Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 2, Issue 3: 9-15, DOI: 10.25236/AJCIS.020302

Analysis of Commodity Housing Price Based on Partial Least Squares Regression

Qimeng Tao

Science & Technology College, North China Electric Power University, Baoding, Hebei, 071003, China

Abstract. *With the development of China's economy and the continuous enhancement of its comprehensive strength, China has gradually entered the middle and upper levels of the world.However, residential prices have continued to rise and the number of properties has gradually increased.The increase in residential prices has exceeded the rate of economic growth.This paper mainly uses the idea of partial least squares regression to study the relevant data of a certain city. Calculated* $Q_h^2 = 0.0566 < 0.0975$*, to meet the principle of cross-validation judgment. Finally, the fitting coefficient of 0.88 was calculated, which proved that the equation fitting performance was good.*

Keywords: *partial least squares regression analysis of commodity housing price forecast*

1. Introduction

Since the housing reform in China in 1998, the real estate industry has made great achievements in the market development of 20 years, and it has also produced many problems. One of the incomprehensible phenomena is that house prices only rise or fall.This phenomenon is rare in the world.Although since the introduction of the control measures in March 2016, after more than two years of efforts, China's current housing prices are already in a more subtle state.From the housing prices of 70 large and medium-sized cities, housing prices in first-tier cities have fluctuated around zero growth rate for a year, and the growth rate of housing prices in second- and third-tier cities has stabilized for a long time around 5% and 7%. The income growth rate of China's residents is quite high, which indicates that China's housing price growth has shown signs of relatively stable.However, the price of housing in most areas has risen more than the rate of economic growth.The problem of excessively rapid growth, high prices and unreasonable supply structure in the short term has gradually adversely affected the socio-economic development of Hainan Province.If it is not effectively resolved, it will become the future that affects the development of the economy and the real estate industry.Therefore, a reasonable forecast of housing prices is a problem that must be solved in today's society.

ISSN 2616-5775 Vol. 2, Issue 3: 9-15, DOI: 10.25236/AJCIS.020302

2. Selection of evaluation indicators

Qualitative analysis refers to an analytical method that gives an intuitive and general expression of the essential attributes, development trends, or correlations between things based on reality and subjective and objective analysis methods.Indicators that affect housing prices include economic factors such as gross national product, inflation, and interest rates.Secondly, it also includes indicators of social factors such as population, social stability, and real estate investment.

According to the above discussion, the relationship between each indicator and the price of a commercial house can be divided into a positive correlation and a negative correlation.Among them, the indicators that are positively related to the price of commercial housing include eight positive feedback indicators: urbanization level, community size, construction facility cost, economic development, land price, residential investment, gross national product, and inflation. According to the literature, the indicators that are negatively correlated with the price of commercial housing are: land scarcity, housing supply, household income, real estate speculation, housing supply, and interest rate.

Among them, the housing system and policy system include the real estate tax system, the policy housing financial system, the low-rent housing system and the real estate market regulation policy.Currency-based regulation of the real estate market, interest rate policy, financial supervision, etc., in addition to focus on stabilizing social and psychological expectations, especially in the context of inflation, the appropriate tightening of monetary policy, which effectively support the consumer psychology, to ensure good market operation.

To sum up, the price of commercial housing involves a wide range of factors, which are based on five aspects: natural, economic, social, administrative and other factors.The complexity of the factors affecting residential prices is reflected in the impact of residential prices on both supply and demand.

Therefore, the dependent variable is confirmed based on the selected indicator above. x_1 For gdp, x_2 For urban per capita disposable income, x_3 For the urban population, x_4The amount of investment for real estate development, x_5For the rate of urbanization, x_6 For inflation, x_7 For interest rates, x_8 For residential land variability prices, x_9 For psychological expectations, x_{10} For financial expenditure, x_{11}In order to use foreign capital, there are 11 factors in total.

3. Commercial housing price model based on partial least squares regression

3.1 Model preparation

(1) Standardization of data processing

ISSN 2616-5775 Vol. 2, Issue 3: 9-15, DOI: 10.25236/AJCIS.020302

Assume that there are dependent variables y and m independent variables $x_1, x_2, \ldots x_m$, the total number of samples is n, x_j The mean and standard deviation are $\bar{x}_j$, s_j. The mean and standard deviation of y are $\bar{y}$, s_y.

