Real estate price prediction in Nanning based on grey model

Hao Cheng [*a], Jian Yang [b], Chengyuan Fan [a], Hengfeng Xiang [c], Wenli Zhou [a], Ren Lin [b]

[a*]School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin, Guangxi, People's Republic of China, 541004
[b]School of Mathematics and Computing Science, Guilin University of Electronic Technology, Guilin, Guangxi, People's Republic of China, 541004
[c]School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin, Guangxi, People's Republic of China, 541004

*Corresponding author: 916988974@qq.com

Abstract. With the rapid development of society and economy, the development of the real estate industry has always been in a leading position, but the change of housing prices has led to housing problems affecting every Chinese people. This paper predicts the trend of real estate prices in Nanning in the next 10 years. Firstly, a grey model is established, the simulation sequence is obtained by substituting the serial number of the relevant year, and the predicted value is compared with the original value by the posterior difference test method. Then, a grey GM(1, 1) model is established for the four main factors affecting housing prices respectively to predict the statistical values of these four factors in the next ten years. For the overfitting phenomenon existing in the prediction of some factors, linear regression fitting is used to predict the data in the next ten years after verifying the correlation. Finally, GM(0, N) estimation formula is used to forecast the real estate price in Nanning in the next ten years.

Keywords: linear regression moving average method GM(0,N) model posterior difference test

1 Introduction

With the rapid social and economic development, the development of the real estate industry has been in a leading position, but the change of housing prices has led to housing problems affecting every Chinese people. Since the real estate industry has become the core of China's economic development, housing prices have shown a rising trend, but due to the impact of the epidemic in the past three years, housing prices have also been changing[1]. There is a close relationship between the stability of real estate prices and the healthy development of the real estate market and social economy is particularly important. As the capital of Guangxi Autonomous Region, Nanning is one of the key developing cities in southwest China, and the stable development of its real

L. Moutinho et al. (eds.), *Proceedings of the 2023 International Conference on Management Innovation and Economy Development (MIED 2023)*, Advances in Economics, Business and Management Research 260,
https://doi.org/10.2991/978-94-6463-260-6_72

estate market is more important in the overall development of Southwest China. Therefore, the analysis and prediction of housing prices can provide effective help and development ideas for the development and construction investment of the real estate market, so as to further ensure the healthy and stable development of the economy[2].

This paper analyzes the historical data, tries to analyze the trend of real estate prices in Nanning in the next ten years, and gives a reasonable forecast method.

2 Model assumptions and symbol description

1. Assume that house buyers have no preference for the surrounding environment of the house and ignore the psychological expectations of house buyers for housing prices;
2. Assume that the housing cost is mainly reflected in the transaction price of the developer's land, ignoring the impact of some supporting facilities on the housing cost;
3. Assume that housing supply and demand are stable;
4. The impact of data mutation caused by mutation of other secondary factors is ignored.

Table 1. Symbol specification

Symbol	Description
t	time
X_t	Data at time point t
SMA_t	Simple moving average at point t in time
Y	Reference sequence
X_i	Comparison sequence
ξ_i	Grey coefficient
a	Minimum difference between two poles
b	Maximum difference between two poles

3 Establishment and solution of the model

3.1 Establishment of grey GM(1,1) model

GM(1,1) model is one of the most widely used grey dynamic prediction models in grey system theory, which is composed of a single variable first order differential equation. It is mainly used to fit and predict the eigenvalue of a dominant factor in complex systems, so as to reveal the change rule of the dominant factor and the future development trend[3].

In order to ensure the feasibility of the modeling method, it is necessary to check the known sequence. Let the reference data is $x^{(0)}=(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n))$, Calculate the level ratio of the series $\lambda(k)$

$$\lambda(k)=\frac{x^{(0)}(k-1)}{x^{(0)}(k)},\ k=2,3,\cdots,n \tag{1}$$

Logarithm series $x^{(0)}$ Do the necessary transformation processing to make it fall within the tolerable coverage $\left(e^{-\frac{2}{n+1}},e^{\frac{2}{n+2}}\right)$. Take the appropriate constant c, do a translation transformation. If all stage ratios fall within tolerable coverage, the series can be grey predicted as data for the model GM(1,1).

