Research of the Influence Factors of Housing Price - Take Singapore as an Example

Zirui Liu¹, Yixiao Pan^{2,a,*}

¹Damai Secondary School, Singapore, 479229, Singapore ²Zhenghia High school of Ningbo city, Ningbo, 315000, China a. chenjie@zhzx.net.cn *corresponding author

Abstract: This paper is to identify the influencing factors which impacts the house resale price in Singapore. 3000 samples from 2022 to 2023 had been selected from the data set. The relationship between 6 variables and the resale price of houses is analysed by the Multiple Linear Regression model. The correlation analysis and VIF value check are introduced, aimed to detect insignificant or highly correlated factors. To eliminate the potential interaction effect, the correlation is carried out. The result shows that the remaining lease and the commence lease are highly correlated to each other. One of them is eliminated in the improved model. By doing VIF check, the result shows that the flat model is an insignificant factor which will not influence the resale prices of houses. The floor area, flat type, also known as room allocation, storey, remaining lease and commencing lease of houses are significant factor, and the commencing lease of the houses, the correlated factor are eliminated. The final model explains how the four factors, floor area, room allocation, storey and remaining lease can influence the resale prices of houses in Singapore.

Keywords: Multiple linear regression, housing price, factor analysis.

1. Introduction

Global housing prices have risen steadily for years, and this pushes up the cost of living for many people around the world [1]. In Singapore, the housing prices have experienced significant fluctuations. From 2006 to 2012, both the Housing Development Board (HDB) flats and the private residential property prices rose steeply before stabilizing [2]. From 1980 to 2015, the average annual housing prices in Singapore for the private sector has a growth rate of 7.02% while for the secondary public housing sector is 6.81% [3]. Since 2020, prices started to climb sharply once again, due in part to the COVID-19 pandemic [4]. As a result, the rising prices directly affect the expenses of individuals and families, making it more challenging to meet their housing needs. However, the public does not realize the influencing factors of housing prices. Hence, this article aims to use the Singapore housing price research to help people assess the expected purchase of a house based on the different underlying factors that may contribute to the house price.

Some researchers are studying the factors influencing housing prices. To track the price reaction of existing homes to the quantity of new units introduced by Singaporean home builders, Joseph and Thao used Vector Autoregression (VAR) models. They discovered that marginal supply Granger-

 $[\]bigcirc$ 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

cause current home prices favorably, defying the negative reaction predicted by the "competition" concept [5]. However, the data that they used was from 1996 to 2009. The result may be out of date, which will lower the accuracy of the results. Bian et al. applied econometric analyses and machine learning approaches, using a hedonic model, least absolute shrinkage and selection operator (LASSO), random forest, and artificial neural networks to get deeper insights into the importance of determinants of property prices [6]. They find that property prices are mostly affected by key macroeconomic factors, such as the time of sale, the size, and the floor level of the property. The machine learning approaches that they used can give them more accurate predictions.

Similarly, Cheng and Liu analyzed the effects of macroeconomic factors, supply factors, and alternative housing prices on housing prices in Singapore [7]. They used the two-stage least squares method to estimate the regression equations for the public resale market, condominium, townhouse, semi-detached, and detached housing markets. Tu modified the dynamic stock-flow model and applied it to the Singapore private housing market [8]. Yong got the result that the movements in the real (Gross Domestic Product) GDP per capita and the total housing stock were found to significantly impact real housing prices in the long run. Zhang et al. got the survey results of a positive correlation between property cost and housing price [9]. However, it just focused on the private housing market, and the results may not be suitable to be used in the HDB market. Gang et al. used the decision tree approach, examining the relationship between house prices and housing characteristics are a factor influencing the housing price.

Thus, this essay will apply the multiple linear regression model to learn the effect of these six factors on the resale price of houses in Singapore. This paper focuses on six variables (Flat Type, Storey Range, Floor Area, Remaining Lease, Lease Commence Date, and Flat Model) and further finds a suitable model to find the relationship between these factors and housing prices.

2. Methodology

2.1. Data Source

The data is extracted from the website Kaggle for dissertation. The data set of the housing price from 2017 to 2022 was published by Singapore's Housing Development Board (HDB). The data includes 10 types of metrics such as month registered for resale, town, flat types and flat area in Singapore. There are 134168 groups of data in the dataset, this survey chooses 3000 of them as samples.

