2014

ISSN : 0974 - 7435 Volume 10 Issue 13

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(13), 2014 [7016-7024]

Real estate price forecast based on multiple linear regression analysis

Yifei Ma[1]*, Jun Ma[2]

[1]Aecom, Beijing 100024, (CHINA)

[2]Nankai University, Tianjin 300071, (CHINA)

E-mail : yifeima1998@163.com

ABSTRACT

A new real estate forecast mode is proposed to predict the right price, which is based on gray theory and multiple linear regression prediction models. Firstly, GM (1,1) model is used to predict the residential land price, average wage, per capita disposable income of urban residents. Then classical multiple linear regression model is used to predict the housing price. It can be seen from the empirical analysis, commercial housing sales price can be predicted more accurately and can provide strong evidence for real estate developers in order to make the right decision analysis.

KEYWORDS

Real estate price forecast; Gray theory; Multiple linear regression.

INTRODUCTION

Classical multiple linear regression model is a kind of commonly used multivariate statistical methods, which has clear principle, simple model and convenient application and is very widely used in the industrial and agricultural production and scientific research. For example, in medicine and health care, weather forecast, geological exploration and many other fields use the classical multiple linear regression model.

Since the housing reform in 1998, our country's real estate industry has been growing by leaps and bounds, has made a significant contribution for the people economic growth and gradually become the pillar industry of national economy in our country[1,2]. But due to a late start of the real estate market in China, combined with not perfect development, it presents the obvious ups and downs and some other more typical primary market characteristics. In this case, the government often takes some measures to guide the real estate developers to make the right investment[3,4]. A forecast comparison of residential housing prices by parametric versus semi-parametric conditional mean estimators was proposed by Gencay. R[5]. The study about the theory of the commodity residential house price influence factors and quantitative analysis was proposed by Jia Shenghua[6]. The analysis of the influence factors of urban housing price change-based on the research data of 31 cities was proposed by Zheng Yao[7]. Model to predict the real estate price index of Changchun based on ARMA was proposed by Ou Yanhao[8]. Boom-bust cycles and the forecasting performance of linear and non-linear models of house prices was proposed Miles. M[9].

The paper is organized as follows. In the next section, the principle of multiple linear regression model is investigated. In Section 3, real estate price forecast based on linear regression and gray theory[10-12] is proposed. In Section 4, in order to test the performance of proposed algorithm, GM (1,1) model is used to predict the residential land price, average wage, per capita disposable income of urban residents and classical multiple linear regression model is used to predict the housing price. Finally, we conclude our paper in section 5.

PRINCIPLE OF MULTIPLE LINEAR REGRESSION MODEL

Y is linearly related to $x_1, x_2, \ldots x_m$. $(y_t, x_{t1}, x_{t2}, \ldots, x_{tm})$ represents n groups of data, which meets the following regression model.

$$y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_m x_{tm} + \varepsilon_t (t = 1, 2, \ldots, n)$$

$$E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma^2, Cov(\varepsilon_i, \varepsilon_j) = 0 (i \neq j)$$

or $\varepsilon_t \square N(0, \sigma^2), (t = 1, 2, \ldots, n)$, which are independent of each other.

$$C = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1m} \\ 1 & x_{21} & x_{22} & \ldots & x_{2m} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nm} \end{bmatrix} = (1_n \mid X)$$

$$Y=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \beta=\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}, \varepsilon=\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

The matrix form of regression model is

$$\begin{cases} Y=C\beta+\varepsilon \\ E(\varepsilon)=0_n, D(\varepsilon)=\sigma^2 I_n \end{cases} \quad \textbf{(1)}$$

This model is called typical multivariate linear regression model. Y represents random vector, which can be observed. ε represents random vector which can not be observed. C represents known matrix. β and σ^2 represents unknown parameters. $n>m$, $rank(C)=m+1$.

