

Research on the Influencing Factors of Rental House Prices

Xiaoyang Mao

Chemistry Department, Denison University, Granville, 43023, The US

mao_x1@denison.edu

Abstract. Since the 1990s, China's real estate market has gradually recovered, and has developed into a large and mature trading market today, as the housing price increase, the demand of housing rental also increases a lot. Based on this this article aims to identify the factors that have an impact on rental price. The method of multi-linear regression is used to analyze the factors with over 10,000 samples in Guangzhou from May 2020 to June 2020 from the website of LianJia, one of the biggest real estate agencies in China. The dataset contains 10 variables in total, this research uses VIF, Pearson correlation coefficient to narrow the variables down to District, Rent Type, Area, Elevator Flag and Buld Floor these five variables and considered that they have a significant linear relationship with rental price. By focusing on these variables this study provides a view of the relationship of demand and supply of China's renting market.

Keywords: Rental price; Linear regression; Pearson correlation coefficient.

1. Introduction

Since the housing reform in the 1990s, China's housing rental market has also begun to recover under the strong market, demographic fundamentals and strong support from the government [1]. In recent years, under the influence of many factors, the rental population in the country shows an increasing trend year by year [2]. With the expansion of the rental market and the continuous growth of the number of tenants, the study of the house rental price has become increasingly important. This will not only help strengthen the supervision of the rental market, but also provide tenants with better pricing references.

In recent years, with the promotion of national policies and the continuous improvement of the market, the housing rental market in Guangzhou has gradually matured, but it also faces many challenges [3]. On the one hand, the fluctuation of housing rental prices is affected by many factors, such as geographical location, housing quality, supporting facilities, etc. [4]. On the other hand, the demand for housing is increasingly diversified, and there are differences in the expectations and affordability of housing rental prices. Therefore, this paper aims to provide a reference for the government to formulate relevant policies, for developers to optimize product positioning, and for residents to rationally choose housing by systematically analyzing the influencing factors of housing rent prices in Guangzhou.

In related research fields, many scholars have conducted in-depth discussions on the influencing factors of housing rental prices. From the perspective of building characteristics, factors such as house size, orientation, floor, house age, elevator, and property management will all have an impact on rental price [5]. Another key factors that significantly impact rental prices in Guangzhou is transportation convenience. The city's robust public transportation system, including its extensive subway network and bus routes, has a profound effect on the accessibility and attractiveness of different neighborhoods. Properties located near major transportation hubs tend to command higher rents due to their superior connectivity and convenience. This is particularly true for commuters and young professionals who value efficiency and time savings [6]. In addition, environmental factors also play a crucial role in determining rental prices. The accessibility of parks and green spaces, as well as the level of industrial pollution, can significantly influence the desirability and pricing of housing units. Residents are increasingly aware of the importance of living in a healthy and sustainable environment, and this is reflected in the rental market. Properties situated in areas with



good air quality, ample green spaces, and low noise pollution tend to attract higher rents [7]. Moreover, different groups have distinct housing preferences that further contribute to the uncertainty in rental prices. For instance, students and young professionals may prefer apartments located near universities or business districts, while families with children may prioritize neighborhoods with good schools and amenities. These preferences create demand for specific types of housing, which in turn affects the pricing dynamics [8]. Lastly, China's unique school district housing policy has a significant impact on real estate rental prices, including in Guangzhou. The policy links school enrollment to property ownership or lease, creating a strong demand for housing units within desirable school districts. This demand often drives up rental prices in these areas, especially during peak enrollment periods [9, 10].

To sum up, this paper will comprehensively analyze the influencing factors of housing rental prices in Guangzhou from multiple perspectives. First, this paper will review the domestic and foreign research literature on the influencing factors of housing rent prices, summarize the achievements and shortcomings of previous studies, and provide theoretical support for this study. Secondly, this paper will use the multi-linear regression, combined with the actual situation of Guangzhou, to build the index system of the influencing factors of rent price, and use the econometric model for empirical analysis. Finally, this paper will put forward corresponding policy suggestions and practical suggestions according to the research results, to provide useful reference for the healthy development of the Guangzhou housing rental market.

2. Methods

2.1. Data Source

The data for this literature is collected from Kaggle website, which is scraped from LianJia website on 23rd May 2020, 3rd June 2020, 12 June 2020 and 20 June 2020. The dataset contains more than 10,000 records, and this research selected 500 of them as a sample. The original dataset remained in .csv format.

2.2. Variable Selection

The data uses in this paper count a total of 11128 individuals. From Table 1, the data contains ten variables (district, rent type, area, price, unit price, house floor, buld floor, elevator flag, community ID and Geometry). This paper does a correlation test to select and check the variables which have high correlation with rental price. The raw data contains some difficult elements, such as the house floor data contains not only specific floors, but also a lot of ambiguous floor information. In order to deal with these fuzzy data, the dataset authors used high, medium and low terms instead, which confused the data analysis, so this variable was excluded. In addition, geographical location variables are also excluded because they contain latitude and longitude information and are difficult to incorporate into multiple linear regression models.