Normalize the data to obtain an independent variable matrix V_0 and dependent variable matrix W_0

$$V_0 = \left[\frac{x_{ij} - \bar{x}_j}{s_j} \right], \quad W_0 = \left[\frac{y_i - \bar{y}}{s_y} \right]$$

(2) Extracting the main component

From V_0 Extracting the first principal component $t_1 = V_0 w_1$, in t_1 Carried on V_0 with W_0 Return

$$V_0 = t_1 p_1 + V_1 \quad W_0 = t_1 r_1 + W \tag{1}$$

In the formula, p_1, r_1 For the regression coefficient, the residual matrix is

$$V_1 = V_0 - t_1 p_1 \quad W_1 = w_0 - t_1 r_1 \tag{2}$$

(3) Check convergence

Residual matrix V_1 with W_1 instead V_0 with W_0 If the absolute value of the element in the residual matrix tends to zero, then the accuracy of the regression equation is considered to be satisfactory.If the condition is not met, continue to use the residual matrix V_1 with W_1 instead V_0 with W_0 Repeat the above steps and get h principal components $\rho = t_1 t_2, \ldots, t_m$

3.2 Calculation method of partial least squares regression analysis

After implementing the above steps, get

$$W_0 = t_1 r_1 + t_2 r_2 + \cdots t_m r_m \tag{3}$$

Which is

$$y^* = c_0 + c_1 x_1^* + c_2 x_2^* + \cdots c_n x_n^* \tag{4}$$

In the middle x_i^* the regression coefficient is $a_i = \sum_{h=1}^{m} h w_{hi}^*$, w_{hi}^* for w_h^* The i-th component

3.3 Cross-validity principle

When the principal component is extracted for X at the 0th time, the squared error of the prediction error of y and the sum of the squares of the error of y are obtained.

ISSN 2616-5775 Vol. 2, Issue 3: 9-15, DOI: 10.25236/AJCIS.020302

$$PRESS(h)=\sum_{i=1}^{n}\left(y_i-\hat{y}_h(-i)\right)^2 \quad SS(h)=\sum_{i=1}^{n}\left(y_i-\hat{y}_h(i)\right) \tag{5}$$

Cross-validity

$$Q_h^2=1-\frac{PRESS(h)}{SS(h-1)} \tag{6}$$

If you find it in step k $Q_h^2<1-0.95^2$, then the accuracy of the model is considered to be required; otherwise, the k+1th step should be performed.

4. Example simulation analysis

4.1 Standardization of data processing

Take a city as an example, according to the relevant data, obtain the data of the region for nearly ten years, and use the annotation processing formula.

$$x^*=\frac{x-\min}{\max-\min} \tag{7}$$

The processing results are shown in Table 1.Remove the maximum value of each indicator to get the virtual optimal indicator.

Table 1 Normalization of indicators

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
2008	0.0	0.0	0.0	0.0	1.0	0.8	0.1	0.0	1.0	0.7	0.8
2009	0.1	0.0	0.1	0.1	1.0	1.0	1.0	0.1	0.0	0.8	0.7
2010	0.1	0.1	0.2	0.3	0.8	0.0	0.1	0.2	0.0	0.9	0.6
2011	0.2	0.2	0.3	0.5	0.4	0.7	0.5	0.3	0.0	1.0	0.4
2012	0.4	0.4	0.4	0.6	0.3	0.9	0.0	0.4	0.0	0.5	0.8
2013	0.5	0.5	0.4	0.7	0.1	0.5	0.0	0.5	0.0	0.6	0.3
2014	0.7	0.6	0.5	0.8	0.1	0.5	0.4	0.8	0.0	0.7	1.0
2015	0.8	0.8	0.6	0.9	0.1	0.4	0.3	1.0	0.0	0.4	0.9
2016	0.9	0.9	0.5	0.9	0.0	0.2	0.6	1.0	0.0	0.4	0.1
2017	1.0	1.0	1.0	1.0	0.3	0.5	0.3	1.0	0.0	0.0	0.9

4.2 Score situation

According to the method of the section in this article, using matlab programming, when proposing two principal components t_1-t_2, Weights w_h Characteristic value w_h^* The scores are shown in Table 2.