2.2. Variable Description

The housing price will be predicted based on the following 6 variables, which are showed by the Table 1. The ranges of these variables are also showed in the table. In this research, 6 variables are chosen, they are flat type, storey range, floor area, remaining lease, lease commence date and flat model. Apart from that, the independent variable is the release price.

Variable	Logogram	range
Resale Price	Y	218888-1418000
Flat Type	X ₁	0-5
Storey	X ₂	1-43
Floor Area	$\tilde{X_3}$	34-192
Remaining Lease	X ₄	43-95
Lease Commence Date	X ₅	1967-2019
Flat Model	X ₆	1-19

Table 1:	The	variables	used	in	the	model	

2.3. Model Instruction

Multiple linear regression is used to find the influencing factors. There is a dependent variable, resale price of the house, and five independent variables, room flat type, storey range, floor area, remaining lease, lease commence date and flat model. This paper aims to analysis how the six factors (X) influence the house resale price(Y) by using multiple linear regression. The equation can be generated:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5$$
(1)

3. Results and Discussion

3.1. Data Analysis

Table 2 shows the analysis of the original data, giving minimum and maximum values, mean, median and standard deviation of each variable.

Itoma	Min	Mox	Maan	۶D	Madian
Items	IVIIII	IVIAX	Mean	2D	Median
resale price	218888.000	1418000.000	552114.861	167014.195	528000.000
flat type	0.000	5.000	3.720	1.279	4.000
storey	1.000	43.000	7.727	5.973	7.000
floor area sqm	34.000	192.000	97.685	23.678	93.000
remaing lease	43.000	95.000	74.247	14.619	74.000
lease commence date	1967.000	2019.000	1997.722	14.593	1997.000
flat model	1.000	19.000	5.199	3.235	7.000

3.2. Correlation Analysis

As can be seen from the Figure 1 and Table 3, correlation analysis was used to study the correlation between six items resale price and storey range, flat type, floor area, remaining lease, lease commence date and flat model. Pearson correlation coefficient is used to indicate the strength of the correlation.



Figure 1: Pearson correlation visualization

All of the six variables have positive correlation with the resale price (dependent variable), and they all have the level of 0.01 significance. From Figure 1, both X_1 and X_3 have higher correlation values relatively, with the value of 0.70 and 0.69 respectively.

By testing the multi-collinearity of six variables, it is clear that in Table 3 most of them are not closely related. Nevertheless, there are still some variables that are correlated to others. Take X_4 and X_5 as an example, the multi-collinearity is 0.999, indicating that X_4 and X_5 are similar to each other, and they influence the result of the model. Hence, it is necessary to delete one of them when modeling. (For convenient, only show X_4 and X_5).

	Mean Value	Standard Divination	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
X ₁	3.150	0.912	1					
X ₂	3.242	1.991	-0.019	1				
X ₃	97.685	23.678	0.954**	- 0.061**	1			
X ₄	273.341	142.186	0.158**	0.278**	0.078**	1		
X ₅	1997.722	14.593	0.148**	0.278**	0.071**	0.999**	1	
X ₆	5.199	3.235	0.079**	0.066**	0.132**	0.346**	0.343**	1

Table 3: Multi-collinearity

* p<0.05 ** p<0.01

3.3. Liner Regression Model

Table 4 shows the relationship of six factors and the house resale price. Variables with VIF higher than 5 mean that they are highly correlated to each other. Lease commence date and remaining lease have VIF value higher than 5, suggesting they are correlated. One of them should be eliminated to improve the model. By focusing on the p value, lesser the value of p, more the significant of variable is. Those with p value higher than 0.05 mean that they are insignificant and will not influence the dependent variable. Table 4 suggests that flat model, lease commence date and remaining lease are not significant.

To improve the model, the insignificant and correlated variables are eliminated. In table 5, all variables' p values are equal to 0, which shows that they are significant and will influence the resale price of houses in Singapore. The VIF values of them shows that are all below 5, which means that they are not correlate to each other. By looking at their B values, storey, flat area and remaining lease have positive influence on the resale price while flat type has a negative influence. With the increase storey and flat area, the resale price will increase. The remaining length of lease means how long they can own their houses. The longer time they can own, the higher the resale price. The flat type is also known as room allocation. The number of rooms increases will cause price to decrease. Compared to the previous model, the regression coefficients have changed slightly. There is no covariance issue in the improved model. The relationship can be explained in the following equation.

$$Y = -6236.981x_1 + 9622.849 + 44877.685x_3 + 2512.839x_4 - 162082.198$$
 (2)

The R^2 value represents the accuracy of the model. Both the models have the R^2 values of 0.674. Both of them have same accuracy, but the improved one is better as is no correlated variables and the insignificant variable is eliminated.

