A Novel Divide-and-Conquer Model for CPI Prediction Using ARIMA, Gray Model and BPNN

Wei Xu
weixu@ruc.edu.cn
http://lisa.ruc.edu.cn
School of Information, Renmin University of China

Authors: Y. Du, Y. Cai, M. Chen, W. Xu, H. Yuan, and T. Li
Supported by Beijing Social Science Fund (No. 13JGB035)
2014.06.03

中国人民大學
RENMIN UNIVERSITY OF CHINA

中国人民大学
RENMIN UNIVERSITY OF CHINA

Content

1 **Research Background**

2 **Theoretical Foundation**

3 **The Proposed Method**

4 **Experimental Results**

5 **Conclusions & Future Work**

Research Background

✓ **The Consumer Price Index (CPI) is a widely used measurement of cost of living**

✓ **Predicting the change of CPI is significant to both the public and policy makers**

中国人民大学
RENMIN UNIVERSITY OF CHINA

Research Background

✓ **CPI consists of eight sub-indexes:**

1. Food
2. Articles for Smoking and Drinking
3. Clothing
4. Household Facilities, Articles and Maintenance Services
5. Health Care and Personal Articles
6. Transportation and Communication
7. Recreation, Education and Cultural Articles and Services
8. Residence

Index	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8
Weight	31.79%	3.49%	8.52%	5.64%	9.64%	9.95%	13.75%	17.22%

✓ **The eight sub-CPIs are integrated with a specific formula as follow to form the value of CPI.**

$$TI = \sum_{i=1}^{8} W_i I_i$$

Theoretical Foundation

GM(1,1)

The grey system theory is first presented by Deng in 1980s. In the grey forecasting model, the time series can be predicted accurately even with a small sample by directly estimating the interrelation of data. In this theory, a system in which information is completely known is defined as a white system while the system in which information is partly known is named a black box.

The GM (1, 1) model is one type of the grey forecasting which is widely adopted. It is a differential equation model of which the order is 1 and the number of variable is 1, too. The differential equation is:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u$$

$x^{(1)}$ is a sequence generated after accumulating, t is time, a and u are parameters to be estimated.

Theoretical Foundation

GM(1,1)

The process of GM (1, 1) model is represented in the following steps:

Step1: The original sequence is

$$x^{(0)}=\{x^{(0)}(1),x^{(0)}(2),x^{(0)}(3),\cdots,x^{(0)}(n)\},i=1,2,\cdots,n$$

$$x^{(1)}(i)=\sum_{m=1}^{i}x^{(0)}(m),\quad i=1,2,\cdots,n$$

Step2: Estimate the parameters a, u using the ordinary least square(OLS)

$$B=\begin{bmatrix} -\frac{1}{2}[x^{(1)}(1)+x^{(1)}(2)] & 1 \\ -\frac{1}{2}[x^{(1)}(2)+x^{(1)}(3)] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}[x^{(1)}(n-1)+x^{(1)}(n)] & 1 \end{bmatrix},$$

$$y_n=\left[x^{(0)}(2),x^{(0)}(3),\cdots,\ x^{(0)}(n)\right]^T$$

a、u: $$\hat{a}=\begin{bmatrix} a \\ u \end{bmatrix}=(B^TB)^{-1}B^Ty_n$$

Step3: Equation solving

$$\hat{x}^{(1)}(i+1)=(x^{(0)}(1)-\frac{u}{a})e^{-ai}+\frac{u}{a},$$

$$\begin{cases} \hat{x}^{(0)}(1)=\hat{x}^{(1)}(1) \\ \hat{x}^{(0)}(i)=\hat{x}^{(1)}(i)-\hat{x}^{(1)}(i-1),i=2,3,\cdots,n \end{cases}$$

Step4: Test the model using the variance ratio and small error possibility.

Theoretical Foundation

ARIMA

ARIMA model was first put forward by Box and Jenkins in 1970; therefore, it is also called" Box-Jenkins" model. ARIMA stands for "Autoregressive Integrated Moving Average." The model has been well acknowledged as a foremost method to forecast the future trend of a time series by taking full advantage of time series data in the past and present. ARIMA model is usually described as ARIMA (*p, d, q*), p refers to the order of the autoregressive variable, while *d* and *q* refer to integrated, and moving average parts of the model respectively. When one of the three parameters is zero, the model is changed to model "AR", "MR" or "ARMR". When none of the three parameters is zero, the model is given by:

$$(1-\sum_{i=1}^{p}\phi_i L^i)(1-L)^d X_t = \delta + (1+\sum_{i=1}^{q}\theta_i L^i)\varepsilon_t$$

Where L is the lag number, ε_t is the error term.