$$y^{(0)}(k)=y^{(0)}(k)+c\ ,\ k=1,2,\cdots,n \tag{2}$$

Then make the sequence of numbers $y^{(0)}=(y^{(0)}(1),y^{(0)}(2),\cdots,y^{(0)}(n))$ level ratio is:

$$\lambda(k)=\frac{y^{(0)}(k-1)}{y^{(0)}(k)},\ k=2,3,\cdots,n \tag{3}$$

Let the reference number corresponding to the fourth influencing factor be listed

$$x_i^{(0)}=(x_i^{(0)}(1),x_i^{(0)}(2),\cdots,x_i^{(0)}(n)) \tag{4}$$

Find the mean series:

$$z_i^{(1)}(k)=0.5x_i^{(1)}(k)+0.5x_i^{(1)}(k-1)\ ,\ k=2,3,\cdots,n \tag{5}$$

Grey differential equation is established:

$$x_i^{(0)}(k)+az_i^{(0)}(k)=b\ ,\ k=2,3,\cdots,n \tag{6}$$

$u=(a,b)^T$, $Y=(x_i^{(0)}(2),x_i^{(0)}(3),\cdots,x_i^{(0)}(n))^T$, $B=\begin{bmatrix}-z_i^{(1)}(2) & 1\\ -z_i^{(1)}(3) & 1\\ \vdots & \vdots\\ -z_i^{(1)}(n) & 1\end{bmatrix}$,

By least square method $u=(a,b)^T=(B^TB)^{-1}B^TY$

The whitening differential can then be solved to obtain:

$$\hat{x}_i^{(0)}(k+1)=\left(x_i^{(0)}(1)-\frac{b}{a}\right)e^{-ak}+\frac{a}{b}\ ,\ k=1,2,\cdots,n \tag{7}$$

The calculation results of the above formula can be obtained by reducing and restoring them:

$$\hat{x}_i^{(0)}(k+1)=\hat{x}_i^{(0)}(k+1)-\hat{x}_i^{(0)}(k)\ ,\ k=1,2,\cdots,n \tag{8}$$

3.2 Model solving

Considering the main factors of Nanning's urban per capita disposable income, gross regional product, commercial housing sales area and registered population at the end of the year, the system characteristic variable $x_i^{(0)}(i=1,2,3,4,5)$ is set as the average sales price of commercial housing (housing price), urban per capita disposable income of Nanning, gross regional product, commercial housing sales area and registered population at the end of the year, respectively[4].

Among them, the average sales price of commercial housing (housing price) is the main behavior of the system, and each variable takes the data of 20 years as the relevant factor sequence. Namely

$$x_i^{(0)}=(x_i^{(0)}(1),x_i^{(0)}(2),\cdots,x_i^{(0)}(20)) \tag{9}$$

Plugging the data into the formula gives us:

$$\begin{cases} x_1^{(0)}=(2252,2761.11,2605.03,\cdots) \\ x_2^{(0)}=(192.2,333.67,455.72,\cdots) \\ x_3^{(0)}=(641.67,648.85,659.54,\cdots) \\ x_4^{(0)}=(502.53,588.86,723.36,\cdots) \\ x_5^{(0)}=(9162,8059,9203,\cdots) \end{cases} \tag{10}$$

In addition to the above formula, the differential equation of variable in whitening form is established

$$x_1^{(1)}=a+b_2x_2^{(1)}(k)+b_3x_3^{(1)}(k)+\cdots+b_{19}x_{19}^{(1)}(k) \tag{11}$$

Among them,

$$B = \begin{bmatrix} 169.3 & 641.67 & 502.53 & 9162 \\ 478.29 & 1290.52 & 1091.39 & 17221 \\ 898.28 & 1950.06 & 1814.75 & 26424 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \tag{12}$$

$$Y = [2252 \quad 5013.11 \quad 7618.14 \quad \cdots]^T \tag{13}$$

Plug in the parameter column $u = [a, b_2, \cdots, b_{19}]$ The least squares estimation formula is obtained:

$$u = [0.2752 \quad -0.8038 \quad 0.2158 \quad -1.4648 \quad 0.27] \tag{14}$$

Therefore, there is the GM (0, N) estimator:

$$\hat{x}_1^{(1)}(k) = -0.8038x_2^{(1)}(k) + 0.2158x_3^{(1)}(k) - 1.4648x_4^{(1)}(k) + 0.27x_4{}^{(1)}(k) + 0.2752 \tag{15}$$

