

Commodity House Price Analysis and Forecast based on Regression Analysis

Cheng Xiuli

Shandong Huayu University of Technology, Shandong, China

Keywords: multiple linear regression; grey correlation analysis; retarded growth; time series prediction; SPSS

Abstract: In recent years, the housing price has been one of the most concerned issues for people. This research takes Hainan province (mainly Haikou and Sanya) as an example to analyze the main influencing factors of commodity housing price and predict the commodity housing price in the future.

First of all, it can find out the factors affecting the commodity housing price, and make qualitative and quantitative analysis of the influencing factors. Through the grey relational analysis of twelve factors, five factors with the highest significance are selected. Then the correlation test is carried out and the correlation between the factors is given. Secondly, necessary data processing and correlation substitution are carried out for the five selected factors. It select three influencing factors, which are commercial housing sales area, real estate development investment and total income of tourism. Finally, multiple linear regression model is established.

$$Y = 9080.950348 + 5.701387807X_5 - 29.49014423X_3 + 46.73616269X_{13}$$

The time series model is used to predict the monthly data of the three influencing factors from June 2018 to May 2019, and the data is substituted into the regression analysis model, then it can predict the commodity housing price from June 2018 to May 2019.

1. Symbol description

The symbol description is shown as the table 1.

Table 1: The symbol description

Symbol	Description
X_1	Gross National Product (GNP)
X_2	population
X_3	Commercial housing sales area
X_4	Land prices
X_5	Real estate development investment
X_6	Consumer spending by urban residents
X_7	Consumer price index
X_8	Total retail sales of social commodities

X_9	Average wage of on-job employees
X_{10}	Per capita disposable income of urban residents
X_{11}	Real estate developer enterprise completed housing area
X_{12}	Number of real estate development enterprises
X_{13}	Total tourism revenue
Y	Average selling price of commercial housing
β	Regression coefficient

2. The factors affecting the commodity housing price

The impact of various factors on the commodity housing price is different, Therefore, it need to process the data and select the most relevant factors as the final independent variables for the model establishment.

Firstly, grey relational degree analysis ^{[1][2]} is conducted on all factors involved in question 1, and the analysis is shown as table 2.

Table 2: The significance of the various factors on the price of commercial housing data

Factors index	X_1	X_2	X_3	X_4	X_5	X_6
significant	0.7882	0.7701	0.7728	0.8462	0.7701	0.7117
Factors index	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}
significant	0.7681	0.7766	0.3513	0.5630	0.7703	0.7942

3. Establishment of regression analysis model

Based on the analysis of the screening results, pick out the five most significant factors with five data sets of 0.8462, 0.7942, 0.7882, 0.7766 and 0.7728, that is, land price(X_4), number of investment by developers(X_{12}), gross national product(X_1), total retail sales of social commodities(X_8) and sales area of commercial houses(X_3). However, in the study of the data after preliminary screening, it is found that these data are incomplete, inconsistent statistical caliber and other problems. In view of the above situation, before using these data to test the regression model, we carried out correlation substitution and other necessary technical processing on the relevant data, which is mainly to ensure the validity and availability of the data.

Through the correlation analysis of each factor, we can get the relationship between the investment amount of real estate development and the number of real estate development enterprises, which is shown as the figure 1.

At the same time, considering that tourism income will also have a great effect on fluctuations in house prices in recent years after 2010 year, especially in Sanya and Haikou of Hainan province, and tourism revenue also has correlation with regional GDP, which is shown as the figure 2, so it can use the tourism revenues (X_{13}) (main considering Sanya and Haikou) instead of regional GNP.

To sum up, we finally determine three factors (X_3, X_5, X_{13}) as the three independent variables of the model establishment through grey relational degree analysis and screening and substitution of linear correlation factors. After determining the independent variables, it establishes a ternary linear regression model

$$Y = \beta_0 + \beta_1 X_5 + \beta_2 X_3 + \beta_3 X_{13} + \varepsilon$$

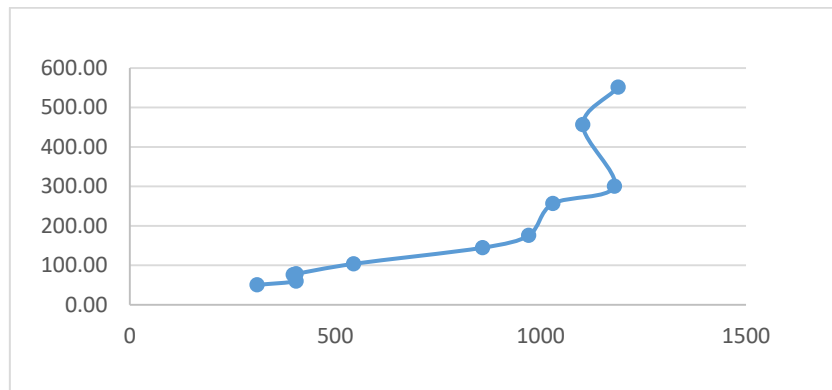


Figure 1: The relationship between the investment amount of real estate development and the number of real estate development enterprises