REAL ESTATE PRICE FORECAST BASED ON GRAY THEORY AND LINEAR REGRESSION

Grey generation

$x_r(k^{'})$ and $x_r(k^{"})$ represent original data. $x_T(k)$ represents grey generation data. O represents operation. The general formula of grey generation is

$$A_{LGO}: x_r(k^{'})Ox_r(k^{"})=x_T(k) \quad \textbf{(2)}$$

The aim of accumulated generation is to generate new data, which is different from original regulation. $x^{(0)}$ represents extendible original sequence.

$$x^{(0)}=(x^{(0)}(1),x^{(0)}(2),\ldots,x^{(0)}(n)),$$

$$\forall x^{(0)}(k)\in x^{(0)} \Rightarrow k\in K=\{1,2,\ldots,n\},$$

$\forall k^{*}\in[k-1,K]$ has definition and K represents extendible data set. If formula (2) meets the following condition.

$$O\Rightarrow +$$

$$k^{'}=k, k^{"}=k-1$$

$$x_r(k^{'})=x^{(0)}(k),\ x_r(k^{"})=x^{(1)}(k-1),\ x_T(k)=x^{(1)}(k)$$

The formula of accumulation generation operation is

$$AGO: x^{(0)}(k) + x^{(1)}(k-1) = x^{(1)}(k) \quad \textbf{(3)}$$

The generated sequence of $x^{(0)}$ corresponding to $x^{(1)}$ is

$$x^{(1)} = AGOX^{(0)}. \quad \textbf{(4)}$$

Formula (4) can be expressed as follows.

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n))$$

$$\forall x^{(1)}(k) \in x^{(1)} \Rightarrow k \in K = \{1, 2, \ldots, n\}$$

$$x^{(1)}(k) = x^{(1)}(k-1) + x^{(0)}(k), \forall k \in K^{'} \subset K$$

$K^{'} = \{2, 3, \ldots, n\}$, $x^{(1)}(1) = x^{(0)}(1)$.

If $x^{(0)}$ is original sequence,

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)),$$

$$x^{(1)} = AGOX^{(0)},$$

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n)),$$

then $\forall x^{(1)}(k) \in x^{(1)} \Rightarrow k \in K = \{1, 2, \ldots, n\}$

$$x^{(1)}(k) = \sum_{m=1}^{k} x^{(0)}(m)$$

GM (1,1) model

An albino background value is called $z^{(1)}(k)$ and definition of grey model GM (1,1) is

$$x^{(0)}(k) + az^{(1)}(k) = b \quad \textbf{(5)}$$

An albino background value sequence is

$$z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \ldots, z^{(1)}(n)),$$

$$z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1).$$

Each value is mean of $x^{(1)}(k)$ and $x^{(1)}(k-1)$, so $z^{(1)}$ is labeled as $MEANx^{(1)}$.

$$z^{(1)} = MEANx^{(1)},$$

$\forall z^{(1)}(k) \in z^{(1)} \Rightarrow k \in K^{'} = \{2,3,\ldots,n\}, n \geq 4$,

$\forall x^{(0)}(k) \in x^{(0)} \Rightarrow k \in K^{'} = \{1,2,3,\ldots,n\}, n \geq 4$.

The first level parameter packet is (a,b) labeled as P_1, which has the following formula under least square criterion.

$$P_1 = \begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T y_N \quad \textbf{(6)}$$

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix},\ y_N = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}.$$

Above all, the parameter packet of GM (1,1) is as follows.

The second level parameter packet is $P_1 = (a,b)$.

$$a = \frac{CD-(n-1)E}{(n-1)F-C^2},\ b = \frac{DF-CE}{(n-1)F-C^2} \quad \textbf{(7)}$$

The second level parameter packet is $P_2 = (C,D,E,F)$.

$$C = \sum_{k=2}^{n} z^{(1)}(k),\ D = \sum_{k=2}^{n} x^{(0)}(k),\ E = \sum_{k=2}^{n} z^{(1)}(k)x^{(0)}(k),\ F = \sum_{k=2}^{n} z^{(1)}(k)^2. \quad \textbf{(8)}$$

The prediction process of housing sales price is as follows.

(1)Generate AGO in accumulation way.

(2)Calculate $MEAN$.

(3)Calculate the second level parameter packet of the model GM (1,1).

(4)Calculate the first level parameter packet of the model GM (1,1).

(5)Set up model GM (1,1).

(6)Residual test on the model.

(7)Carry out prediction based on multiple linear regression model.

