

Exploring Internal Factors that Affect Real Estates Prices: Evidence from Singapore

Jiatong Jiang

Vanke Meisha Academy, Shenzhen, China

jiangjiatong@stu.vma.edu.cn

Abstract. The purpose of this paper is to analyze the internal factors that generate the difference in real estate prices in Singapore. The objectives are to achieve a deeper understanding of the reason why prices of real estate are different at the same period and to provide a standard for investors to detect real estate with prices that are underestimated and consider whether to invest or not. This paper examines intrinsic factors that mainly affect real estate prices. Regression analysis is the core methodology to analyze the factors; the author used normalization to process the data. The result shows there is a strong relationship between the price of real estate and house age, distance to the nearest MRT station, the number of convenience stores, and the latitude. It can be discovered that the price of real estate negatively correlates to the house age and the distance to the nearest MRT station; positively corresponds to the number of convenience stores and the latitude.

Keywords: multiple regression analysis, real estate prices, P-value.

1. Introduction

In an era with an abundance of uncertainty and risk, rational analysis of potential factors that lead to differentiation in the same industry is required for economic decisions. Real estate is pivotal component in the market. For example, the global real estate market, which used to be valued at \$379.7 trillion in 2022, has reposited almost the greatest wealth in the Global market [1]. Real estate price is a vital economic indicator; it affects people's commercial behaviors, investment decisions, and employment opportunities. Many investors and economists focus on factors that lead to the difference in real estate prices. Moreover, government tends to purchase real estate for their portfolio.

Despite an abundance of external factors, like GDP, inflation rate, exchange rate, interest rate, and crude oil price [2], the geographic position, the age of the residential property, the internal structure of real estate, and many other interior factors are extremely significant. However, there are few researchers who explore the inner reasons why the prices in different places are different. To accomplish the goal of filling the research margin, this paper examines the internal factors that generate the difference in real estate prices.

2. Literature Review

2.1. Former Researchers' Methods

Much research has been conducted to evaluate the factors that affect real estate prices. Regression analysis is one of the most common methods to calculate the relationship between prices and other factors. For example, Guo Jianxiao, Wang Hongli, and Han Hongchen have used a stochastic gradient regression model to analyze influencing factors in real estate prices [3]. Spatial analysis is also a proper method. A spatial model was applied by Hui Zhi in 2019 to investigate whether a region's real estate prices may be affected by prices of the near region [4] Geographically Weighted Regression (GWR), a specialized method of regression analysis, has been used in the analysis of geographic determinants [5]. Another method that can be applied to analyze non-numeral variables is Qualitative Analysis. Qualitative Analysis is for analyzing how government's policies and regulation or a certain event affect the price of real estate. For instance, the "zero threshold" household entry policy has been tested if it correlates to house prices negatively or positively by using qualitative analysis [6]. After

evaluating the feasibility and accuracy, linear regression analysis is utilized as the only method this research adopts.

2.2. Factors from Previous Research

The previous researchers gained the result that the number of convenience stores near the community positively corresponded to the residential property prices [7], but there is little research showing that real estate prices are positively related to the number of convenience stores. Some research results indicated the railway distance insignificantly corresponded to the prices of the commercial estate, compared to other variables, like distance to the hospital or building structures [8]. One study indicates that aging houses will experience depreciation without redevelopment [9]. Another study indicates that real estate prices are unlikely to diminish as the house age is greater when the land is scarce [10]. The accommodation with higher temperature is less likely to be chose by the consumers [11]; therefore, the latitude might associate with the real estate prices. The previous research inspires the author to further explore factors whether the previous research result is right, and to what extent these factors affect real estate prices.

3. Hypothesis

After the house is built, besides the natural erosion, the use and damage done by the owner, lead to the aging of the building. The old-designed structures, appearance, and infrastructures of the real estate are incomparable to the new one, causing trivial things like pests and frequent pipe maintenance requirements. As a result, consumers' decisions to purchase real estate to a certain extent restricted to the age of a house; hence the paper develops hypothesis 1:

H1: the house age is negatively related to the price.

Convenience is one of the crucial criteria for choosing an accommodation or a company. People prefer to live in an environment with well-equipped facilities and convenience stores to reduce the concern about living quality. Then, the hypothesis 2 is:

H2: the number of convenience stores is positively correlated to the real estate price.

Since there has been research demonstrating the insignificant relationship between real estate prices and the distance to the metro station; therefore hypothesis 3 is:

H3: the distance to the metro station does not correlate to the real estate prices.

Since people are unlikely to live in a location with extremely high temperatures, it can be inferred that house prices are relatively low in areas were located near the equator; according to the common sense, hypothesis 4 can be developed:

H4: real estate prices are positively correlated with latitude.

It seems that there are no factors about longitude that affect real estate prices; moreover, the impact that different longitude has on weather, or living conditions is insignificant, so here is hypothesis 5:

H5: real estate prices do not correlate to longitude.