Theoretical Foundation

ARIMA

The steps of ARIMA model are:

Step1: Test the Stationarity of the time series

Step2: Build the Model by the Regulations

Step3：White Noise Test

Step4：Analyzing and Forecasting

The constructed model is:

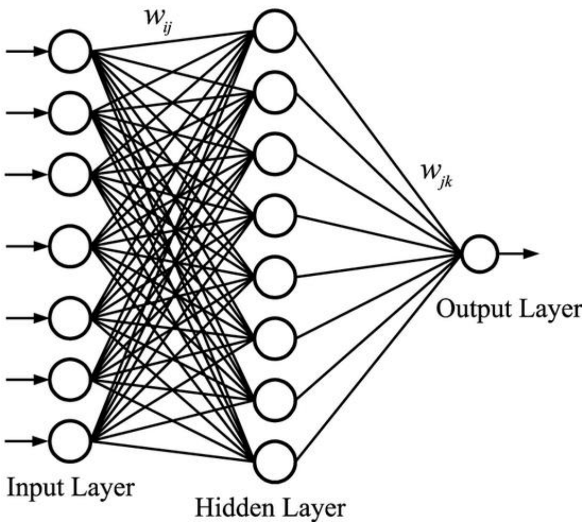
$$(1-\sum_{i=1}^{p}\phi_i L^i)(1-L)^d X_t = \delta + (1+\sum_{i=1}^{q}\theta_i L^i)\varepsilon_t$$

Where L is the lag number, ε_t is the error term.

Theoretical Foundation

BPNN

Artificial Neural Network (ANN) is a mathematical and computational model which imitates the operation of neural networks of human brain. ANN consists of several layers of neurons. Neurons of contiguous layers are connected with each other. The values of connections between neurons are called ‘weight’. Back Propagation Neural Network (BPNN) is one of the most widely employed neural network among various types of ANN. BPNN was put forward by David Rumelhart and J.McClelland in 1985. It is a common supervised learning network well suited for prediction. BPNN consists of three parts, one input layer, several hidden layers and one output layer, as is demonstrated in Figure 1. The learning process of BPNN is modifying the weights of connections between neurons based on the deviation between the actual output and the target output until the overall error is in the acceptable range.



Back-propagation Neural Network

The Proposed Method

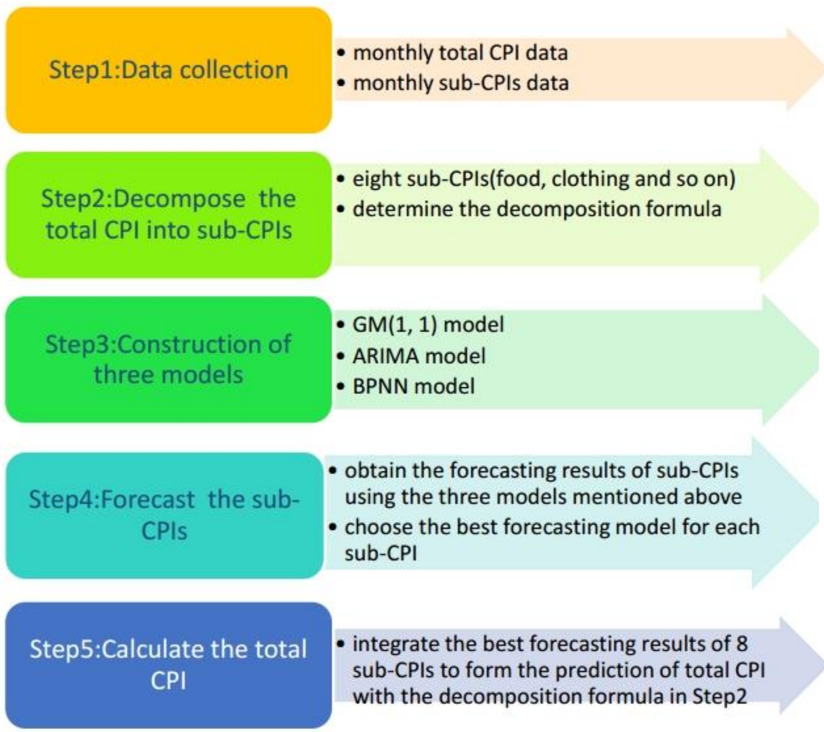


Figure 1.The Framework of the dividing-conquer model

The decomposition formula in step 2 and step 5 is:

$$Y = \sum_{i=1}^{8} c_i x_i$$

Where Y refers to the predicted rate of the total CPI, c_i is the weight of the sub-CPI which will be showed in table 1 andx_i is the predicted value of the sub-CPI which has the minimum error among the three models mentioned above.

