Factors Affecting Changes in House Prices:
A Multiple Linear Regression Analysis

Yuehan Chen [1],

Zihan Guo [2, *],

Sijia Song [3]

[1] Wanke Meisha Academy, Guangdong, 518000, China
[2] Jinqiu International, Shandong, Shandong, 266000, China
[3] Wuhan Britain-China School, Wuhan, 430030, China
*2285944547@qq.com

Abstract:

Housing prices, as a key indicator of macroeconomics and livelihoods, are influenced by a variety of economic factors. Exploring the impact of unemployment, building permits, housing subsidies, and mortgage rates on the US housing price index helps shed light on the mechanisms of the real estate market and provide insights for policymaking. This paper, utilizing US housing-related economic data from the Kaggle platform, employs a multivariate linear regression model to investigate the impact of GDP, building permits, housing subsidies, and mortgage rates on the housing price index (HPI). The results showed that in the univariate regression, the unemployment rate had a significant negative impact on the housing price index (R^2=0.256), while building permits and housing subsidies showed significant positive effects, with housing subsidies having the strongest explanatory power (R^2=0.687). Mortgage rates exhibited a significant negative correlation (R^2=0.048). In the multivariate regression, all four variables maintained statistical significance, and the overall model fit was high (R^2=0.945), indicating that the model's predictive power is enhanced when multiple factors are considered.

1. Introduction

As the economic landscape evolves, the real estate sector has shifted toward stable growth, with housing prices fluctuating due to various influencing factors. According to the National Bureau of Statistics of China's Basic Situation of the National Real Estate Market in the First Half of 2025," national real estate development investment
reached 466.58 billion yuan, a year-on-year decrease of 11.2%. According to the National Bureau of Statis-

tics' press conference on the national economic operation for January-February 2025, while housing prices remained generally stable in 2025, some regions faced pressure to stabilize their real estate markets [1]. This marked the end of China's welfare housing policy and an important step in China's real estate development. With the release of the policy, the market demand for housing began to rise dramatically. This change greatly boosted the economy and caused housing prices to rise more and more. Nowadays real estate prices are crucial for economic growth and real

GDP in any country.

The purpose of this paper is to analyze the deeper reasons affecting the fluctuation of real estate prices through geographic location, economic situation and related policies, to promote the healthy and stable development of the real estate industry, so that real estate prices remain within a stable range.

2. Methods

2.1 Data Description

This study analyzes the data obtained from Kaggle platform on different factors and changes in house prices in the United States. The file format of the data is CSV. the author of the data is Jyotsna Gurjar. the publication period is within the last two years and the scope of the data is the whole continent of the U.S.A. explored the effect of GDP, building permits housing subsidies and mortgage rate on the house price, since the model is based on multiple linear regression model with multiple independent variables [2]. Since the dependent variable is house price and there are multiple independent variables, this study is based on multiple linear regression model by using SPSSAU to explore and exclude samples with more than ten percent missing values.

2.2 Model

The core logic of multiple regression is: assuming that there is a linear relationship between the dependent variable y and k independent variables x_1, x_2, ..., x_k, the optimal linear equation is estimated through sample data, so that the error between the predicted value of the dependent variable by the equation and the actual observed value is minimized.

Compared with simple linear regression (with only 1 independent variable) or other univariate analysis methods, the advantages of multiple regression lie in being closer to real-world scenarios" and more accurate causal/correlation inference". Specifically, they can be summarized into the following 4 points:

2.2.1 Control confounding variables and isolate the net effect

In reality, changes in the dependent variable are often influenced by multiple factors, and there may be a confounding relationship" between independent variables (that is, the influence of one variable may be masked or exaggerated by another variable). By incorporating multiple variables into the model, multiple regression can control the interference of other variables, thereby accurately estimating the net impact" of a single independent variable on the dependent variable.

2.2.2 Improve prediction accuracy and reduce errors

For the need to predict the dependent variable through independent variables" (such as predicting housing prices and sales volume), multiple regression can more comprehensively explain the variation of the dependent variable by integrating information from multiple related independent variables, thereby significantly reducing prediction errors and being more accurate than univariate prediction.

2.2.3 Quantify the comprehensive impact of multiple variables to support decision-making

Multiple regression can not only analyze the impact of a single variable, but also quantify the comprehensive impact on the dependent variable when multiple variables act together" through the regression equation, providing a clear quantitative basis for practical decision-making.

2.2.4 Flexible expansion to adapt to complex scenarios

Multiple linear regression is the foundation of the linear model family and can be simply expanded to adapt to more complex data relationships.

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+?+\beta_kX_k+? \qquad (1)$$

Where, y is dependent variable (must be a continuous variable, such as sales volume, housing price, weight"); x_1, x_2, ..., x_k: Independent variables (can be continuous or categorical; categorical variables need to be processed through dummy variables, such as gender, educational background); β_0: Intercept term (the theoretical mean value of the dependent variable y when all independent variables $x_1...x_k$ are 0); β_1, β_2, ..., β_k: Regression coefficients (core parameters), representing when all other independent variables are controlled to remain unchanged, the average number of units by which the dependent variable changes for each 1-unit increase in this independent variable (i.e., the net effect"); ε: Random error term (unobservable random factors, which satisfy the basic assumptions of mean 0, constant variance, and independent and identical distribution, and are the premise for the rationality of the model).

