ISSN 2616-5902 Vol. 4, Issue 3: 83-86, DOI: 10.25236/AJBM.2022.040315

Analysis of Commodity Housing Price in Shanghai Based on Multiple Linear Regression Model

Xinyu Shen

University of Shanghai for Science and Technology, Shanghai, China
shenmonkey@126.com

Abstract: *As Shanghai is committed to building a vibrant, thriving home for all residents, the pursuit of a high quality of life is correspondingly strengthening. In particular, housing has probably become an important indicator to measure the quality of people's lives and the house price is the key to the housing problem. This article analyses the annual date of the commodity house prices in Shanghai from 2001 to 2020 and aims to explore the affecting factors of commodity housing price, which include the number of ordinary high schools, medical and health institutions and cultural institutions in Shanghai, urban green coverage and residents' disposable income. Therefore, a multiple linear regression model of commodity house price has been established and successfully underwent the test of multicollinearity, heteroscedasticity and sequence correlation.*

Keywords: *House Price; Multiple Linear Model; Impact Factor*

1. Introduction

House price has always been a hot issue concerned by the society and the public. The price of commodity house has always been closely linked with China's economic development and social livelihood. The "14th Five-Year Plan" of Shanghai Housing Development released in 2021 vigorously advocate solving the housing problem, thus creating a more comfortable environment for the real estate industry in Shanghai. Therefore, it is of great importance for resident investment and government decision to scientifically understand the he development law of commodity house price in Shanghai and analyze the affecting factors. The research on the real estate industry in China has been quite comprehensive, so this article specifically focuses on the aspect on "the supporting services of residential community " in the policy mentioned above, which takes education, sanitation, environment and consumption capacity into account, to analyze its joint impact on the commodity house price in shanghai.

2. Data analysis

2.1 Data collection

Multiple linear regression model is established to analyze the affecting factors of house price in Shanghai. As a result, the commodity house price in Shanghai has been taken as the dependent variable, and the number of ordinary high schools, medical and health institutions and cultural institutions, urban green coverage and residents' disposable income has been taken as independent variables. The annual data from 2001 to 2020 is selected, which all comes from National Bureau of Statistics of China.

2.2 Index explanation

2.2.1 The number of ordinary high schools

The number of schools has been chosen as an important indicator to measure the regional education level. Generally speaking, the larger number of schools reveals the more educational resources. In addition, "School District Housing" policy has greatly strengthened the influence of education on the house price.

2.2.2 The number of medical and health institutions

The number of medical and health institutions in Shanghai is selected as an important index to measure the regional medical level. It is generally acknowledged that the larger number of the medical

ISSN 2616-5902 Vol. 4, Issue 3: 83-86, DOI: 10.25236/AJBM.2022.040315

and health institutions reflects the higher level of medical care. Strengthening the health care functions as an effective way to improve housing quality.

2.2.3 The number of cultural institutions

The sum of museums and public libraries is selected to measure the regional cultural level, which plays an indispensable role in improving the happiness of housing.

2.2.4 Urban green coverage

Urban green area is selected to measure the regional environment. The environmental factor is of great importance in increasing the livable level of housing.

2.2.5 Residents' disposable income

Residents' disposable income emphasizes the part of income used for free consumption, which indirectly reflects the concern of consumers on housing prices.

3. Empirical research

3.1 Model theory

Multiple linear regression model is a model for modeling and analyzing multiple independent and dependent variables through least square function. Its common formula is:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \mu_i \quad (i = 1,2,\ldots,n)$$

In the formula above, figures like $\beta_1,\ldots,\beta_k$ are considered as regression coefficients and μ_i is a constant term, which represents a series of minute factors that affect the change. The model illustrates the degree to which the explained variable Y_i changes when the explanatory variable X_{ki} changes by one unit.

3.2 Model establishment

Variables in this articles are set in the following table:

Table 1: Variable setting

Y	The annual commodity house price in Shanghai
X_1	The number of ordinary high schools
X_2	The number of medical and health institutions
X_3	The number of cultural institutions
X_4	Urban green coverage
X_5	Residents' disposable income

Therefore, the formula can be written as:

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

In the formula above, figures likeβ1,…, β5 are considered as regression coefficients and β0 is a constant term. Residual is ε, representing several minor factors that affect the change.

