The 6th International Conference on Economic Management and Green Development (ICEMGD 2022)
DOI: 10.54254/2754-1169/4/20221059

Influence of Objective Factors on House Prices

Lyuyangguang Lyu[1,a,*], Xinyu Song[2,b]

[1]*International College, Wenzhou Business School, Wenzhou, China*
[2]*College of Accounting, Shandong Technology and Business University, Yantai, China*
a. 1477608853@qq.com, b. sxydyx5757@163.com
**corresponding author*

Abstract: With the development of economy and the improvement of people's living standards, many people will try to buy houses. But when it comes to buying a house, people will also pay attention to some objective factors such as the size of the house and supporting facilities. This paper uses the statistical analysis software SPSS linear regression analysis, factor analysis, comparative analysis and other methods to analyze the impact of objective factors on house prices, through data analysis, the fitting degree also reached about 0.88. In the experimental results, it is found that taxi distance, market distance, hospital distance, carpet area, built up area, parking type, city type, rainfall have a significant impact on house prices. Because in real life, These objective factors will be considered when people choose to buy a house.

Keywords: data processing, house prices, infrastructure.

1. Introduction

House prices have been widely concerned by people, but this problem has not been well solved. House price prediction has also become a hot topic of concern for governments, markets and people all over the world. In a complete market environment, housing prices are determined by both demand and supply. However, in the current society, housing has both residential and investment effects, which makes the research on housing prices very complex. Predicting long-term housing prices has become an impossible task [1]. Before the housing crisis in 2007-2008, most people believed that house prices had been rising for many years, and those who invested in real estate would certainly be rewarded. But this is not the case, because banks are basically approving loans to people who are unable to pay for their houses, and even financial institutions approve loans to ordinary individuals at variable interest rates (that is, the interest rates will change according to the current market interest rates). When the crisis occurs, many ordinary individuals are unable to repay their mortgages. Of course, there are other reasons leading to the financial crisis, such as the introduction of complex financial instruments (derivatives are still not widely understood), hedging financial instruments (credit default swaps) [2] and the deregulation of the entire financial industry. Although we can discuss the factors that led to the financial crisis, the main purpose of this article is to determine which possible characteristics will have a real impact on the overall value of housing.

In academic circles, many scholars have also analyzed and predicted the impact of objective factors on house prices. Supporting facilities are an important factor for house prices. From the perspective of urban resource allocation [3,4], such as medical care, education and market, constitute the basic

© 2023 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/).

DOI: 10.54254/2754-1169/4/20221059

elements of housing price differentiation [5]. Feng et al. believe that the premium effect of public service facilities on the housing market is the key determinant of housing prices, which leads to competition among different social groups in the housing market and intensifies spatial inequality [6]. Chen et al. mainly studied the impact of the geographical distance between Shanghai and the city center (whether the supporting facilities are complete) on the housing sales price, and discussed how the price gradient in different directions in Shanghai changes [7]. Lou et al. believe that house prices are not only affected by the quality of houses, but also closely related to geographical location. After the research and analysis, it is found that schools, subway stations and green spaces will have a positive impact on the house price in the study area, while business centers and hospitals will have a negative impact [8]. Moreover, Heyman et al. believed that location is the main determinant for people to estimate the value of houses [9].

The research content is arranged as follows: Chapter 2 is the comparative analysis of the original data, normality test, independent sample t-test, factor analysis. Chapter 3 is the model establishment of linear regression. Chapter 4 summarizes the full text and draws a conclusion.

2. Exploratory data analysis

Variable meaning: Taxi_dist (Taxi distance): Market_dist(Market distance)； Hospital_dist(Hospital distance)； Carpet_area(Indoor area/Carpet area); Builtup_area(Construction area), Parking_type (Parking permit), City_type(City type)

2.1. Normality test

All variables were tested for normality, and the results are shown in the following table 1:

Table 1: This caption has one line so it is centered.

variable	sample size	average value	Kolmogorov-Smirnov test	
			Statistics D value	*p*
Taxi_dist	919	8229.728	0.012	0.991
Market_dist	919	11018.753	0.037	0.004**
Hospital_dist	931	13072.092	0.022	0.318
Carpet_area	924	1511.863	0.262	0.000**
Builtup_area	917	1794.925	0.108	0.000**
Parking_type	932	2.840	0.241	0.000**
City_type	932	1.902	0.231	0.000**
Rainfall	932	785.579	0.021	0.389
Price_house	932	6084695.279	0.248	0.000**

The normality test is carried out for Taxi_dist, Market_dist, Hospital_dist, Carpet_area, Builtup_area, Parking_type, City_type, Rainfall, Price_house. It can be seen from the above table that the sample size of the research data is all greater than 50, so the K-S test is used. Specifically, a total of 6 items of Market_dist, Carpet_area, Builtup_area, Parking_type, City_type, Price_house showed significant ($p<0.05$), meaning that Market_dist, Carpet_area, Builtup_area, Parking_type, City_type, Price_house did not have normality characteristics. In addition, the three items of Taxi_dist, Hospital_dist, Rainfall did not show significant ($p>0.05$), which means that Taxi_dist, Hospital_dist, Rainfall have normal characteristics.