4. Methodology

4.1. Variables

Table 1. Three Scheme comparing

	MIN	MAX	standard deviation	P50	Mean
x_{age}	0	43.8	11.392	16.1	17.713
x_{MRT}	23.383	6488.02	1262.110	492.231	1083.89
x_{store}	0	10	2.946	4	4.094
x_{lat}	24.932	25.0146	0.012	24.971	24.969
x_{lon}	121.474	121.566	0.015	121.539	121.533

x_{age} represents the house age; x_{MRT} shows the distance from the real estate to the nearest MRT station; x_{store} indicates the number of convenience stores around the community. x_{lat} and x_{lon} represent the latitude and the longitude (Table 1). Y is the house price of the unit area. All variables have been normalized by using the formula:

$$(x_n - x_{n \min}) \div (x_{\max} - x_{n \min}) \tag{1}$$

4.2. Assumption

The transaction dates of each Y are slightly different. To control the external factors effect, the transaction dates are assumed to be in one period, economic indicators are fixed; the only variables are the inner differences of each real estate

4.3. Formula

The following formula is applied to calculate:

$$Y = \beta_1 x_{age} + \beta_2 x_{MRT} + \beta_3 x_{store} + \beta_4 x_{lat} + \beta_5 x_{lon} \tag{2}$$

After analyzing the 5 variables, and calculating the P-value, it can be found that there is no significant relationship between variable5 and the prices of real estate; the variable is excluded from the data set, and the following formula is used to measure the coefficient of the rest of variable:

$$Y = \alpha_1 x_{age} + \alpha_2 x_{MRT} + \alpha_3 x_{store} + \alpha_4 x_{lat} \tag{3}$$

One variable showed contradictory results in the first regression analysis and the second one. The variable is deleted from the data set, the following formula is used to measure the final three related variables:

$$Y = \gamma_1 x_{age} + \gamma_2 x_{MRT} + \gamma_3 x_{store} \tag{4}$$

4.4. Tools and Data Source

Excel is the main software to conduct regression analysis. The data are from kaggle, which is six years ago in Singapore [12].

5. Result and Discussion

Table 2. Regression Analysis Result 1.0

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-4945.595	6211.157	-0.796	0.426	-17155.46	7264.269	-17155.46	7264.269
x_{age}	-0.269	0.0389	-6.896	0.000	-0.346	-0.192	-0.346	-0.192
x_{MRT}	-0.004	0.001	-5.888	0.000	-0.006	-0.003	-0.006	-0.003
x_{store}	1.163	0.190	6.114	0.000	0.789	1.537	0.789	1.537
x_{lat}	237.767	44.948	5.290	0.000	149.409	326.126	149.409	326.126
x_{lon}	-7.805	49.149	-0.159	0.874	-104.422	88.811	-104.422	88.811

The R-square in the first regression model is 0.571, which is relatively high in a regression model. Table 2 shows that the P-values of variables 1~4 are smaller than 0.5, which means there is a significant relationship between the real estate prices and these variables. Variables 1 and 2, which are house age and the distance to MRT show a negative correlation. Variable 1 with the coefficient of -0.269, shows a stronger correlation with the prices, compared to -0.004; Variable 3 and 4, which are the number of convenient stores and the latitude, show a positive correlation. However, the P-value of variable 5, the longitude is approximately 0.876, which is far higher than 0.5 and indicates an insignificant relationship with real estate prices. Therefore hypothesis 5 is verified because that longitude only affects the time zone, which may not correlate to real estate prices. In the next test, the variable of longitude is excluded.

Table 3. Regression Analysis Result 2.0

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-9539.623	12237.590	-0.780	0.436	-33596.045	14516.799	-33596.045	14516.799
x_{age}	-11.132	1.759	-6.329	0.000	-14.589	-7.674	-14.589	-7.674
x_{MRT}	-37.375	4.429	-8.438	0.000	-46.082	-28.668	-46.082	-28.668
x_{store}	12.861	1.949	6.598	0.000	9.029	16.693	9.029	16.693
x_{lat}	-39.439	50.367	-0.783	0.434	-138.450	59.576	-138.450	59.572

After excluding Variable 5, the result which shows in Table 3 above, is similar to the result that is analyzed in Table 2 and Table 3. The R-square is around 0.541. The p-value of each variable is even smaller. Still, house age and the distance to MRT negatively correlate with real estate prices. the number of convenience stores shows a positive correlation with real estate prices. However, the coefficient of latitude is a negative number, which contradicts the previous analysis. Moreover, the P-value is very close to 0.5, indicating that the relationship is not very strong. Therefore hypothesis 4 is invalid. A former researcher has explored the relationship between latitude and house prices. The result shows that in the range of 30.2° to 39.0°, the house price is positively correlated to the latitude; however, in the range of 39.0° to 39.7°, there is a negative correlation [13]. It can be the reason why the P-value is not very low in this simple regression model.