Table 1. All variables and their meanings in the dataset

Variable	Meaning
District	Administrative district of the house
RentType	The preferred types of rent. (whole/share)
Area	The size of the house in square meter. for a shared type of rent.
Unitprice	Calculated by price/area
Price	The rental prices that are asked for by the landlords, in cny.
HouseFloor	The floor of the house. some of the records do not present the exact floor, but rather label them as l (low).
BuldFloor	The highest floor of the building where the houses locate.
ElevatorFlag	Indicator of whether the building has at least one elevator.
Community ID	ID representing the community where the houses locate.
Geometry	Geographic coordinates of the communities, in EPSG 4326.

2.3. Method Introduction

This paper uses Pearson correlation coefficient to test the correlation between the price and other variables. This test is mainly aim for the correlation between variables and select the variables can be used in the further multi-linear regression model. The multi-linear regression model is also used in this paper. It is a linear regression model with more than one variable. It is used in explain the linear relationship between one variable and multiple related variables to make the model closer to the reality.

$$E(Y) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p + e \quad (1)$$

In the above formula: β_0 is a constant term, and e is a residual term.

3. Results and Discussion

3.1. Multi-linear Regression

This analysis shows that there are many factors can impact on rental price. As Table 2 shows:

Table 2. Multi-linear Regression analysis on six variables

	Nonnormalized Coefficient		Normalized Coefficient	t	p	Collinearity	
	B	S.E.	Beta			VIF	Tolerance
Constant	-360.505	377.670	-	-0.955	0.340	-	-
District (x1)	84.865	10.472	0.057	8.104	0.000**	1.016	0.984
RentType (x2)	-70.768	184.130	-0.003	-0.384	0.701	1.025	0.975
Area (x3)	41.938	0.445	0.664	94.142	0.000**	1.030	0.970
BuldFloor (x4)	35.954	3.518	0.089	10.221	0.000**	1.559	0.642
ElevatorFlag (x5)	129.228	81.439	0.014	1.587	0.113	1.521	0.658
CommunityID (x6)	0.009	0.027	0.003	0.348	0.728	1.099	0.910

Based on the analysis above the relevant multi-linear regression equation is available:

$$Price = -360505 + 84.865x_1 - 70.768x_2 + 41.938x_3 + 129.228x_4 + 35.954x_5 + 0.009x_6 \quad (2)$$

It is showed in the regression analysis that the coefficient value of District is 84.865, and t is equal to 8.104, p is 0.000, smaller than 0.01, which indicate that District has a significant positive impact on Price. The coefficient value of Rent Type is -70.768, and t is -0.384, p is 0.701, larger than 0.05, it shows that Rent Type has no significant effect on Price. The coefficient value of Area is 41.938, with t equals 94.142 and p equals 0.000, which is samller than 0.01, indicating that Area has a significant positive effect on Price. The coefficient value of Buld Floor is 35.954, t is 10.221, and p is 0.000, smaller than 0.01, which means that Buld Floor has a significant positive effect on Price. The coefficient value of Elevator Flag is 129.228 it has a t equal 1.587 and p equals 0.113, which is larger than 0.05, which indicates that Elevator Flag has no significant effect on Price. The coefficient value of Community ID is 0.009 with a t value of 0.348 and p value of 0.728, larger than 0.05, this shows that Community ID doesn't have a significant influence on Price. To sum up, District, Area and Buld Floor have a significant positive impact on Price, while Rent Type, Elevator Flag and Community ID have no significant impact on Price

3.2. Data Relevance

The regression analysis shows that only district, area and buld floor have a significant positive impact on the rent price. Pearson Correlation Coefficient is introduced to double check the correlation between rent price and the independent variables. From Table 3, it demonstrates the Pearson correlation coefficient of the independent variables with respect to price. Given the complexity of geometry and house floor in the dataset, they are not incorporated in the analysis. Furthermore, the unit price equates to price per area and is solely reliant on price and area, excluding it from the test. The table above allows for determining the correlation between price and the independent variables. The correlation between price and region is 0.070, significant at the 0.01 level, indicating a significant positive correlation between the two. The correlation between price and lease type is 0.086, also significant at the 0.01 level, suggesting a significant positive correlation. The correlation between price and elevator sign is 0.107, significant at the 0.01 level, signifying a significant positive correlation. The correlation between price and area is 0.671, significant at the 0.01 level, indicating a significant positive correlation. The correlation between price and building floor is 0.147, significant at the 0.01 level, suggesting a significant positive correlation. The correlation between price and community ID is -0.001, near 0, with a P-value of 0.877, greater than 0.05, suggesting no correlation between price and community ID. In conclusion, district, rent type, elevator flag, area, and building floor all exhibit a significant correlation with the rent price. However, the community ID doesn't show any significant correlation to the rent price, so it can be ignored in further analysis.