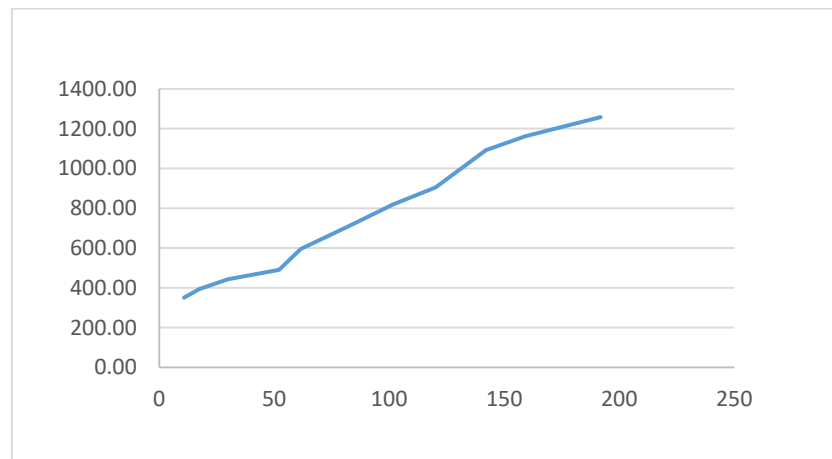


Figure 2. The correlation between total tourism revenue and regional GNP

MATLAB software is used to establish the regression coefficients of the three independent variables, as shown in table 3.

Table 3: The regression coefficient values

Regression coefficient	β_0	β_1	β_2	β_3
The values	9080.950348	5.701387807	-29.49014423	46.73616269

If the regression model is meaningful, the determination coefficient must meet the following values, which is shown as the Table 4.

Table 4: The determination coefficient value

R^2	F	p	msq
0.959942121	55.91571464	0.000029482	281940.8646

The output results show that the regression model ^[3] is

$$Y = 9080.950348 + 5.701387807 X_5 - 29.49014423 X_3 + 46.73616269 X_{13}$$

Because the determination coefficient is

$$R^2 = 0.959942121, p = 0.000029482 < 0.05$$

So, the regression model makes sense.

4. The solution of regression analysis model

The time series prediction model ^[4] ^[5] of SPSS software was used to predict the data of independent variables, and the predicted results are shown in table 5.

Table 5: The predicted results

Forecast time	Housing sales area	Real estate development investment	Total revenue of tourism
2018.6	46.67	45.05	21.14
2018.7	49.87	60.02	21.9
2018.8	57.82	80.75	20.96
2018.9	71.15	57.97	22.33
2018.10	54.41	66.94	22.9
2018.11	51.8	64.29	24.04
2018.12	60.56	70.41	29.46
2019.1	62.95	81.35	31.68
2019.2	73.31	100.48	37.18
2019.3	65.24	75.18	28.34
2019.4	70.11	64.59	28.53
2019.5	60.36	61.95	28.15

The predicted value of commodity housing price from June 2018 to May 2019 can be obtained by substituting the prediction data of independent variables into the established mathematical model of Y and X_5, X_3, X_{13} . The forecast house prices are shown in table 6.

Table 6: The forecast house prices

Forecast month	Forecast house prices
2018.6	14790.96
2018.7	15355.57
2018.8	15920.19
2018.9	16484.8
2018.10	17049.41
2018.11	17614.02
2018.12	18178.63
2019.1	18743.24
2019.2	19307.85
2019.3	19872.47
2019.4	20437.08
2019.5	21001.69

References

- [1] Si Shoukui. *Mathematical modeling algorithm and application*. National defense industry press.2011.
- [2] Tan Yongji. *CAI zhijie. Mathematical model*. Fudan university press.
- [3] Liu Laifu. Huang haiyang. Zeng yiyi. *Mathematical model and mathematical modeling*. Beijing normal university press.2000.
- [4] Wang Libin. *Multivariate statistical analysis: model, case and SPSS application*. Beijing: economic science press. 2010.3.
- [5] Liu Renyun. *Mathematical modeling methods and mathematical experiments*. China water conservancy and hydropower press.