The initial fitting result by EViews is:

$$Y = -21652.791739 + 63.1981587306X_1 + 0.136351699144X_2 - 33.6869939482X_3 + 0.448916830447X_4 + 549.278596182X_5$$

(-4.292422) (4.239873) (0.641946) (-2.324919) (13.11730) (3.422254)

$$R^2 = 0.992050 \quad \bar{R}^2 = 0.989211 \quad F = 349.3979$$

The result shows that nearly 99% of commodity house price in Shanghai from 2001 to 2020 can be explained by these 5 variables. The signs and values of the regression coefficient are overall reasonable, but the number of cultural institutions in Shanghai is negatively correlated with the house price, which needs further research. The figure of R^2 and $\bar{R}^2$ is both closed to 1, showing that the model has superior fitting degree. The F statistic is 349.3979 and F0.05(5,14)=2.958, of which the former exceed the latter. It is obvious that the model has successfully passed the F-test and is statistically significant overall with an absolute linear relationship. While in the t-test, all the variables except X2 have the significant effect

ISSN 2616-5902 Vol. 4, Issue 3: 83-86, DOI: 10.25236/AJBM.2022.040315

on the commodity house price.

To sum up, the problem that the sign of the regression coefficient β_3 is negative and X2 fails to pass the t-test needs further modification and confirmation.

4. Model testing and modification

4.1 Test and correction of multicollinearity

4.1.1 Test by correlation coefficient method

The correlation coefficients between each variables have been calculated out with the help of EViews, of which the lower limit is close to 0.5 and the upper limit is up to 0.9. It shows that the explanatory variables can be approximated linearly by other explanatory variables, that is to say, multicollinearity does exist.

4.1.2 Modification stepwise regression method

Stepwise regression method contributes hugely to the model modification. The adjusted model is:

$$Y = -22365.8506557 + 0.443462520738X_4 - 31.1550522467X_3 + 66.2686010996X_1 + 595.75633593X_5$$

(-4.637125) (13.64684) (-2.279678) (4.788884) (4.243004)

$$R^2 = 0.991816 \ \bar{R}^2 = 0.989633 \ F = 454.4562$$

The figure of R^2 and $\bar{R}^2$ is both closed to 1, showing that the model has superior fitting degree. The F statistic is 454.4562 and $F_{0.05}(4,15)=3.056$, which reveals that the model has successfully passed the F-test. Additionally, the probability of t statistic of each explanatory variable is below 0.05, thus leading to the passing of t-test. As a result, the multicollinearity of the model has been well resolved.

4.2 Test of heteroscedasticity

White test has been adopted to check whether the heteroscedasticity exists in the adjusted model. At a significance level of 5%, the probability of the F statistic and chi-square both exceeds 0.05, which confirms the fact that there is no heteroscedasticity in this model.

4.3 Test of autocorrelation

Dubin-Watson test is used in the adjusted model to test the autocorrelation. The Durbin-Watson statistic of the model above is 2.221419 and under the 5% significance points its dL statistic is 0.684 and the dU statistic is 1.567. As the Durbin-Watson statistic of the adjusted model is between dU and 4-dU, autocorrelation does not exist.

5. Conclusion

The article establishes a multiple linear regression model with the data of potential factors that affect the commodity house price and successfully undergoes the test of multicollinearity, heteroscedasticity and autocorrelation. Therefore, the optimal model of influencing factors of commodity house price in Shanghai is obtained: $Y = -22365.8506557 + 0.443462520738X_4 - 31.1550522467X_3 + 66.2686010996X_1 + 595.75633593X_5$. The model illustrates that the commodity house price is closely linked with the number of schools and cultural institutions, disposable income of residents and urban green coverage. And education, income and environment are positively correlated with the house price, of which the tiny rise would probably lead to the huge change of the house price. These three influencing factors should be given priority in the future formulation of house price policies. However, there is a negative correlation between the number of cultural institutions and the housing price, which is not necessarily absolute. It is highly likely that the public pay less attention to culture compared with education and environment. In view of the incomplete linear relationship between commodity house price and various variables and the fact that government policies fail to be quantified easily, there will be some errors between the model prediction and the real value. However, this model has passed a series of tests, thus contributing hugely to the economy.

ISSN 2616-5902 Vol. 4, Issue 3: 83-86, DOI: 10.25236/AJBM.2022.040315

References

[1] Ma Lili. Influencing factors and Empirical Analysis of real estate price in Kunming city: A Case study of housing market [D]. Kunming: Yunnan University of Finance and Economics,2008.Fangfang. Research on power load forecasting based on Improved BP neural network [D]. Harbin Institute of Technology, 2011.

[2] Dai Lei, Li Xueting. Analysis of Influencing Factors of Second-hand House Price Based on Multiple Linear Regression Model -- A Case study of Certain District in Chengdu city [J]. Henan Building Materials, 2019 (05): 80-82.

[3] Liu Meifang, Liu Weiyue. An Empirical Study on the Influencing Factors of Commercial housing Price in Tianjin -- Based on multiple Linear Regression Model [J]. Times Economics and Trade, 2017(16): 84-87.