Investigation and comparison between GM(1,1) and BPANN forecast models in Shanghai low-rent housing families

Zhuo Li
The Research Center for East-West Cooperation in China,
East China Normal University, 200062
Shanghai, People's Republic of China
lzecnu@163.com

Jianhua Xu
The Research Center for East-West Cooperation in China,
East China Normal University, 200062
Shanghai, People's Republic of China
jhxu@geo.ecnu.edu.cn

Qing Wei
Shanghai Civil Affairs Bureau, 200002
Shanghai, People's Republic of China
davidweilaw@yahoo.com.cn

***Abstract*—Based on the data of household income of Shanghai low-rent housing families, a GM(1,1) forecast model and a Back-Propagation Artificial Neural Network (BPANN) forecast model are established respectively to predict the average household income of low-rent housing families. The comparison between the GM(1,1) and the BPANN model showed that the BPANN model is better than the GM(1,1) model at the aspects of prediction accuracy and data adaptability. The BPANN model could be applied successfully to predict the average household income of Shanghai low-rent housing families in a short-term and it will provide scientific and effective basis for formulate policy on low-rent housing.**

Keywords-Low-rent housing families; GM(1,1); BPANN; prediction model; comparison

I. INTRODUCTION

Low-rent housing is the social security housing which is provided by the government to urban residents who live below the minimum living security standard line and have housing problems. The policy was usually implemented in two ways: providing houses for rent or providing housing subsidy. To some extent, low-rent housing system not only provides a policy-supporting for the low-income families who have the difficulties in the housing, but also alleviates the house price rising excessively. It also helps adjusting the interest relations among social classes, promoting the social justice and maintaining the social stability. Discussing current low-rent housing families' income, we could find out its internal rules and then forecast the future trend, which not only help the government to formulate scientific policies, but also explore more comprehensive low-rent housing exit mechanisms, strength the government's actual control of city low rent houses.

In low-rent housing research area, there are few researches and mainly focus on the families' living conditions and policies' problems in qualitative analysis, lacking of data supporting [1,2,3]. But, there are many mature methods in the low-income groups research area, for example: time series method [4,5], regression method [6], markov method [7,8]. These methods discuss the prediction problem from different perspectives, but they are based on traditional prediction process, requiring high quality simple data. In recent years, with the development of intelligence methods such as grey method [9,10] and artificial neural network [11,12], many shortcomings of traditional methods to be overcome. They have fewer requirements on the data quality and quantity. Meanwhile, they are performed as a black box without considering the internal structure. They are widely applied in business and scientific research areas. As an important component part of low-income groups, we select GM(1,1) model and Back-Propagation Artificial Neural Network (BPANN) model to predict the household income of Shanghai low-rent housing families. By comparison, we select a more appropriate model to quantitatively describe the future trend.

II. MATERIALS AND METHODS

A. Materials

Shanghai is one of the first cities to implement low-rent housing policy in China, since "Interim Measures of Shanghai urban low-rent housing" issued in 2000, a low-rent housing supply system has been established in Shanghai. And in 2008, the "residents income verification system" has operated in 160 streets, which provide a clear and scientific evidence for the low-rent housing policy. This paper counts the household income of low-rent housing families from December 2008 to February 2010, a total of 15 months. Based on the statistics, GM(1,1) model and BPANN model are established to predict the average household income of Shanghai low-rent housing families.

In order to construct prediction model and evaluate model accuracy at the same time, low-rent housing household income data are divided into two parts, one part (from December 2008-December 2009, 86.7% of the entire sample) is used to construct the prediction model, and the other (from January

978-1-4244-7941-2/10/$26.00 ©2010 IEEE

2010 to February 2010, 13.3% of the entire sample) is used to evaluat the model accuracy.

B. Methods

1) GM(1,1) Model

GM(1,1) model is the basic model of grey system theory dealing with complex system who lacks information or characteristics are only partially known or known with uncertainty. It converts the original sequence data into new data series in order to weak the randomness and enhance its regularity by AGO transformation (Accumulated Generating Operation), then fit the forecast model by a first single-variable equations. The prediction values are obtained by returning an AGO's level to the original level using IAGO (Inverse Accumulated Generating Operation). The information of household income in Shanghai low-rent housing families is insufficient and low reliability, so GM(1,1) maybe provides us one method to study the system.

According to the grey system theory, the GM(1,1) model is defined as (1).