ISSN 2616-5775 Vol. 2, Issue 3: 9-15, DOI: 10.25236/AJCIS.020302

Table 2 Score situation

	r_6	r_8	r_6	r_8
x_1	0.15	0.24	0.59	-0.01
x_2	-0.44	-0.16	-0.39	0.08
x_3	0.08	0.04	-0.81	0.01
x_4	0.1	-0.18	0.31	-0.16
x_5	-0.5	0.41	-0.46	-0.05
x_6	0.04	0.17	0.07	-0.07
x_7	0.24	-0.23	0.67	0.21
x_8	0.25	0.74	0.27	0.25
x_9	0.52	-0.14	0.7	-0.24
x_{10}	-0.01	-0.23	-0.96	-0.03
x_{11}	0.34	0.15	0.67	0.02

Calculated, Q_2^2 The principle of cross-validation is satisfied, so the two principal components extracted are sufficient.Finally, the normalized regression coefficient is converted into the regression coefficient of the original data, which is obtained as shown in Table 3.

Table 3 Regression coefficient

Independent variable	Standardization coefficient	Original regression coefficient
X_1	2735.03028685	1.00000000
X_2	0.00475467	0.66893600
X_3	38.13358531	0.67355136
X_4	-0.00175945	0.66893515
X_5	0.00014458	0.66893539
X_6	-5526.28817583	0.00000000
X_7	48.10888836	0.67475880
X_8	-577.38676310	0.59904497
X_9	-0.04504905	0.66892995
X_{10}	-0.00079810	0.66893527
X_{11}	0.00004914	0.66893539

Get the original regression equation as

$$y = 2735x_1 + 38x_3 - 5526x_6 + 48x_7 - 577x_8$$

ISSN 2616-5775 Vol. 2, Issue 3: 9-15, DOI: 10.25236/AJCIS.020302

4.3 Model fitting situation

According to the above model, the original data is predicted and analyzed. The fit between the actual dependent variable value and the predicted value is shown in Figure 1.

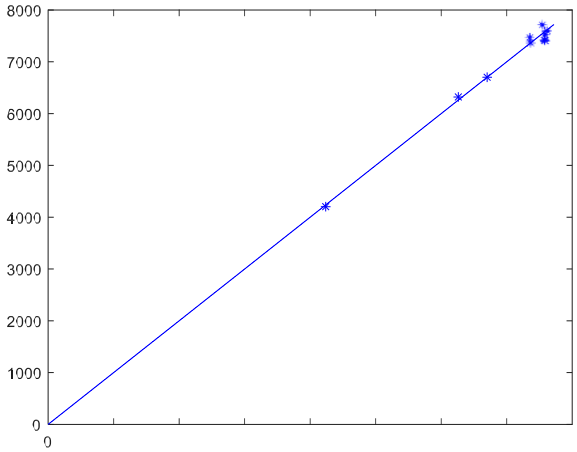

Figure. 1 Commodity house price forecast chart

Figure 2 is a histogram of regression coefficients

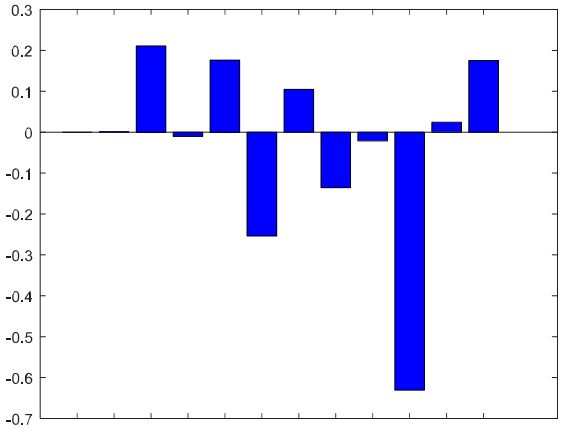

Figure. 2 regression coefficient histogram

ISSN 2616-5775 Vol. 2, Issue 3: 9-15, DOI: 10.25236/AJCIS.020302

5. Conclusion

It can be seen from Fig. 1 and Fig. 2 that the prediction result of the model is less than the true value, the equation fitting coefficient is 0.88, and the fitting effect is good, so the model is accurate.However, when the index of the model is selected, there is no horizontal comparison between the indicators and the weight of each indicator changes with the change of the sample. The weight depends on the sample and is limited in application.Therefore, in the real prediction, it must be combined with the actual situation of the local housing price macro-control policy to get more accurate predictions.

References

[1] Ding Feng. Study on influencing factors and prediction of real estate prices [D]. Anhui University of Finance and Economics, 2014.
[2] Liu Hui. Influencing factors and Forecasting Research on the price change of commercial housing in Wuhan [D]. Wuhan University of Science and Technology, 2010.
[3] Chen Senjun. Influencing factors analysis and price forecast of commodity house prices in China [D]. Huazhong University of Science and Technology, 2008.