3. Result

Table 1 presents the results of a linear regression analysis with a sample size of 240. The dependent variable is the home price index (HPI), and the independent variable is building permits (BP).

Table 1. Correlation between Unemployment Rates and Housing Price Index

Results of linear regression analysis (n=240)							
Constant	242.334	7.196	-	33.674	0.000**	-	-
Unemployment rate (UR)	-10.259	1.134	-0.506	-9.046	0.000**	1	1
R 2	0.256						
Adjustment R 2	0.253						
F	F (1,238)=81.824,p=0.000						
¥D-W value	0.043						
Note: Dependent variable=me_price_index							
* p<0.05 ** p<0.01							

As can be seen from the above table, UR is taken as the independent variable, while HPI is taken as the dependent variable for linear regression analysis, and as can be seen from the above table, the model formula is: HPI=242.334-10.259*UR, and the model R The square value is 0.256, which means that UR can explain 25.6% of the change in HPI. The F-test of the model found that the model passed the F-test (F=81.824, p=0.000<0.05), which means that UR must have an impact on HPI relationship, the final specific analysis can be seen:

The value of regression coefficient of UR is -10.259 (t=-9.046, p=0.000<0.01), which means that UR will have a significant negative impact on HPI. To summarize the analysis, it can be seen that all of the UR will have a significant negative impact on HPI. This table shows the results of a linear regression analysis with a sample size of 240, where the dependent variable is the HPI, and the independent variable is building permits.

Table 2. Correlation between Building Permits and Housing Price Index

Results of linear regression analysis (n=240)							
	B	Standard Error	Beta	T	P	VIF	Tolerance
Constant	137.757	7.183	-	19.179	0.000**	-	-
BP	0.033	0.005	0.381	6.36	0.000**	1	1
R 2	0.145						
R2 Adjustment	0.142						
F	F (1,238)=40.444,p=0.000						
D-W Value	0.005						
Note: Dependent variable = HPI							
* p<0.05 ** p<0.01							

From Table 2, BP is taken as the independent variable while HPI is taken as the dependent variable for the linear regression analysis, from the above table, the model equation is: HPI = 137.757 + 0.033*BP, and the model R-squared value is 0.145, implying that BP explains 14.5% of the variation in HPI. The F-test of the model found that the model passes the F-test (F=40.444, p=0.000<0.05), which means that BP must have an impact relationship on HPI, which can be seen in the final specific analysis:

The value of regression coefficient of BP is 0.033 (t=6.360, p=0.000<0.01), which means that BP will have a significant positive impact on HPI. To summarize the analysis, it can be seen that all of BP will have a significant positive impact on HPI.

Table 3 presents the results of a linear regression analysis with a sample size of 240, where the dependent variable is the home price index (HPI), and the independent variable is housing subsidies (HS).

Table 3. Correlation between housing subsidies and Housing Price Index

Results of linear regression analysis (n=240)							
	B	Standard Error	Beta	T	P	VIF	Tolerance
Constant	-16.769	8.765	-	-1.913	0.057	-	-
HS	5.693	0.249	0.829	22.858	0.000**	1	1
R 2	0.687						
Adjustment R 2	0.686						
F	F (1,238)=522.506, p=0.000						
Note: Dependent variable= HPI							
* p<0.05 ** p<0.01							

From Table 3, HS is taken as the independent variable, while HPI is taken as the dependent variable for the linear regression analysis, from the above table, we can see that the model formula is: HPI = -16.769 + 5.693*HS, and the R-square value of the model is 0.687, which means that HS can explain 68.7% of the variation in HPI. The F-test of the model found that the model passes the F-test (F=522.506, p=0.000<0.05), which means that HS must have an impact relationship on HPI, which can be seen in the final specific analysis:

The value of regression coefficient of HS is 5.693 (t=22.858, p=0.000<0.01), which means that HS will have a significant positive influence on HPI. To summarize the analysis, it can be seen that all of HS will have a significant positive impact on HPI. Table 4 shows the results of a linear regression analysis with a sample size of 240. The dependent variable is the home price index (HPI), and the independent variable is the mortgage rate (MR).

