one relation show linear shape, linear regression analysis will be used in this step. In linear regression analysis, coefficients of intercept and every independent variable will be concluded based on historical data. And then the model has been made.

Third, model test is to be done. In this step, T test will be done to test independent variable significance. R-squared test will be to do test the fitting effect of this formula. F test will be used in testing the whole model significance. And practical test will also be done to ensure the model confirms to objective law. After all those tests, the model is proved to be reasonable and can be applied. In addition, when the data's inherent relationships defy a linear pattern, nonlinear regression becomes the method of choice.

3. Results and discussion

3.1. Descriptive Analysis

Before doing the regressions, Pearson correlation test should be done to identify the key influencing factors to make them be as independent variables. Pearson correlation coefficient is a statistical measure that determines the trend and advantage of the linear relationship in those factors or variables such X and Y. It is in the rage of -1 to 1. And the place -1 indicates a good correlation of negative way. And 1 indicates a good positive way of correlation. As to the 0, that shows nothing in the correlation.

Table 2. Pearson value analysis

Indies	Independent variable	Pearson value
CPI	X1	0.889
GDP	X2	0.838
Housing subsidies	X3	0.829
Income	X4	0.788
Urban population	X5	0.764
Total house	X12	0.720
Building permits	X6	0.381
Construction unit	X11	0.269
Interest rate	X9	0.135
House for sale or sold	X8	0.122
Mortgage rate	X10	-0.219
Delinquency rate	X7	-0.536

As Table 2 shown, CPI GDP, Housing subsidies, Income and Urban population has positive correlation with Home price index. But Pearson value of GDP and CPI is strong, while that of other factors is relatively weak. Thus, this paper should remove the factor of Housing subsidies, income and urban population before linear regression in case col-linearity occurs. Col-linearity occurs when a factor is so highly correlated with one or more of the other factors that make some factors positive correlation turn into negative correlation in linear regression formula. And the col-linearity will lead to the final model weak in explanatory power.

The Pearson value from CPI and GDP is about 0.889(rounded to 3 digits, the same as follows) and about 0.838 respectively. That means CPI and GDP has relatively strong positive correlation with housing price. Figure 1 can also show this correlation and trend accordingly.

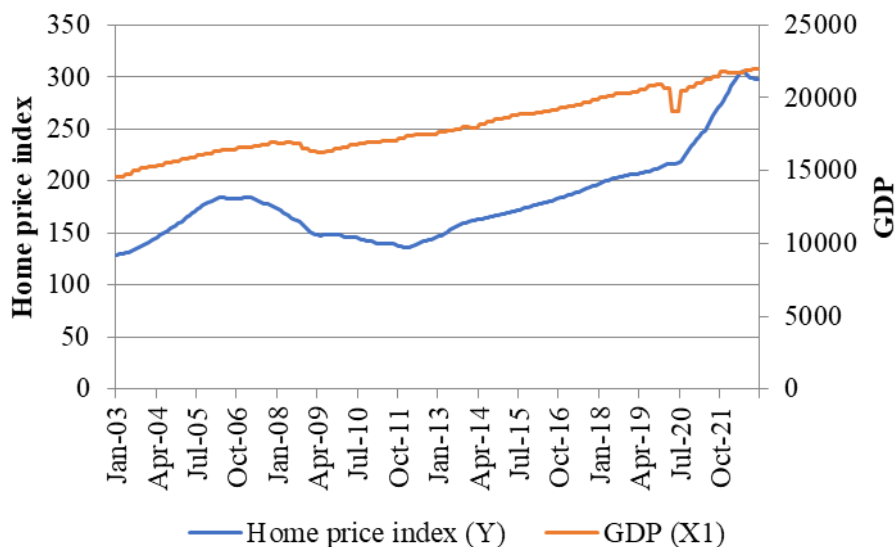


Fig. 1 Home price index and GDP trend

The home price increases or decrease generally according to the same trend of GDP in Figure 1. And the home price also shows the same fluctuation trend as CPI in Figure 2.

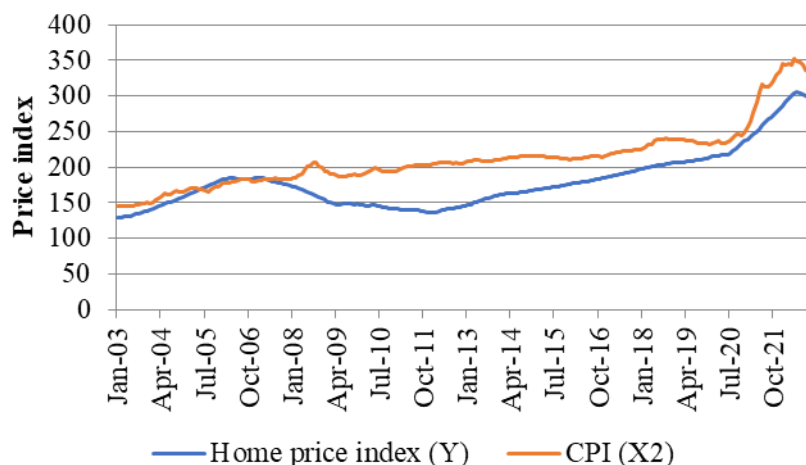


Fig. 2 Home price index and CPI trend

From Pearson value analysis and trend analysis, CPI and GDP actually work as key influencing factors in housing price. Compare with other influencing factors, CPI and GDP has relatively highly correlation. Therefore, CPI and GDP will be chosen to do further research in this article and they will be used to do linear regression analysis independent variables.