Table 3. Pearson correlation coefficient analysis on variables

	Price
District (x1)	0.070**
RentType (x2)	0.086**
Area (x3)	0.671**
BuldFloor (x4)	0.147**
ElevatorFlag (x5)	0.107**
CommunityID (x6)	-0.001

3.3. Re-build Multi-linear Regression Model

According to the Pearson correlation coefficient a new multi-linear regression model was built. District, Rent Type, Area, Elevator Flag and Buld Floor are the remaining independent variables (Table 4).

Table 4. Multi-linear Regression Analysis five remaining variables

	Nonnormalized Coefficient		Normalized Coefficient	t	p	Collinearity	
	B	S.E.	Beta			VIF	Tolerance
Constant	-346.032	375.352	-	-0.922	0.357	-	-
District (x1)	84.809	10.471	0.057	8.100	0.000**	1.016	0.984
RentType (x2)	-68.823	184.037	-0.003	-0.374	0.708	1.024	0.976
Area (x3)	41.947	0.445	0.664	94.328	0.000**	1.027	0.974
ElevatorFlag (x4)	125.881	80.864	0.013	1.557	0.120	1.500	0.667
BuldFloor (x5)	35.757	3.471	0.088	10.300	0.000**	1.518	0.659

Based on the new regression model, new equation can be obtained:

$$Price = -346.032 + 84.809x_1 - 68.823x_2 + 41.947x_3 + 125.881x_4 + 35.757x_5 \quad (3)$$

In the table 4, no variable has a VIF greater than 5, which indicate that there is no covariance problems occur.

Table 5. Multi-linear Regression Model Summary

R	R2	Adjusted R2	RMSE	DW	AIC	BIC
0.681	0.464	0.463	3252.976	1.570	211469.338	211513.238

According to the data in the above table 5, the square value of model R is 0.464, which indicates that factors such as District, Rent Type, Area, Elevator Flag and Buld Floor can explain 46.4% of the change in Price. Moreover, during the F-test of the model, the result shows that the model passes the F-test ($F=1920.929$, $p=0.000<0.05$), which indicates that the construction of the model has practical significance.

4. Conclusion

The study selected 11128 samples from LianJia from May to June 2020, which contains 10 variables. This paper uses multi-linear regression and Pearson correlation coefficient combine with logical reasoning to reduce the variables to 5 and conclude that District, Rent Type, Area, Elevator Flag and Buld Floor have a significant impact on the rent price.

It cannot be denied that due to the change of the policy, supply and demand relationship and customer's expectation this model cannot perfectly fit in the real market situation. Also, the data for the district is not evenly distributed, which may also make the model has a district preference. However, this study still has a lot of value in some aspects. First of all, the method used in this study is innovative, the use of multi-linear regression is to find out the relationship between variables and the rental price. Compare to the single factor analysis method it makes the analysis more comprehensive and closer to the reality. It provides an insight into the proportionate weight of location and the building itself in rental pricing. On one hand, the analysis of this model reflects customer's demand for rent from the side, which provides the direction for the housing agent when choosing the high-quality housing. This can improve the relationship between supply and demand of rent market, making the housing agent choose more targeted housing and lower down the cost of choosing low demand housing. On the other hand, this study shows the relationship between variables and rental price, which also gives the customer a measuring standard when renting house. This can help the customer balance the housing demand and cost, lower down the waste of money on unwanted requirements. In general, this study gives the supplier and customer in the housing rent market a guidance on the demand and supply, which can improve the market efficiency and reduce the cost of unwanted demand.

References

- [1] W. Y. Zhang, People Living in rental residential housing in China 2017-2023, Statista, (2022).
- [2] Y. Wang, et al. Spatial Differentiation and Influencing Factors Dataset of Housing Rents in Guangdong-Hongkong-Macao Greater Bay Area, *Journal of Global Change Data & Discovery*, 6 (1) (2022) 37-44.
- [3] Y. Wang, et al. The core influencing factors of housing rent difference in Guangzhou's urban district, *Acta Geographica Sinica*, 76(8) (2021).
- [4] L. Q. Gu, On the perfection of China's Long-term Rental System, *Frontiers in Business, Economics and Management*, (2022).
- [5] T. Hu, The rise of China's rental housing market, Jones Lang LaSalle, (2018).
- [6] D. G. Wang, S. M. Li, Socio-economic differentials and stated housing preferences in Guangzhou, China, *Habitat International*, (2006).
- [7] Y. X. Xu, S. J. He, J. X. Qian, An investigation on the emerging housing rental market in a studentified village in Guangzhou: a new institutional economics perspective. University at Albany, (2020).

- [8] Y. L. Ouyang, K.Y. Sun, Empirical Evidence Based on Geographic Regression Discontinuity Analysis of Housing in Guangzhou School District. International Conference on Finance, Investment and Business Analysis (FIBA 2022), (2022).
- [9] Y. Gao, C. H. Wang, H. H. Qiao, P. Yin, The spatial differentiation pattern and influencing factors of housing prices in Shanghai's tourism and accommodation industry, *Geography*, 42 (8) (2022) 11.
- [10] C. Y. Zhou, C. Y. Li, Research on Regional Second hand Housing Price Prediction Method Based on PSO-LSTM, *Modern Information Technology*, (2024).