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = \mu \tag{1}$$

Where a is defined as development coefficient, μ is defined as grey action quantity, we can obtain the parameter estimators in terms of the least square estimate method. And the sequence $X^{(1)}$ can be obtained from applying the AGO transformation on sequence $X^{(0)}$.

Then, the approximate time response function is derived as following,

$$X^{(1)}(t) = \left[X^{(0)}(1) - \frac{\mu}{a} \right] e^{-a(t-1)} + \frac{\mu}{a} \tag{2}$$

The predicted value of $X^{(0)}(t)$ can be restored as

$$X^{(0)}(t) = X^{(1)}(t) - X^{(1)}(t-1) \tag{3}$$

In order to improve the prediction accuracy of GM(1,1) model, it is necessary to fit a residual correction function defined as (4), the parameter a_0,b_0,c_0 can be calculated by "cftool" toolbox in the MATLAB.

$$f(t) = a_0 * \sin(b_0 * t + c_0) \tag{4}$$

From (3) and (4), we can get the final prediction equation,

$$X^{(0)}(t) = X^{(1)}(t) - X^{(1)}(t-1) + f(t) \tag{5}$$

Here the t is the time series.

2) BPANN Model

BPANN is the most widely used network and is considered the workhorse of ANNs. Because of its simplicity and its power to extract useful information from samples, the application of BPANN model is very widely used recently. BPANN model eliminates the limitations of the traditional methods, and establishes the mapping accurately between the input and output variables.

The BP algorithm can be described as Figure 1. In the first phase, the actual outputs of the network are computed forward from the input layer to the output layer. While in the second phase, the descent gradient is calculated in a back-propagation fashion, which makes it possible to adjust the weights in a descent direction. This procedure is repeatedly performed for each training pattern until all error signals between fitted values and actual values are sufficiently small

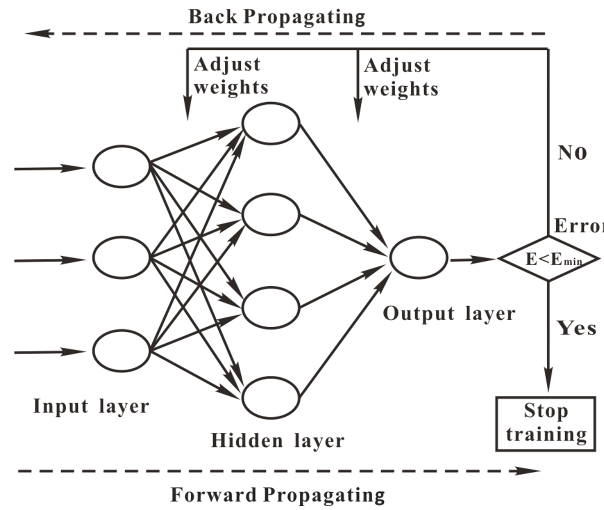

Figure 1. The structure of BPANN model

According to the given data, BPANN model is designed as shown in Figure 1, network parameters are designed as follows: in the input layer, there are 3 nodes, which represent the household income of month t-3,t-2,t-1 respectively; one hidden layer is utilized, and there are 4 nodes are given, the decision is made according to combination of previous researches and actual experiment. The output layer has 1 node, represent the data of month t. The transformation function between input layer and hidden layer is "tansig", between hidden layer and output layer is "purelin". The training function is "traincgf". The minimum training rate is set as 0.01, the number of iterations is 1000, and allowing minimum mean square error (E_m) is 0.00001.

III. ANALYSIS AND DISCUSSION

A. Model Inspection

Based on the household income from December 2008 to December 2009, GM(1,1) model and BPANN model are established and then inspected respectively.

1) GM(1,1) Model Inspection

The prediction accuracy can be examined by relative error, error probability and relative coefficient. In GM(1,1) model, the average relative error is 2.46% , which is lower than 5%. It points out the fitting accuracy is high; parameter P is 0.923077 and parameter C is 0.354759, the accuracy degree of GM(1,1) model can be defined as qualified from the table I. The relative coefficient is 0.6891, which is higher than 0.60 that indicating the predicted value has a good correlation with the actual value. In a word, the accuracy of GM(1,1) model is qualified. It can be used to predict the household income in Shanghai low-rent housing families.

TABLE I. THE ACCURACY DEGREE OF GM (1,1) MODEL

Accuracy degree	*P*	*C*
Good	>0.95	<0.35
Qualified	>0.80	<0.5
Just qualified	>0.70	<0.65
Disqualification	≦0.70	≧0.65

2) BPANN Model Inspection

Taking the income data (from December 2008 to December 2009) as sample pattern, the network is trained in the light of learning rule of error back propagation network after 237 times repetitions. The average relative error of the predicted values is 2.46%. Using this model to predict the test samples (from January 2010 to February 2010), the average relative error is 3.25%. Two values are both lower than 5%. The network is excellent. In a word, this BPANN model can predict the future value.