3.2. Linear Model Results

After the key factors have been chosen based on Pearson value, the author can come to linear regression analysis. In this section, CPI and GDP work as independent variable X1 and X2 respectively. Home price index works as variable Y. And the model can be concluded as follows through linear regression.

Table 3. Result of linear regression analysis

Item	Coefficients	Standard error	P-value
Intercept	-31.193	14.568	0.033
GDP (X1)	0.0039	0.001	0.006
CPI (X2)	0.665	0.0633	0.000

Based on Table 3, the linear regression analysis comes to the following formula,

$$Y = 0.04X_1 + 0.665X_2 - 31.193 \quad (2)$$

After the model has been created, the coefficient test is widely regarded as significant. When the regression model produces a T-test probability that is below the 0.05 threshold, it signifies a highly significant result. This suggests that the explanatory variables are adept at accounting for the fluctuations in Y, thereby affirming the robustness of the model's construction. In Table 3, the P-values are less than 0.05. And this model is OK based on T test.

R-squared test will be done to compare the degree of match between their evaluation results and the actual occurrences. Typically, several models are tested simultaneously, and the one with a better fit is chosen for trial use. And R-squared value is commonly used as an important index to do this.

The R-squared value, synonymous with the coefficient of determination, serves as a pivotal statistical metric that quantifies the extent to which the variance of a dependent variable is accounted for by the independent variables within a regression model. Spanning a scale from 0 to 1, an elevated R-squared value is indicative of a model's superior alignment with the underlying data. A perfect score of 1 on the R-squared scale denotes an impeccable model that encapsulates the total variance of the dependent variable, whereas a score of 0 implies that the model fails to elucidate any variance at all. In practical application, R-squared value more than 0.7 is generally considered that the model fits well, indicating that the model can explain most of the variability in the dependent variable.

From Table 4, R^2 is about 0.797. And the Adjusted R^2 is about 0.795. Both of them are more than 0.7. That is to say, the model concluded from linear regression is reasonable and it is capable of explaining a significant portion of the variability in the dependent variable. It means that 79.5% of the values fit the regression analysis model. In other words, 79.5% of the dependent variables (y-values) are explained by the independent variables (x-values).

Table 4. Linear regression statistic

Item	Value
Multiple R	0.892
R Square	0.797
Adjusted R Square	0.795
Standard error	18.688
Observed value	240

3.3. Model Test

F test is also done in this research to do further model test. F test is a number test. And it is adopted in doing the test of hypothesis. And that decides whether or not factors, for example, two populations and two factors are the same. The distribution of F is where the information in the test of F facing to. Through separating the factors into two indices, the test makes them comparison by adopting the statistic of F. Relying on the details of the situation one, the test of F can be made or two-made. And

that formula can be considered good based on F test when F value is more than 2 and Significance F is less than 0.05. In Table 5, F value is about 463.895, much more than 2. And Significance F is much less than 0.05. And that indicates that the model is significant.

Table 5. F test index

Item	Freedom	Sum of Squares	Mean Square	F	Significance F
Linear regression	2	324017.133	162008.566	463.895	0
Residual	237	82768.828	349.236	-	-
Total	239	406785.960	-	-	-

4. Conclusion

From this research, influencing factors of housing prices are many, including economic factors, demographic factors, and government police factors, etc. Those factors include many indices such as CPI, GDP, Urban population, Delinquency rate, etc. But not all those indices are key independent variables. Most of those indices work as positive correlation with housing price, while Delinquency rate has negative correlation. In the correlation degree, CPI and GDP are relatively highly. Based on CPI and GDP, multi linear regression comes to a formula as a model for explanation of relationship of CPI, GDP and housing price. Through T test, R-squared test and F test, the model as follows is reasonable.

In practical use, GDP is economic coincident indicator and CPI is economic lagging indicator. This model can be used to evaluate whether housing price runs normal level to support housing regulation making.

References

- [1] Ouyang Xinyu. Nanning Real Estate Price Influencing Factor Linear Regression Model Analysis. China Academic Journal Electronic Publishing House, 2020, 110-112.
- [2] Peng Zhen, Huang Qiang, Han Yincheng. 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, 2019.
- [3] Pan Jia, Luan Yaoyao, Hong Xiaoqing, Li Min. Analysis and prediction of second-hand housing prices in Qingdao based on integrated algorithms. Advances in Applied Mathematics, 2023, 12(4): 1671-1682.
- [4] Wang Xiaojuan. Research on the impact of second-hand housing prices in Chongqing. Journal of Langfang Normal University (Natural Science Edition), 2019, 19(3).
- [5] Zheng Yongfeng. Research on the spatial difference of housing prices in different urban areas of Hangzhou. Economic Forum, 2007, 20: 32-34.
- [6] Lv Chenyue, Liu Yingxin, Wang Lidong. Analysis and Forecast of Influencing Factors on House Prices Based on Machine Learning. Proceedings of 3rd International Symposium on Information Science and Engineering Technology, 2022, 117-121.
- [7] Wang Nan, Wu Wei, Hu Xiying, et al. 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, 2018, 10: 123-130.
- [8] Hu Qiang. Analysis of housing price factors based on the SVAR model. Times Finance, 2017.
- [9] Yang Dianxue, Zhang Zhimin. An empirical study on incorporating housing price factors into China's CPI. Statistics & Information Forum, 2013, 28(3).
- [10] Johansen S. Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. Econometrica, 1991, 59(6): 1551-1580.