Experimental Results

We construct the tree introduced models and get the experimental results as follows (take the Residence sub-CPI as example):

Table 2.

Forecasting results of Residence CPI by GM(1,1)

Date	*Actual value*	*Predicted value*	*Residual error*	*Relative error*
2013-4	102.9	103.0619	-0.1619	0.0016
2013-5	103.0	103.1621	-0.1621	0.0016
2013-6	103.1	103.2624	-0.1624	0.0016
2013-7	102.8	103.3628	-0.5628	0.0055
2013-8	102.6	103.4632	-0.8632	0.0084
2013-9	102.6	103.5638	-0.9638	0.0094

Table 3.

Forecasting results of Residence CPI by ARIMA

Date	*Actual value*	*Predicted value*	*Residual error*	*Relative error*
2013-4	102.9	102.8817	0.0183	0.0002
2013-5	103.0	102.7096	0.2904	0.0028
2013-6	103.1	102.8695	0.2305	0.0022
2013-7	102.8	103.0393	0.2393	0.0023
2013-8	102.6	102.7573	0.1573	0.0015
2013-9	102.6	102.5265	0.0735	0.0007

Table 4.

Forecasting results of Residence CPI by BPNN

Date	*Actual value*	*Predicted value*	*Residual error*	*Relative error*
2013-4	102.9	102.869784	0.030216	0.000294
2013-5	103.0	102.870656	0.129344	0.001256
2013-6	103.1	102.866411	0.233589	0.002266
2013-7	102.8	102.865677	-0.065677	0.000639
2013-8	102.6	102.864506	-0.264506	0.002578
2013-9	102.6	102.872821	-0.272821	0.002659

Table 5.

Error of Sub-CPIs of the 3 Models

	No.1	*No.2*	*No.3*	*No.4*	*No.5*	*No.6*	*No.7*	*No.8*
GM(1,1)	0.0378	0.0009	0.0075	0.0031	0.0010	0.0047	0.0061	0.0047
ARIMA	0.0059	0.0025	0.0028	0.0016	0.0010	0.0072	0.0026	0.0016
BPNN	0.0110	0.0193	0.0041	0.0036	0.0058	0.0053	0.0029	0.0016

Experimental Results

- ✓ We re-model the forecasting approach for sub-index 1 and sub-index 6
- ✓ For predicting sub-index 1, we use BPNN for prediction based on first-order difference series of sub-index 1
- ✓ We offer MA model for predicting sub-index 6

The improved dividing-integration model

	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8
Model	BPNN	GM(1,1)	ARIMA	ARIMA	GM(1,1)	MA	ARIMA	BPNN
Error	0.0028	0.0009	0.0027	0.0016	0.0010	0.0039	0.0026	0.0016

After calculating, the forecasting error of the total CPI by the improved dividing-integration model is 0.0023

Comparison

The original dividing-integration model

	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8
Model	ARIMA	GM(1,1)	ARIMA	ARIMA	GM(1,1)	GM(1,1)	ARIMA	BPNN
Error	0.0059	0.0009	0.0028	0.0016	0.0010	0.0047	0.0026	0.0016

After calculating, the forecasting error of the total CPI by the original dividing-integration model is 0.0034

The baseline models

Model	SARIMA	BRF NN	Verhulst	ARIMA	GM(1,1)	VAR
Error	0.0038	0.0057	0.0035	0.0038	0.0090	0.0179

Conclusions

- ✓ **In this paper, the prediction of national CPI is transformed into the forecasting of 8 sub-CPIs**
- ✓ **In the prediction of 8 sub-CPIs, we adopt three widely used models: a GM (1, 1) model, a ARIMA model and a BPNN model, we can obtain the best forecasting results for each sub-CPI**
- ✓ **We adjust the forecasting methods of sub-CPIs which predicting results are not satisfying enough and get the improved prediction results**
- ✓ **The improved predicting results of the 8 sub-CPIs are integrated to form the forecasting results of the total CPI**

Future Work

- ✓ **The proposed model only uses the time-series information of the CPI, and some factors can be considered to improve the accuracy and stability of the model, especially web information, such as web article news, and micro-blogging data.**
- ✓ **The forecasting performance of sub-CPIs can be improved by some new methods, such as SVM and DWD to deal with high dimension -small sample datasets**

中国人民大學
RENMIN UNIVERSITY OF CHINA

Thanks for your attention!