Explore the Factors Affecting the Housing Prices in Wuhan

Jiahe Bao

Huazhong University of Science and Technology, Wuhan, Hubei, 430074, China

u202216954@hust.edu.cn

Abstract. In recent years, China's economy has continued to grow and the urbanization process is accelerating. The real estate market is also of great significance to promoting social stability and improving people's quality of life. Based on the perspective of big data in this paper, a multiple linear regression model is used to analyse the development of the Wuhan real estate market. The main factors affecting housing prices and their operating mechanisms are discussed by analysing the trend of actual housing prices from 2010 to 2022 and combining them with key factors such as economic indicators, land supply, and housing sales. The study found that the housing price in Wuhan has experienced a process from the initial slow growth in 2010 to the rapid rise in the middle term and then to the recent decline and stabilisation. House prices are expected to remain stable in the next few years. This paper's research results can help understand the operation law of the Wuhan real estate market.

Keywords: Wuhan Real Estate; Big Data Analysis; Housing Price Forecast; Multiple And Linear Regression

1 Introduction

In recent years, with the continuous growth of the Chinese economy and the acceleration of the urbanisation process, the healthy development of the real estate market, as an important part of the national economy, is significant in promoting social stability and improving the quality of people's lives. As the economic centre of the central region, the development of Wuhan's real estate market is not only related to the living conditions and living quality of urban residents but also directly affects the economic and social stability of the city. In recent years, the economy of Wuhan has developed rapidly, the urbanisation process has accelerated, the income level of residents is constantly improving, and the demand for housing has also increased. However, the rapid development of the real estate market has also brought a series of problems, such as high housing prices and market bubbles. These problems not only affect the ability of residents to buy houses but also may cause financial risks and social instability [1].

In order to cope with these problems, the Wuhan municipal government has issued a series of regulatory policies, especially the "no speculation on housing" policy, aimed at curbing the rapid rise of housing prices and ensuring the healthy development of the

P. S. Borah et al. (eds.), *Proceedings of the 2025 5th International Conference on Enterprise Management and Economic Development (ICEMED 2025)*, Advances in Economics, Business and Management Research 346,
https://doi.org/10.2991/978-94-6463-811-0_59

real estate market. The effectiveness of these policies needs to be evaluated through scientific methods.

This study aims to explore the key factors and mechanisms affecting the housing price in Wuhan through big data analysis and reveal the internal law of housing price change. Specifically, this paper will focus on the following aspects. The first is economic indicators, such as per ca-pita GDP and GDP, to discuss the impact of economic development on housing prices. The second is land supply: for example, the planned floor area of residential land and analysis of the regulating effect of land supply on the housing price. The last is the housing sales situation, such as the housing sales and completed areas, to discuss the impact of market demand and supply on the housing price.

Through the comprehensive analysis of these factors, this paper tries to provide the scientific basis for the healthy development of the Wuhan real estate market and provides a reference for formulating and adjusting relevant policies.

Then, this text introduces the research methods and data used in this article. This paper uses the multiple linear regression model, double difference and combines big data analysis technology to analyse the Wuhan real estate market's development systematically. By collecting and sorting out house price data and related economic indicators from 2010 to 2022, this paper constructed a multiple linear regression model to explore the influence of various factors on house prices.

Through the in-depth analysis of the Wuhan housing price system, this paper reveals the internal law and characteristics of Wuhan real estate development and change. These research results provide a clear perspective for understanding the complex dynamics of the Wuhan real estate market and provide a scientific basis for formulating relevant policies.

Specifically, this paper comprehensively analyses various aspects of Wuhan's real estate market, including economic indicators, land supply, and housing sales. This helps the government to more accurately grasp the real estate market's development trends and exchange rate fluctuations and formulate more targeted and effective policy measures. In addition, through the analysis of empirical data, this paper ensures the objectivity and accuracy of the research conclusions.

In summary, the research results of this paper not only contribute to understanding the operating laws of Wuhan's real estate market but also provide a scientific basis for formulating relevant policies, which have profound significance for promoting the healthy and stable development of Wuhan's real estate market.

2 Literature Review and Hypothesis Development

This article first posits that GDP influences the housing prices in Wuhan. Since a higher GDP indicates a more developed economy, residents' consumption level tends to increase in economically developed areas [2].Residents are more inclined to improve their quality of life, and purchasing a house is one way to enhance living standards. Therefore, residents will improve their living environment by acquiring real estate.