B. Model Comparison

In order to compare the model accuracy and select the better one between GM(1,1) model and BPANN model, we select some indicators as follows.

Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | x_i - \hat{x}_i | \tag{6}$$

Mean Proportional Error (MPE):

$$MPE = \frac{1}{n}\sum_{i=1}^{n} | \frac{x_i - \hat{x}_i}{x_i} | \tag{7}$$

Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_i - \hat{x}_i)^2} \tag{8}$$

x_i *(i=1，2…n)* is the actual value; $\hat{x}_i$ *(i=1，2…n)* is the fitting value; *n* is the number of sample dates.

According to the indicators, the comparative results are shown in table II. Because the predictive methods are different, so the prediction time interval is different.

Table II tells us both MAE and MPE of GM(1,1) model are higher than BPANN model, which indicating the fitting error of GM(1,1) model is higher than that of BPANN model. The RMSE and maximum relative error represent the dispersion of the fitted values, the lower the value is, the lower the dispersion is. From the table II we can see the RMSE and maximum relative error are 8.68 and 2.62%, which are lower than those of GM(1,1) model respectively. So the BPANN is closer to the true value.

It is found that the fitting accuracy of BPANN model is higher than GM(1,1) model. However the fitting results only represent the fitting accuracy but not the prediction accuracy. In order to reveal the prediction accuracy of GM(1,1) model and BPANN model, we compare the prediction value and the actual value in January 2010 and February 2010. The results are shown in table III.

TABLE II. COMPARISON OF GM(1,1) AND BPANN MODEL ON FITTING ACCURACY

Month	*Actual value*	*GM(1,1) model*		*BPANN model*	
		Predicted value	*Relative error*	*Predicted value*	*Relative error*
2009.1	743.61	728.89	1.98%		
2009.2	729.45	738.86	1.29%		
2009.3	688.91	725.14	5.26%	689.64	0.11%
2009.4	694.28	697.57	0.47%	681.31	1.87%
2009.5	672.77	673.00	0.03%	676.12	0.50%
2009.6	655.84	666.80	1.67%	666.74	1.66%
2009.7	667.70	685.11	2.61%	650.22	2.62%
2009.8	682.23	721.79	5.80%	686.84	0.67%
2009.9	765.82	761.47	0.57%	766.29	0.06%
2009.10	790.37	787.32	0.39%	780.23	1.28%
2009.11	783.50	789.46	0.76%	773.20	1.31%
2009.12	751.30	769.99	2.49%	759.59	1.10%
MAE		13.65		7.92	
MPE		1.94%		1.12%	
RMSE		18.32		8.68	

TABLE III. COMPARISON OF GM(1,1) AND BPANN MODEL ON PREDICTION ACCURACY

Month	*Actual value*	*GM(1,1) model*		*BPANN model*	
		Predicted value	*Relative error*	*Predicted value*	*Relative error*
2010.1	724.31	741.89	2.43%	724.79	0.07%
2010.2	664.53	722.51	8.72%	668.41	0.58%

As shown in table III, the relative error of GM(1,1) model in February 2010 is 8.72% ,which is larger than 5%. There are two reasons for that, on the one hand, GM(1,1) model can reflect the overall tendency and suitable for the single exponential growth model, but it is affected easily by some uncertain factors which will cause data increasing or reducing suddenly. On the other hand, the GM(1,1) model takes the predicted values as input variables, which will create large cumulative error.

The relative error of BPANN model in January 2010 and February 2010 are 0.07% and 0.58%, they are much smaller than GM(1,1) model, which indicating the BPANN model has a higher precision accuracy on the non-sample data. So this model can be used to predict the income in Shanghai low-rent housing families.

In addition, from the perspective of reasonable comparison, GM(1,1) model translates random information into strong regular information to establish continuous different equations. GM(1,1) model could reflect the overall tendency, but it is always affected by some uncertain factors. For example, in February 2010, because there are a biggest data changing and the changing direction is contrary to prediction tendency of GM model when the household income changed from 724.31 Yuan to 664.53 Yuan, the relative error of prediction value is very large. And since the low-rent housing policy was implemented in 1999, the access standards have became more and more lenient from then on, as a result, the household income is changing. It always has been an uncertain factor for the predictive efficiency of GM(1,1) model. However, BPANN model is able to eliminate the limitations of the non-linear models and approach any non-linear mapping. Although there are many uncertainties and few samples in household income of low-rent housing families, BPANN model could control the error in a certainly reasonable extent and get a better prediction.