Table 4. Regression Analysis Result 3.0

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	42.852	1.378	31.108	0.000	40.144	45.559	40.144	45.559
x_{age}	-11.075	1.757	-6.3045	0.000	-14.528	-7.622	-14.528	-7.622
x_{MRT}	-34.774	2.929	-11.874	0.000	-40.531	-29.017	-40.531	-29.017
x_{store}	12.97	1.943	6.678	0.000	9.155	16.794	9.155	16.794

Excluding the Variable of latitude and longitude, it can be observed from Table 4 that the P-value is far below 0.5. The R-square is approximately 0.541, showing that the degree of fitting is improved and in a reasonable range. Other figures are normal and stable, there is no contrary. The house age is negatively related to the real estate prices, with a coefficient of -11.075; Hypothesis 1 is verified, due to the maintenance and aging problems. The distance to the MRT station is negatively correlated to the price, with an even more significant effect than the variable of house age. This result is contradictory to the former research result. One possible reason may be that the accessibility to transportation is more demanded by residents. Hypothesis 3 is not verified. The number of convenience stores is positively related to house prices, for its convenience to purchase commodities. so, it can be proven that Hypothesis 2 is true (Table 4).

6. Conclusion

To conclude, by using a popular method, regression analysis, the author estimates the inner factors that affect real estate prices. The research result shows the price of real estate has a negative relationship with the house age and the distance to the nearest MRT station and has a positive relationship with the number of convenience stores. The relationship between real estate prices with latitude or longitude is insignificant or uncertain.

Investors can concentrate on the four factors and detect underestimated real estate property, making smarter and more rational economic decisions. As for government's perspective, it will be beneficial for it to consider the distribution of its portfolio and stabilize the real estate prices.

There are a few limitations in the research. The first is about the date. It is extremely hard to collect all house prices of the unit area at the same moment. The real estate prices might be impacted by external factors. There may be small errors. Secondly, the data set provides only a small amount of data for the researchers to conduct regression analysis. Finally, there are substantial factors that lead to the fluctuation of real estate prices; It is almost impossible to consider all of them. Moreover,

merely fundamental regression analysis is applied, disregarding robustness test. Moving forward, the research will be further expanded by utilizing varies of more complicated models to prove the result.

References

- [1] Tang Maohua, Yi Yonggang. Real Estate Commodity Attributes and Value Speculation in Valuation Practice [J/OL]. *Price Theory and Practice*, 1-4 [2400-08-28] <https://doi.org/10.19851/j.cnki.CN11-1010/F.2024.04.185>.
- [2] Alkali, Musa Abubakar, Ibrahim Sipan, and Muhammad Najib Razali. "An overview of macro-economic determinants of real estate price in Nigeria." *International Journal of Engineering & Technology* 7.3.30 (2018): 484-488.
- [3] Jianxiao, Guo, Wang Hongli, and Han Hongchen. "Analysis of influencing factors in real estate prices based on stochastic gradient regression model." 2009 16th International Conference on Industrial Engineering and Engineering Management. IEEE, 2009.
- [4] Zhi, Hui. *An Enquiry into the Chinese Housing Market: Evidence from Spatial Analysis and Real Options*. Diss. University of Portsmouth, 2019.
- [5] Sisman, Suleyman, and Arif Cagdas Aydinoglu. "A modelling approach with geographically weighted regression methods for determining geographic variation and influencing factors in housing price: A case in Istanbul." *Land use policy* 119 (2022): 106183.
- [6] Cui, Faqiang. "Quantitative study on factors affecting the price of residential real estate multiple linear regression model." *Journal of Physics: Conference Series*. Vol. 1629. No. 1. IOP Publishing, 2020.
- [7] Chiang, Ying-Hui, Ti-Ching Peng, and Chin-Oh Chang. "The nonlinear effect of convenience stores on residential property prices: A case study of Taipei, Taiwan." *Habitat International* 46 (2015): 82-90.
- [8] Berawi, Mohammed Ali, et al. "Impact of rail transit station proximity to commercial property prices: utilizing big data in urban real estate." *Journal of Big Data* 7 (2020): 1-17.
- [9] Wang, Wen-Kai, and I-Chun Tsai. "House age and housing prices: a viewpoint of the optimal time for land redevelopment." *International Journal of Strategic Property Management* 26.3 (2022): 172-187.
- [10] Lee, Bun Song, Eui-Chul Chung, and Yong Hyun Kim. "Dwelling age, redevelopment, and housing prices: The case of apartment complexes in Seoul." *The Journal of Real Estate Finance and Economics* 30 (2005): 55-80.
- [11] Ma, Li, and Yildiray Yildirim. "High temperature, climate change and real estate prices." *Climate Change and Real Estate Prices* (June 2, 2023).
- [12] BRUCE, ALGOR. "Real Estate Price Prediction." Kaggle: Your Machine Learning and Data Science Community, kaggle, 2018, www.kaggle.com/datasets/quantbruce/real-estate-price-prediction. Accessed 10 Aug. 2024.
- [13] Oladunni, Timothy, and Sharad Sharma. "Spatial dependency and hedonic housing regression model." 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2016.