D	Std. Error	Beta	t	р	VIF	tolerance			
- 330772	6734729.760	-	-0.495	0.621	-	-			
- 910.481	1588.535	-0.053	-4.350	0.000	1.360	0.735			
624.689	305.237	0.344	31.532	0.000	1.095	0.914			
886.325	74.609	0.693	65.492	0.000	1.028	0.973			
924.163	3498.074	0.081	0.264	0.792	861.307	0.001			
647.510	3500.970	0.144	0.471	0.638	859.733	0.001			
572.127	657.466	-0.011	-0.870	0.384	1.490	0.671			
\mathbb{R}^2					0.674				
Adj R ²					0.674				
F					32.368,p=0.	000			
	- 330772 910.481 624.689 886.325 924.163 647.510 572.127	$\begin{array}{c} - & 6734729.760 \\ 330772 & 1588.535 \\ 910.481 & 305.237 \\ 624.689 & 305.237 \\ 886.325 & 74.609 \\ 924.163 & 3498.074 \\ 647.510 & 3500.970 \\ 572.127 & 657.466 \\ R^2 \\ Adj R^2 \\ F \end{array}$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$			

Table 4: Linear regression model 1

 Table 5: Linear regression model 2

Parameter Estimates								
	В	Std. Error	Beta	t	р	VIF	tolerance	
Constant	-162082.1	11649.634	-	-13.913	0.000	-	-	
flat type	-6236.981	1396.406	-0.048	-4.466	0.000	1.051	0.951	
storey	9622.849	305.167	0.344	31.533	0.000	1.095	0.914	
floor area sqm	4877.685	74.032	0.692	65.886	0.000	1.012	0.988	
remaining lease	2512.839	126.981	0.220	19.789	0.000	1.135	0.881	
\mathbb{R}^2				0.674				
Adj R ²				0.674				
F	F (4,2995)=1548.811,p=0.000							

4. Conclusion

The paper has selected 3000 samples of resale prices of houses in Singapore with 6 factors which are the storey, the flat model, the room allocation, the floor area, the lease commencing date and remaining lease date. The correlation is introduced first to eliminate the correlated factor. The lease commencing date and remaining lease are highly correlated to each other. By using the multiple linear regression model and checking the VIF value of variables, the flat model is proven not influencing the resale price of the houses. Hence, in the improved model, only the storey, the room allocation the floor area and the remaining lease are used. The first model and the improved model both can explain 67.4% of the resale prices of the houses in Singapore. People can take these factors into consideration when they purchase houses. However, there are still deficiencies which can be improved. More factors should be taken, for instance, the district of the houses. In addition, more samples can be selected to improve the accuracy of the model.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

- [1] Renaud, B. and Kim, K. (2007) The global housing price boom and its aftermath. Housing Finance International, 22(2), 3.
- [2] Lim, A. (2024) Housing cost statistics in Singapore. SmartWealth Singapore.
- [3] Chia, W., Li, M. and Tang, Y. (2017) Public and private housing markets dynamics in Singapore: The role of fundamentals. Journal of Housing Economics, 36, 44-61.
- [4] Lim, A. (2024) Average house price in Singapore: HDB, condo & landed. SmartWealth Singapore.
- [5] Ooi, J.T.L. and Le, T.T.T. (2012) New Supply and Price Dynamics in the Singapore Housing Market. Urban Studies, 49(7), 1435-1451.
- [6] Bian, T., Chen, J., Feng, Q. and Li, J. (2020) comparing econometric analyses with machine learning approaches: a study on Singapore private property market. The Singapore Economic Review, 67(06), 1787-1810.
- [7] Tan, J. T. and Liu, Y. (1998) Factors affecting housing prices in Singapore. WORLD SCIENTIFIC, 74-87.
- [8] Tu, Y. (2004) The Dynamics of the Singapore Private Housing Market. Urban Studies, 41(3), 605-619.
- [9] Zhang, J. Z., Ye J. S, Fang Y. and Zhang Y. J. (2023). The impact of green buildings, parks and other factors on housing prices and property fees: A case study of Singapore's private residential market. Architecture and Culture (5), 27-30.
- [10] Fan, G.Z., Ong, S. E. and Koh, H.C. (2006) Determinants of House Price: A Decision Tree Approach. Urban Studies, 43(12), 2301-2315.