In conclusion, Market_dist, Carpet_area, Builtup_area, Parking_type, City_type, Price_house do not have normality characteristics. Also, Taxi_dist, Hospital_dist, Rainfall have normality traits.

DOI: 10.54254/2754-1169/4/20221059

2.2. Non-parametric Test

From the normality test, it can be seen that Market_dist, Carpet_area, Builtup_area, Parking_type, City_type showed significant ($p<0.05$), so these variables were tested nonparametrically. The values in Price_house that are less than the mean are marked as 0, and the values greater than the mean are marked as 1.

Table 2: Non-parametric test.

	Price_house median M(P25，P75)		p
	0.0(n=508)	1.0(n=424)	
Market_dist	10763.000(8950.0,12407.0)	11544.500(9684.8,13156.8)	0.000**
Carpet_area	1463.000(1305.0,1626.0)	1496.000(1346.0,1682.0)	0.007**
Builtup_area	1758.000(1570.0,1948.0)	1794.000(1613.5,2015.5)	0.013*
Parking_type	3.000(2.0,4.0)	3.000(2.0,4.0)	0.593
City_type	2.000(2.0,3.0)	1.000(1.0,2.0)	0.000**

As can be seen from the above table 2, the non-parametric test is used to study the difference of Price_house to Market_dist, Carpet_area, Builtup_area, Parking_type, City_type a total of 5 items , the summary analysis shows that: using the MannWhitney test statistic for analysis, the analysis shows that: different Price_house samples will not show significant difference for a total of 1 Parking_type ($p>0.05$), and the Price_house sample has a total of 4 for Market_dist, Carpet_area, Builtup_area, City_type Items showed significant differences ($p<0.05$).

2.3. Independent Sample T-Test

From the normality test, it can be seen that Taxi_dist, Hospital_dist, Rainfall a total of 3 items did not show significant ($p>0.05$), so the independent sample T-test was performed. The values in Price_house that are less than the mean are marked as 0, and the values greater than the mean are marked as 1.

Table 3: Independent sample t-test.

		F	salience	t	degrees of freedom	Significance (two-tailed)
Taxi_dist	Assume equal variances	11.634	.001	-3.454	917	.001
	do not assume equal variances			-3.401	816.263	.001
Hospital_dist	Assume equal variances	8.810	.003	-3.944	929	.000
	do not assume equal variances			-3.894	842.155	.000
Rainfall	Assume equal variances	.061	.805	.187	930	.852
	do not assume equal variances			.187	901.654	.852

As can be seen from the above table 3, the independent sample ta test was used to study the differences of Price_house for Taxi_dist, Hospital_dist and Rainfall. The summary analysis showed that different Price_house samples did not show significant difference for Rainfall ($p>0.05$). , and the Price_house sample showed a significant difference ($p<0.05$) for a total of 2 items of Taxi_dist and Hospital_dist.

DOI: 10.54254/2754-1169/4/20221059

2.4. D. Factor Analysis

Through the result graph of factor analysis, it can be seen that the significance is 0.000. It is considered that there is a significant difference between the correlation coefficient matrix and the identity matrix.The extraction of variables in the common factor variance table is basically above 0.7 and most of them are above 0.85, indicating that the common factor is well expressed. Five factors extracted from the total variance interpretation table jointly explained 89.383 of the total variance, so the original data lost less information, and the result of factor analysis was ideal.

Table 4: Rotated component matrix.

	1	2	3	4	5
Carpet_area	.991				
Builtup_area	.941				
Price_house	.911				
Hospital_dist		.936			
Taxi_dist		.872			
Market_dist		.780			
City_type			.993		
Rainfall				.997	
Parking_type					.997

As can be seen from the above table 4, the rotated component matrix can be seen that Carpet_ area, Builtup_ area, Price_house are located on the first factor and interpreted as housing information; Hospital_dist, Taxi_dist, Market_dist are located on the second factor and interpreted as the distance of infrastructure; City_type is located on the third factor On the factor and is the interpretation of the city type; Rainfall is on the fourth factor and is the interpretation of the rainfall; Parking_type is on the fifth factor and is interpreted as the parking space information. It can be seen that the meaning of the factors is very clear.

3. Model Building

Assume that Taxi_dist, Market_dist, Hospital_dist, Carpet_area, Builtup_area, Parking_type, City_type, Rainfall variables can predict house prices.