From the analysis above, we can draw the conclusions that (1) BPANN model is more accurate than GM(1,1) model, (2) BPANN model has a better adaptability under the influence of uncertain factors. So BPANN model can be used in the prediction of the household income of low-rent housing families and it will have a good application prospect.

C. Results

Using BPANN model to forecast the household income of Shanghai low-rent housing families from March 2010 to August 2010, the results are shown in table IV.

TABLE IV. HOUSEHOLD INCOME PREDICTION OF SHANGHAI LOW-RENT HOUSING FAMILIES

Month	***2010.3***	***2010.4***	***2010.5***	***2010.6***	***2010.7***	***2010.8***
Prediction	650.30	651.77	749.38	778.07	792.45	772.41

Table IV tells us the household income increases month by month from March 2010 to July 2010 and then decreases slowly in August 2010. Taking the data from December 2008 to February 2010 into consideration, we find wave-like rise in the income data. It is forecasted that the average income will continue to show the undulation condition and the margin of fluctuation is slightly higher than before in the coming months.

IV. CONCLUSIONS

Based on data conditions and model characteristics, this paper select the GM(1,1) model and BPANN model for comparison. These models have different features and different effects. GM(1,1) model has fewer requirements on the data and the mount of calculations is small. But this method ignore some uncertain factors, only reflects the general trend. It has a poor data fitting capability. BPANN model using the inner function strictly approximates the data plot, it could fit the value affected by uncertain factors. However, BP neural network model does not reflect the long-term trends and prone to the phenomenon of over-training.

In consideration of GM(1,1) model is affected easily by some uncertain factors and the relative error of prediction value is larger when the actual value changing exceptionally, GM(1,1) model could not meet the need of prediction. In contrast, BPANN model can be applied successfully in predicting income when the information changing exceptionally and it get a good forecast effect, which provides a scientific basis for the policy making of the low-rent housing families. The prediction results of BPANN model showed that the household income of Shanghai low-rent housing families would maintain relatively fast growth in the short term but increase slowly in the long term.

REFERENCES

[1] D. Y. Qiu, M. L. Wang, Low-rent House-Real Choice for Help to the Urban Vulnerable Groups in House Needs(in Chinese), Joural Of Chongqing Jianzhu University, 2003 25(6), pp. 124-129.

[2] X. G. Zhao,W. Q. Yuan, The Institutional Puzzle Dom and Its Legal Rectification of Cheap Renting House in China(in Chinese), Journal Of Chongqing Jianzhu University, 2007, 29(6), pp. 145-148.

[3] F. L. Wu, Land Development, Inequality and Urban Villages in China(in Chinese), International Journal Of Urban And Regional Research, 2009, 33(4), pp. 885-889.

[4] J. H. Xu, Mathematical Methods in Contemporary Geography(in Chinese), Higher Education Press, Beijing, 2002.

[5] S. Andreas, A. Neil, Time series prediction, Addison-Wesley Pub, 1994.

[6] J. W. Osborne, Prediction in Multiple Regression, Practical Assessment, Research & Evaluation, 2000.

[7] T. Liu, Application of Markov Chains to Analyze and Predict the Time Series, Modern Applied Science, 2010(4),pp. 162-166.

[8] W. K. Ching, E. S. Fung, M. K. Ng, A multivariate Markov chain model for categorical data sequences and its applications in demand predictions, IMA Journal of Management Mathematics, 2002(13), pp. 187–199.

[9] J. L. Deng, Introduction to grey system(in Chinese), Journal of Grey System, 1989, 1(1), pp. 1-24.

[10] J. L. Deng, Grey Prediction and Decision(in Chinese), Huazhong University of Technology, 1986.

[11] I. A. Basheer, M. Hajmeer, Artificial neural networks: Fundamentals, computing, design, and application, Journal of Microbiological Methods, 2000, 43(1), pp. 3-31.

[12] S. Hudon, Y. Yan, W. Kinsner, A comparative study of neural network models, Mathematical and Computer Modeling, 1990,(14),pp. 300–304.

[13] Z. Q. Sun, Neural network theory and MATLABR 2007(in Chinese), Electronics Industry Publishing House, 2007.