Substitute the above estimator to calculate the estimated value. The simulation results and errors are shown in the following table 2:

Table 2. Actual and forecast prices

year	True value	Predicted value	Relative error
2003	2252	2252	0
2004	2761.11	479.7199	2281.39
2005	2605.03	2098.02	507.0098
2006	2872.42	3185.54	-313.12
⋮	⋮	⋮	⋮
2018	7782.17	7268.34	513.8301
2019	8406	9332.712	-926.712
2020	8605	8786.199	-181.199
2021	8301	8635.255	-334.255

Then we forecast the main factors that affect house prices.

Taking the registered population $x_1^{(0)}$ at the end of the year as an example, we forecast the housing price in the next 10 years.

Given the original data is $x_1^{(0)} = (641.67, 648.85, 659.54, \cdots)$, the given original data is accumulated to obtain a new cumulative series $x_1^{(0)} = (641.67, 1290.52, 1950.06, \cdots)$.

The prediction function $\hat{x}_1^{(0)}(k+1) = \hat{x}_1^{(1)}(k+1) - \hat{x}_1^{(1)}(k)$ is established and the predicted value of the known data is solved for 20 years. Here's how the actual data compares to the forecast[5].

Table 3. Actual and projected population

year	True value	Predicted value	Relative error
2003	641.67	641.67	0
2004	648.85	656.1269	-7.27692
2005	659.54	663.5605	-4.02049
2006	671.89	671.0783	0.811722
⋮	⋮	⋮	⋮
2018	770.82	768.2192	2.60085
2019	781.97	776.9227	5.047335
2020	793.1103	785.7248	7.385557
2021	784.8765	794.6266	-9.75009

The calculated results in Table 3 were substituted into the two test criteria respectively to conduct a posterior difference test. The specific results are shown in Table 4:

Table 4. Accuracy test table

	p	C	Overall accuracy class
Index value	0.9831	0.3306	Level 1 (good)
Level	Level 1 (good)	Level 1 (good)	

The accuracy of the model is good and can be predicted.

Set k = 20,21,... The estimate for the next 10 years is obtained by substituting it into the formula and reducing it.

The following is the index value of the other three major factors that can also be predicted in the next 10 years according to the above process. The forecast data of the housing price and its main influencing factors in the next 10 years are summarized as follows:

Table 5. Prediction of influencing factors

year	Sales area	population	Gross product	income
2022	1799.779	803.6293	5260.609	42504.96
2023	1887.308	812.734	5530.486	44463.9
2024	1974.838	821.9419	5800.363	46422.83

2025	2062.367	831.254	6070.24	48381.76
2026	2149.897	840.6717	6340.117	50340.7
2027	2237.426	850.1961	6609.994	52299.63
2028	2324.956	859.8283	6879.871	54258.56
2029	2412.485	869.5697	7149.748	56217.5
2030	2500.015	879.4215	7419.624	58176.43
2031	2587.544	889.3849	7689.501	60135.36

Figure 1 is a comparison of population projections with actual data:

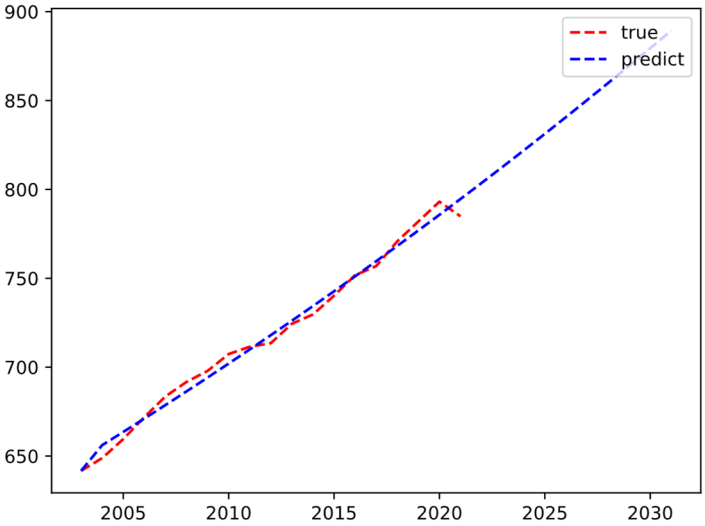

Fig. 1. Comparison of population projections with real data

The accuracy of the model obtained by the posterior difference test is shown in the table 6:

Table 6. Accuracy test

	p	C	Overall accuracy class
Index value	0.9632	0.3201	Level 1 (good)
Level	Level 1 (good)	Level 1 (good)	

The accuracy of the model is good and can be predicted.