THE EMPIRICAL ANALYSIS OF REAL ESTATE PRICE FORECAST

Based on the grey system theory and multivariate linear regression analysis, a real estate price forecast model is proposed. First of all, the new time series values are predicted by the grey system GM (1,1) model to filter the random disturbance contained by original time series. On this basis, the classical multiple linear regression analysis is carried out to establish prediction model. It also can effectively solve multicollinearity problem of the classical multiple linear regression analysis, thus gives more accurate change trend of response variable. The commodity housing sales price of some city in China from statistical bulletin for national economic and social development is shown in TABLE 1. GM (1,1) model is used to predict the residential land price, average wage, urban per capita disposable

income, commercial housing sales price of this city from year 2006 to 2007. The unit of land price is yuan per square meter and the unit of wage is yuan per year.

TABLE 1 : The commodity housing sales price

Year	The residential land price	Average wage	Urban per capita disposable income	Commercial housing sales price
1997	837	6153	3923	956
1998	852	6922	4057	982
1999	893	7764	4529	1097
2000	965	9179	5128	1548
2001	1067	9863	5484	1732
2002	1149	10749	6331	1861
2003	1152	11561	6806	1879
2004	1273	13423	7492.5	2046
2005	1369	15780	8272	3118

The residential land price is predicted as follows.

(1) $$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(9)) = (837, 852, 893, 965, 1067, 1149, 1152, 1273, 1369)$$

(2) AGO $x^{(1)} = AGOx^{(0)}$

$$x^{(1)}(k) = x^{(1)}(k-1) + x^{(0)}(k)$$

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(9)) = (837, 1689, 2582, 3547, 4614, 5763, 6915, 8188, 9557)$$

(3) $z^{(1)} = MEANx^{(1)}$

$$z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1)$$

$$z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \ldots, z^{(1)}(9)) = (1263, 2135.5, 3064.5, 4080.5, 5188.5, 6339, 7551.5, 8872.5)$$

(4) $P_2 = (C, D, E, F)$

$$C = \sum_{k=2}^{9} z^{(1)}(k) = z^{(1)}(2) + z^{(1)}(3) + \ldots + z^{(1)}(9) = 1263 + 2135.5 + \ldots + 8872.5 = 38495$$

$$D=\sum_{k=2}^{9}x^{(0)}(k)=x^{(0)}(2)+x^{(1)}(3)+\ldots+x^{(1)}(9)$$
$$=852+893+\ldots+1369$$
$$=8720$$

$$E=\sum_{k=2}^{9}z^{(1)}(k)x^{(0)}(k)=z^{(1)}(2)x^{(0)}(2)+z^{(1)}(3)x^{(0)}(3)+\ldots+z^{(1)}(9)x^{(0)}(9)$$
$$=45317840$$

$$F=\sum_{k=2}^{9}z^{(1)}(k)^{(2)}=235050000.5$$

(5) $$a=\frac{CD-(n-1)E}{(n-1)F-C^2}$$
$$=-0.0674$$

$$b=\frac{DF-CE}{(n-1)F-C^2}=765.6292$$

GM (1,1) model is as follows.

$$x^{(0)}(k)-0.0641z^{(1)}(k)=1015.1$$

An albino responsive model is

$$\hat{x}^{(1)}(k+1)=(x^{(0)}(1)-\frac{b}{a})e^{ak}+\frac{b}{a}$$

$$\hat{x}^{(0)}(k+1)=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k)$$

(6)Residual test is carried out.

$\hat{x}^{(0)}(1)=837$, $x^{(0)}(1)=837$, $e^{(0)}(1)=0\%$.

$\hat{x}^{(0)}(2)=850$, $x^{(0)}(2)=852$, $e^{(0)}(2)=2.3\%$.

$\hat{x}^{(0)}(9)=1362$, $x^{(0)}(9)=1369$, $e^{(0)}(9)=0.51\%$.

$$e^{(0)}(avg)=\frac{1}{9}\sum_{k=1}^{9}\left|e^{(0)}(k)\right|=1.11\%$$

$$p^{(0)}=(100-e^{0}(avg))\%=99.9889\%$$

The accuracy of model achieves 99.9889%.

$$\hat{x}^{(0)}(10) = \hat{x}^{(1)}(10) - \hat{x}^{(1)}(9)$$
$$(7) = (12196.5 \cdot e^{0.0674 \cdot 9} - 11359.5) - (12196.5 \cdot e^{0.0674 \cdot 8} - 11359.5)$$
$$= 1458.4$$

Residential land price in 2006 is 1458.4 yuan per square meters and in the same way, residential land price in 2007 is 1560 yuan per square meters. The prediction result of land prices, average wage and per capita income is shown in TABLE 2.