Table 4. Correlation between Mortgage rate and Housing Price Index

Results of linear regression analysis (n=240)							
	B	Standard Error	Beta	T	P	VIF	Tolerance
Constant	218.677	11.308	-	19.339	0.000**	-	-
MR	-8.118	2.35	-0.219	-3.455	0.001**	1	1
R 2	0.048						
Adjustment R 2	0.044						
F	F (1,238)=11.938,p=0.001						
Note: Dependent variable= HPI							
* p<0.05 ** p<0.01							

From Table 4, it can be seen that MR as the independent variable, while HPI as the dependent variable for linear regression analysis, from the above table can be seen that the model formula is: HPI = 218.677-8.118*MR, and the model R-squared value is 0.048, which It means that MR can explain the 4.8% change in HPI. F-test of the model found that the model passes the F-test (F=11.938, p=0.001<0.05), which means that MR must have an impact on HPI relationship, the final specific analysis can be seen:

The value of regression coefficient of MR is -8.118 (t=-3.455, p=0.001<0.01), which means that MR will have a significant negative impact on HPI. To summarize the analysis, it can be seen that all of MR will have a significant negative impact on HPI.

Table 5 presents the results of a multiple linear regression analysis with a sample size of 240, where the dependent variable is the home price index (HPI), and the independent variables are MR, UR, BP, and HS.

Table 5. Results of linear regression analysis (n=240)

	B	Standard error	Beta	T		VIF	Tolerance
constant	-136.305	9.555	-	-14.265	0.000**	-	-
MR	12.252	0.788	0.330	15.552	0.000**	1.936	0.517

	B	Standard error	Beta	T		VIF	Tolerance
UR	-2.549	0.407	-0.126	-6.262	0.000**	1.734	0.577
BP	0.023	0.002	0.265	13.606	0.000**	1.631	0.613
HS	7.068	0.144	1.029	49.086	0.000**	1.891	0.529
R^2	0.945						
$AjustnmentR^2$	0.944						
F	$F(4,235)=1017.071, p=0.000$						
D-W Value	0.267						
Note dependent variable = HPI							

4. Discussion

4.1 Advantages

This paper conducted linear regression analyses for both single variables (unemployment rate, building permits, housing subsidies, and mortgage rates) and multiple variables (including all four variables). This allows us to explore the impact of both single and combined factors on the housing price index, providing a relatively comprehensive coverage of key economic factors that may influence housing prices. In each regression analysis, it conducted t-tests (on regression coefficients) and F-tests (on the overall model). Most results were significant at the $p < 0.01$ level, strongly indicating a statistically significant correlation between the variables and the housing price index [3].

4.2 Disadvantages

The Durbin-Watson (D-Watson) statistics for some single-variable regressions (e.g., 0.043 in the unemployment rate regression and 0.005 in the building permit regression) were well below 2, indicating strong positive autocorrelation in the residuals. This suggests that the model may have omitted important variables with time-series correlation characteristics or that the functional form may be inaccurately specified, both of which could affect the validity of the regression coefficients and the model's predictive accuracy. Furthermore, this analysis primarily focuses on the presentation of statistical results, with little elaboration on the theoretical mechanisms by which economic factors influence housing prices (for example, how unemployment affects housing demand and, in turn, housing prices) [4]. This lacks an in-depth interpretation of the results from an economic theoretical perspective.

4.3 Suggestions for Improvement

In the future, consideration could be given to introducing lag terms (for example, a first-order lag of the housing price index) to capture the dynamic impact of variables and reduce autocorrelation. Furthermore, consideration could be given to incorporating more potentially influential variables, such as population mobility, household income levels, and real estate inventory, to further enhance the model's explanatory power for the housing price index [5].

Finally, consideration could be given to nonlinear relationships between variables. Testing for nonlinear effects by taking logarithms of variables or setting interaction terms could be used to better align the model with real-world economic principles.

5. Conclusion

In summary, house prices have a direct bi-directional impact on unemployment, mortgage rates, building permits and government subsidies on house prices. This clearly shows that house price is not an independent economic entity, but a core variable deeply embedded in macroeconomic operation, policy regulation and social development. When house price fluctuates dramatically, it not only affects the prospect of social and economic development, but also the wealth of thousands of families and whether they can live and work in peace and happiness, affecting the harmony of urban development. The research methodology used in this article is not very rigorous, as the The SPSSAU tool supports uploading files up to 10 MB in size, with a maximum of 50,000 rows and 1,024 columns.Therefore, in future research, we can use more accurate data to support our viewpoints.

References

[1] National Bureau of Statistics of China. Changes in the sales prices of commercial residential properties in 70 large and medium-sized cities. 2024-10-18. Available: https://www.stats.gov.cn/sj/sjjd/202410/t20241018_1957033.html

[2] Kaggle. Factors influencing US house prices dataset. 2023-

01-01 [Accessed: 2025-09-01]. Available: https://www.kaggle.com/datasets/jyotsnagurjar/factors-influencing-us-house-prices?select=US_House_Price.csv

[3] Madurapperuma W. The dynamic relationship between economic crisis, macroeconomic variables and stock prices in Sri Lanka. Journal of Money and Business, 2023, 3(1): 25-42.

[4] Li N, Li R Y M, Nuttapong J. Factors affecting housing prices in China: a systematic review of papers indexed in Chinese Science Citation Database. Property Management, 2022, 40(5): 780-796.

[5] Marinković S, Džunić M, Marjanović I. Determinants of housing prices: Serbian cities' perspective. Journal of Housing and the Built Environment, 2024, 39(3): 1601-1626.