The forecast data of the major influencing factors in Table 5 for the next 10 years are substituted into the GM (0, N) estimator we have obtained:

$$\hat{x}_1^{(1)}(k) = -0.8038x_2^{(1)}(k) + 0.2158x_3^{(1)}(k) - 1.4648x_4^{(1)}(k) + \cdots + 0.2752 \tag{16}$$

It can be seen that under the influence of different relevant factors in the next 10 years, the change trend of housing prices is as follows figure 2:

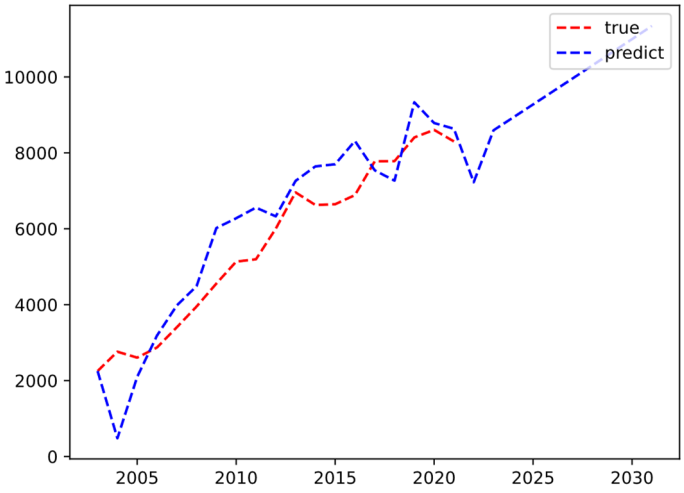

Fig. 2. Comparison of forecast and actual data of house prices

Through the posterior difference test method, the accuracy of the model obtained is as follows table 7:

Table 7. Accuracy test table

	p	C	Overall accuracy class
Index value	0.8571	0.4947	Level 2 (Pass)
Level	Level 2 (Pass)	Level 2 (Pass)	

The accuracy level of the model is qualified and can be predicted.

At the same time, with respect to the trend, except for some years, the predicted trend is generally consistent with the actual trend. This paper predicts that the housing price of Nanning will continue to rise in the next ten years.

4 Conclusion

This paper forecasts the trend of real estate prices in Nanning in the next 10 years. Firstly, a grey model is established, the simulation sequence is obtained by substituting

the serial number of the relevant year, and the predicted value is compared with the original value by the posterior difference test method. Then, a grey GM (1, 1) model is established for the four main factors affecting housing prices respectively to predict the statistical values of these four factors in the next ten years. For the overfitting phenomenon existing in the prediction of some factors, linear regression fitting is used to predict the data in the next ten years after verifying the correlation. Finally, GM (0, N) estimation formula is used to forecast the real estate price in Nanning in the next ten years. This paper predicts that the housing price in Nanning will continue to rise in the next ten years.

References

1. Chen Qingxin, Wen Qinglan, Lu Haihua. Analysis of influencing factors of housing price based on grey correlation degree [J]. Modern Marketing (Management Edition), 2019, No.317(05):66-68.
2. Tang Zhaoguang. Grey correlation degree analysis of influencing factors of housing price -- A case study of Nanning City [J]. Management and Technology of Small and Medium-sized Enterprises (Mid-10 days), 2015, (03):183-185.
3. Sun Shizhong. Discussion on highway toll revenue forecasting method based on moving average and GM (1,1) grey model [J]. Statistics and Management, 2018, No.257(12):16-21.
4. Xie Zhongdarang, Wei Yinkai. Research on rural land asset price prediction based on GM (0,N) and GM(1,1) models [J]. Journal of Intelligent Agriculture, 2019,1(02):1-10.
5. Xu Ruoning. Linear regression model fitting fuzzy observation data [J]. Pure and Applied Mathematics,1997, (02).