Table 5: Model summary table.

	R	R-square	Adjusted R-square	Error in standard estimation	Debin-Watson
1	.905[a]	.818	.818	2176566.578	
2	.942[b]	.887	.887	1720028.281	
3	.943[c]	.889	.888	1706073.726	1.938

As can be seen from the above table 5, the model summary table shows the fitting degree of the model.From the results of the model summary table in the linear regression analysis, it can be seen that there is only a single variable Carpet_ area, the R-square (0.818) is greater than 0.5 and the data fitting is good; When adding a variable Builtup_ area, the explanatory enhancement of R-square becomes 0.887 and the fitting degree also increases; When adding another variable Hospital_ dist, the interpretability and fitting degree of R-square are improved to 0.889. From the table, we can get

DOI: 10.54254/2754-1169/4/20221059

the conclusion that the fitting degree of the whole model is the best when there are three variables Carpet_ area, Builtup_ area and Hospital_ dist in the model at the same time.

Table 6: ANOVA.

Model		Sum of squares	Freedom	Mean	F	Sig.
1	Regression	1.914E+16	1	1.914E+16	4040.085	.000[b]
	Residual	4.245E+15	896	4.737E+12		
	Total	2.338E+16	897			
2	Regression	2.074E+16	2	1.037E+16	3504.576	.000[c]
	Residual	2.648E+15	895	2.958E+12		
	Total	2.338E+16	897			
3	Regression	2.078E+16	3	6.927E+15	2379.994	.000[d]
	Residual	2.602E+15	894	2.911E+12		
	Total	2.338E+16	897			

As can be seen from the above table 6, the ANOVA shows whether the data model is significant. The three groups of models in the ANOVA analysis table passed the F-test and the significance was 0.000 less than 0.05, then the model construction is meaningful. It can be concluded that house prices are significantly affected by three variables Carpet_ area, Builtup_ area and Hospital_ dist.

In the beta values of the regression coefficient table, the values of Carpet_ area and Hospital_ dist are both positive, and only the value of Builtup_ area is negative, indicating that the impact of Carpet_ area and Hospital_ dist on house prices is positively correlated, while Builtup_ area is negatively correlated, it also means that Taxi_dist,Market_dist, Parking_type,City_type，Rainfall variables will not have an impact on house prices. Then, the significance of T-test is 0.000, indicating that the result is very significant. Finally, the VIF values of the three groups of models are between 0-10 and the tolerance is not less than 0.1, which shows that there is no multilinearity. Then the experimental results reject the original hypothesis.

$$Y = 5764.491XC - 7804.659XB + 87.981XH - 3617504.798 \quad (1)$$

4. Conclusion

This paper studies the impact of objective factors on house prices and establishes a model through data, with a fitting degree of about 88%. This also really shows that housing prices are affected by some social and natural factors such as surrounding infrastructure, housing information, etc. this complex influence relationship can be used as a reference for housing builders and buyers and sellers. The study also found that the level of housing prices will be restricted by the type of city. For example, in cities with high-speed economic development, house prices will rise on the basis of a certain minimum price, and generally will not be too low; However, in economically underdeveloped cities, housing prices near the city center may be slightly higher, and relatively speaking, housing prices in other regions will not be particularly high.

References

[1] Zhou Y. Influencing factors of real estate prices in China, Hangzhou: Zhejiang University, 2017
[2] Adam Hayes, Credit Default Swap (CDS), Updated March 02, 2022
[3] Thomas Francis Galligan, An Urban Comprehensive Planning Model for Resource Allocation, May 05, 2020

DOI: 10.54254/2754-1169/4/20221059

[4] Wiley,Markets and Efficient Resource Allocation in Urban Areas.
[5] Shasha Lu, Xingliang Guan, Min Zhou, Yang Wang, Land resources allocation strategies in an urban area involving uncertainty: a case study of Suzhou, in the Yangtze River Delta of China,2014 February 14
[6] Feng L, Qi W, Tao Z, Huili D, Spatial Effects of Public Service Facilities Accessibility on Housing Prices: A Case Study of Xi'an, China,29 November,2018
[7] Jie Chen, Qianjin Hao, The impacts of distance to CBD on housing prices in Shanghai: a hedonic analysis,11 August,2008
[8] Hanning Lou, Wu Xiao, Debin Lu, Understanding the Impacts of Public Facilities on Residential House Prices: Spatial Data-Driven Approach Applied in Hangzhou, China, February 24, 2022
[9] Axel Viktor Heymana, Dag Einar Sommervoll, House prices and relative location, December 2019, 102373Axel Viktor Heymana, Dag Einar Sommervoll, House prices and relative location, December 2019, 102373