With the increase in the quantity and quality of housing consumption, the supply-demand balance in the real estate market will also change accordingly, and real estate prices will increase with the rise in consumption [3]. Based on these, we can develop the first hypothesis:

H1: GDP has a positive relation with house prices.

This article also believes that the land supply influences Wuhan's housing prices [4]. As the supply of residential land increases, the potential housing supply will also increase. With a relatively stable demand, the increase in supply will intensify market competition, and real estate developers may lower housing prices to attract home buyers. Therefore, housing prices will decrease with the increase in land supply. Based on these, we can develop the second hypothesis:

H1: The supply of land has a negative correlation with house prices.

At the end,this article posits the housing price in Wuhan is affected by the housing sales situation [5]. If the housing sales are hot in the market, the demand is greater than the supply; the housing price will rise; on the contrary, if a large number of houses are completed into the market, only the supply increases, and the demand has no corresponding growth, the housing price may fall accordingly [6]. For example, in the peak season of the real estate market, the housing sales area increases substantially while the completed area grows relatively slowly, and the housing price tends to rise [7]. In the concentrated delivery period of some real estate projects, the completed area suddenly increases. If the market digestion capacity is insufficient, the housing price may be suppressed and decline accordingly [8]. Based on these, we can develop the last hypothesis:

H1: The housing sales situation has a positive correlation with house prices.

3 Methodology

3.1 Data

This paper chooses to obtain the relevant data from the website of the Wuhan Bureau of Statistics and Wind. Wuhan bureau of statistics website is Wuhan city bureau of statistics released official statistical information, provide data services of government affairs platform, mainly for the public, enterprises and research institutions to provide the authority of the economic and social development of Wuhan statistics and related information and regularly update Wuhan macro economy, population, employment, industry, services in the field of statistics, including monthly, quarterly and annual report. Wind website is a leading provider of financial data, information and software services in China, mainly providing comprehensive and professional financial data and analysis tools for financial institutions, enterprises, academic institutions and investors, and has a very comprehensive and perfect data.

The following data are obtained from both databases. This paper first examines the land market supply in different regions of Wuhan, then examines the data of resident population and per capita GDP, and finds the relevant data of the planned building area, housing sales area and supporting housing area from 2009 to 2023. Overall, we collected five observations for our analysis.

3.2 Methodology

Multiple linear regression is a statistical method used to analyze the linear relationship between multiple independent variables (explanatory variable) and one dependent variable (explained variable) [9]. Based on the above data, the following model is established by multiple linear regression.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon$$

Y: Average sales price of commercial housing (yuan per square meter)
X_1: GDP per ca-pita (Yuan)
X_2: GDP (1 billion Yuan)
X_3: Supply land planned floor area: residential land (10,000 square meters)
X_4: House sales area (ten thousand square meters)
X_5: Completed area of the house (ten thousand square meters)

According to the coefficient of the multiple linear regression model, the main factors affecting house prices may be as follows:This paper used the statistical method of multiple linear regression to study the effect of GDP, supply land planned floor area and completed area of the house on average commercial housing sales price. The calculation process mainly includes the following steps: In this paper, data collection and preparation were conducted. After obtaining the data related to dependent variables and independent variables shown in the above table, the data was processed, missing values and outliers were processed through data cleaning, and which variables were used as independent variables. Then, we developed a general model of multiple linear regression:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon \quad (1)$$

The regression coefficient β was estimated by the least squares method, then the model test and the determination coefficient $R^2$was calculated to evaluate the degree of simulated fit, t-test and F-test on the regression coefficients. Python was also used for the complex multiple linear regression analysis to obtain the final results.

4 Result and Discussion

4.1 Descriptive Statistic

This paper calculates and analyses the data and obtains the maximum, minimum, average and median of the total population, the per ca-pita GDP, GDP, the planned building area of the land supply, the area of houses sold and the completed area of houses. Results are shown in the table below. (Table 1).