TABLE 2 : land prices, average wage and per capita income

Year	land prices	average wage	per capita income
2006	1458.4	17059	9173
2007	1560	19080	10136

Land prices, average wage and per capita income from year 1997 to 2005 are taken as variables, and average housing sales price from year 1997 to 2005 are taken as responsible variables. EVIEWS is used to calculate regression coefficients. At last, the predictive values and regression coefficients are input to the proposed model.

$$\hat{Y}(t) = \hat{\beta}_0 + \hat{\beta}_1 \hat{X}_1(t) + \ldots + \hat{\beta}_{P-1} \hat{X}_{P-1}(t)$$

Housing sales price of year 2006 and 2007 are shown in TABLE 3.

TABLE 3 : Predictive housing sales price of year 2006 and 2007

	2006	2007
Actual value	3258	3579
Predictive value	3202	3506
Relative error	0.0172	0.0203

Finally, carry out F check for the above model.

$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$, $H_1 : H_0$ is not true.

$F = \dfrac{MSS / 3}{ESS / (9-3-1)} \sim F(3,5)$, MSS represents regression square sum and ESS represents residual square sum. For given $\alpha = 0.05$, $P(F > F_{0.05}(3,5)) = 0.05$. After calculation $F = 237.2 > 216$, when $\alpha = 0.05$, this model is remarkable and the check passes, meaning the model is efficient.

CONCLUSION

This paper presents a new real estate forecast model, which is based on gray theory and multiple linear regression prediction models to predict the right price. Firstly, GM (1,1) model is used to predict the residential land price, average wage, per capita disposable income of urban residents. Then classical multiple linear regression model is used to predict the housing price. The results show that this method can be used to help the realtors forecasting commodity housing sales average price in order to make the right decision analysis.

REFERENCES

[1] C.L.Lee; "Housing price volatility and its determinants," International Journal of Housing Markets and Analysis, **2(3)**, 293-308 **(2009)**.

[2] H.Lind; "Price bubbles in housing markets: concept, theory and indicators," International Journal of Housing Markets and Analysis, **2(1)**, 1753-8270 **(2009)**.

[3] W.McCluskey; "Predictive accuracy of machine learning models for the mass appraisal of residential property," New Zealand Valuers'Journal, July, 7-41 **(1996)**.

[4] D.E.Rapach, J.K.Strauss; "Differences in housing price forecastability across US states," International Journal of Forecasting, **25(2)**, 351-72 **(2009)**.

[5] R.Gencay, X.Yang; "A forecast comparison of residential housing prices by parametric versus semi-parametric conditional mean estimators", Economics Letters, **52**, 129-35 **(1996)**.

[6] Jia Shenghua; "The study about the theory of the commodity residential house price influence factors and quantitative analysis," Renmin university of China, **(2006)**.

[7] Zheng Yao, Huaping Sun, Changchun Feng; "The analysis of the influence factors of urban housing price change-based on the research data of 31 cities," The economic construction, **(2011)**.

[8] Ou Yanhao; "Model to predict the real estate price index of Changchun based on ARMA," Northeast University of finance, **(2003)**.

[9] M.Miles; "Boom-bust cycles and the forecasting performance of linear and non-linear models of house prices," Journal of Real Estate Finance & Economics, **32**, 49-264 **(2008)**.

[10] Chen Jintao; "The Application of Optimized Grey Model in the Load Forecasting of Power System," Journal of Nanjing Institute of Technology, **1(4)**, 1-5 **(2003)**.

[11] Zhao Chengwang, Gu Xingsheng; "Application of Combination Optimized Grey Model in Long-medium Term Power Load Forecasting," Proceeding of the 24th Chinese Control Conference, 1539-1543 **(2005)**.

[12] Tang Zhenjun; "Gray Model Based on Improved Medium and Long Term Load Forecasting," Journal of Anhui Electrical Engineering Professional Technique College, **15(4)**, 63-66 **(2010)**.