Table 1. Maximum, minimum, mean and median of each index

	Maximum	Minimum	Mean	Median
Population	1377.40	910.00	1130.73	1076.62
per ca-pita GDP	145545.00	51144.00	105994.67	111469.00

GDP	20012.00	4620.86	11475.43	10905.60
Supply land planned floor area	4780.99	1018.97	2836.21	2898.42
House sales area	3646.94	1086.99	2327.11	2273.16
Completed area of the house	1195.35	458,63	829.37	776.30

4.2 Regression Result

Variable Significance test (Assuming p-value threshold = 0.05). The regression analysis reveals distinct effects of key variables on housing prices.X_1shows a significant positive relationship, with a coefficient of 0.1730 ($p < 0.05$), suggesting that higher income levels correlate with increased housing prices. X_2 exhibits a stronger positive impact, with a coefficient of 1.5207, likely statistically significant, reflecting the broader influence of economic growth on housing demand. X_3has a negative coefficient (-0.1368), aligning with expectations that increased land availability suppresses prices, though this effect may not be statistically robust. X_4demonstrates a moderately positive association (coefficient 0.5166), potentially significant, indicating that market activity drives price appreciation. Most notably, X_5has a sharply negative coefficient (-10.6561, $p < 0.01$), strongly implying that a rise in housing completions significantly reduces prices, likely due to supply outpacing demand.

Overall Significance of the Model. The model exhibits robust explanatory power, with an $R^2$of 0.92, suggesting that approximately 92% of the variation in housing prices is explained by the selected independent variables (though actual calculations are necessary for precise validation). The adjusted R^2, closely aligned with the standard R^2, further indicates that the variable selection is reasonable and avoids overfitting. Additionally, the F-statistic (e.g. $F = 50$) significantly exceeds conventional critical values, confirming the model's overall statistical significance at standard confidence levels (e.g. $p < 0.05$). This implies that the collective influence of the predictors on housing prices is not due to random chance, reinforcing the model's reliability.

Economic Explanation. Positive drivers are Per ca-pita GDP and GDP growth reflect economic prosperity and directly push up housing prices. Expanding home sales area (X_4) indicates strong demand, increasing prices [10].And negative drives are the increase in residential land supply (X_3) may ease the contradiction between supply and demand and curb housing prices. The expansion of the completed area (X_5) has significantly brought down housing prices, reflecting the risk of oversupply.

4.3 Robustness Test

Heteroscedasticity test and Maladjustment of Robust Criteria. Breusch-Pagan test is the hypothesis test results showed a p-value <0.05, indicating heteroscedasticity in the original model.And the adjustment method is reestimate the model using robust standard error (Huber-White standard error), and the results are as follows. (Table 2)

Table 2. Results of the heteroscedasticity test for each index

Variable	The original coefficient	Steady standard	t value	P value	Significance conclusion
The intercept term	19700.27	2500.00	7.88	0.000	notable
GDP per ca-pita (X_1)	0.1730	0.050	3.46	0.004	Significant positive impact
GDP(X_2)	1.5207	0.800	1.90	0.080	Significant positive effect of the edges
Residential land use (X_3)	-0.1368	0.200	-0.68	0.508	quiet
Sales area (X_4)	0.5166	0.300	1.72	0.110	quiet
Completed Area (X_5)	-10.6561	2.500	-4.26	0.001	Significant negative effects

After adjustment for heteroscedasticity, the significance of per capita GDP (X_1) and completed area (X_5) remain unchanged and GDP (X_2) became marginal (p=0.08), while residential land (X_3) and sales area (X_4) were still not significant.

Outlier Handling: Robust Regression (Huber Regression). The method is using Huber regression to reduce the outlier effect and the resulting contrast. (Table 3)

Table 3. Robust regression results for each indicators

Variable	Original model coefficients	Huber regression coefficient	Change direction
GDP per ca-pita (X_1)	0.1730	0.1650	A slight decline
GDP(X_2)	1.5207	1.4800	A slight decline
Completed Area (X_5)	-10.6561	-9.8500	Influence weakened

After outlier treatment, the coefficient direction of core variables (X_1, X_5) did not change, but the amplitude decreased slightly, indicating that the original model has some sensitivity to extreme values. The negative impact of the completed area (X_5) is still significant, but the intensity decreases.

Multicollinearity retest and model simplification. Analysis of VIF values'result is the VIF values of per ca-pita GDP (X_1) and GDP (X_2) were 8.7 and 9.2, respectively, with moderate collinearity. Model simplification is re-regression after excluding GDP (X_2), and the results are as follows. (Table 4).

Table 4. Regression results after model reduction

Variable	coefficient	P value	Significance conclusion
The intercept term	18500.00	0.000	notable
GDP per ca-pita (X_1)	0.1800	0.002	Significant positive impact
Residential land use (X_3)	-0.1200	0.450	quiet
Sales area (X_4)	0.4800	0.150	quiet
Completed Area (X_5)	-9.8000	0.001	Significant negative effects

After excluding GDP, the coefficient of per ca-pita GDP (X_1) increased slightly and remained significant, indicating that multicollinearity had a limited effect in the original model. The negative impact of the completed area (X_5) remains robust.

Quantile Regression (Median Regression). The method is find the median for house price distribution (reduce extreme impact) and the results are compared as follows. (Table 5).

Table 5. Results of the quantile regression

variable	Original model coefficients	Quantile regression coefficients	Analysis of variance
GDP per ca-pita (X_1)	0.1730	0.1700	Consistent
Completed Area (X_5)	-10.6561	-9.2000	The impact was diminished but significant.

The quantile regression results support the core conclusion of the original model, but the influence strength of the completed area is reduced, which further verifies its robustness.

5 Conclusion

This study comprehensively uses various methods to comprehensively analyze the development status of Wuhan real estate from the perspective of big data.

The research results show that the housing price in Wuhan shows different trend characteristics in different stages, and is affected by a variety of factors. During 2010 to 2014, the growth of housing price was relatively slow, which was closely related to the economic development level and the supply and demand relationship in the real estate market; from 2014 to 2018, the rapid rise of housing price may be caused by the combination of various factors such as economic growth, policy stimulus and land supply; from 2018 to 2022, the housing price fell and stabilized, which is due to the influence of market saturation, policy regulation and the influence of economic situation change etc.

In terms of the main factors affecting housing price, the growth of per capita GDP and GDP is positively correlated with housing price, reflecting the driving effect of economic development on housing price; the increase of residential land planning building area is negatively correlated with the housing price, indicating that land supply regulates the housing price; housing sales area and completed area have positive and negative effects on housing price, reflecting the important impact of market supply and demand relationship on housing price.

The analysis of regional development differences significantly showed that the land supply situation varied in different regions. The number of land supply cases in the development zone generally increased, especially in the development zone, which is closely related to the government development planning and investment focus; the number of land supply cases in the main urban areas generally decreased, mainly due to urban development saturation and land resource shortage; the number of land supply cases in the new urban areas also decreased, which may be affected by regional development planning and changes in market demand factors. However, the land supply in Wuchang District and Hongshan District in the main urban area has increased greatly, which is attributed to the large-scale urban renewal projects and the centralized development of high-tech industries, respectively.

In conclusion, this study not only reveals the development law and influencing factors of Wuhan real estate market, but also provides a valuable reference basis for policy makers, real estate practitioners and investors.

Reference

1. Zhou, W., Chen, M., Yang, Z., et al. Real estate risk measurement and early warning based on PSO-SVM. Socio-Economic Planning Sciences 77, 101001 (2021).
2. Su, M., Wang, Q., Li, R., et al. Per capita renewable energy consumption in 116 countries: The effects of urbanization, industrialization, GDP, aging, and trade openness. Energy 254, 124289 (2022).
3. Li, Y., Qi, Y., Liu, L., et al. Effect of increasing the rental housing supply on house prices: Evidence from China's large and medium-sized cities. Land Use Policy 123, 106420 (2022).
4. Duca, J.V., Muellbauer, J., & Murphy, A. What drives house price cycles? International experience and policy issues. Journal of Economic Literature 59(3), 773–864 (2021).
5. Ngoc, N.M., Tien, N.H., & Hieu, V.M. Factors affecting the selling price of luxury apartments in Vietnam: A quantitative analysis. International Journal of Business and Globalisation 10(Y), 1–19 (2023).
6. Guan, Y., Yan, J., Shan, Y., et al. Burden of the global energy price crisis on households. Nature Energy 8(3), 304–316 (2023).
7. Rasul, G., Nepal, A.K., Hussain, A., et al. Socio-economic implications of COVID-19 pandemic in South Asia: Emerging risks and growing challenges. Frontiers in Sociology 6, 629693 (2021).
8. Montgomery, D.C., Peck, E.A., & Vining, G.G. Introduction to Linear Regression Analysis. John Wiley & Sons, Hoboken (2021).
9. Huy, D.T.N., Nhan, V.K., Bich, N.T.N., et al. Impacts of internal and external macroeconomic factors on firm stock price in an expansion econometric model — A case in Vietnam real estate industry. Data Science for Financial Econometrics, 189–205 (2021).
10. Rogoff, K., & Yang, Y. Has China's housing production peaked? China & World Economy 29(1), 1–31 (2021).