ISSN 2616-5805 Vol. 3, Issue 2: 35-42, DOI: 10.25236/AJMS.2022.030206

Forecast of Consumer Price Index-Take Beijing as an example

Tongtong Jia[1], Kongming Ai[2]

[1]School of Mathematics and Statistics, Liaoning University, Shenyang, Liaoning, 110000, China
[2]College of Science, Xi'an University of Technology, Xi'an, Shaanxi, 710000, China

Abstract: *The consumer price index (CPI) reflects the relationship between the price changes of goods and services related to the life of residents and is an important indicator to evaluate the level of inflation. Because of the high randomness and volatility of CPI under the infectious diseases, it is very difficult to predict its trend accurately.In this paper, we combine the monthly CPI data of Beijing from January 2020 to July 2022, and use the ARIMA model, GM (1,1) and BP neural network model as the basis of the combined model to forecast the CPI of Beijing under the infectious diseases using two methods: ultra-short-term forecasting and conventional forecasting. It is obtained that the combined model has better forecasting effect than the single model, and the ultra-short-term forecasting effect is better than the conventional forecasting. Among them, the combination model using ARIMA-GM-BP for ultra-short-term forecasting is the best. Finally, the model and method were applied to forecast the CPI of Beijing in August as 102.079.*

Keywords: *CPI; Combined model; Ultra-short-term forecasting; ARIMA-GM-BP*

1. Introduction

Consumer Price Index (CPI), a relative number reflecting the trend and degree of price changes of consumer goods and services purchased by urban and rural residents in a certain period of time, can be used as an indicator to judge whether inflation occurs, which affects the formulation of government policies in monetary, fiscal and price aspects, and also directly affects the living standard of residents. Therefore it is of great significance to find out how to accurately tap and forecast the change pattern of CPI.

In order to accurately tap and forecast the pattern of CPI, many domestic scholars have conducted in-depth studies. Xueyan Zhao established a Markov model to forecast the future changes of CPI in Shaanxi Province based on the sample of consumer price index from 1998 to 2019[1] . Li-Min Wang used the metabolic GM (1, 1) model to forecast the future CPI trend in China based on the annual CPI data from 2003 to 2008[2] . Xiaofeng Guo forecasted the future CPI trend in China based on the monthly CPI data from January 2001 to October 2011 in China using ARIMA model[3] . Jinhuan Liu forecasted the future CPI in China based on the monthly CPI data from November 2008 to August 2013 using a combined model ARIMA-BP [4].

However, in recent years, influenced by the infectious diseases around, the CPI has been more volatile compared with the previous ones. If the original models and methods of are still selected for forecasting, the forecasting effect will be unsatisfactory. In this paper, considers the characteristics that the CPI under the infectious diseases can be affected at any time, fluctuate greatly, and have poor forecasting effect using a single model. First, the autoregressive integrated sliding average model (ARIMA), GM(1,1), and BP neural network model are used for ultra-short-term forecasting (predicting only one period) and conventional forecasting, and then the two forecasting methods are compared, and the three models are combined on this basis to obtain the combined ARIMA-GM-BP model with the highest forecasting accuracy. Finally, using this model, the CPI of Beijing for August is forecasted.

2. Architecture of the model

2.1 ARIMA model

The ARIMA model is also called the autoregressive integrated sliding average model, abbreviated

ISSN 2616-5805 Vol. 3, Issue 2: 35-42, DOI: 10.25236/AJMS.2022.030206

as ARIMA (p,d,q). The definition is as follows.

$$\begin{cases} \Phi(B)\nabla^d x_t=\Theta(B)\varepsilon_t \\ E(\varepsilon_t)=0,Var(\varepsilon_t)=\sigma_\varepsilon^2,E(\varepsilon_t\varepsilon_s)=0,s\neq t \\ Ex_s\varepsilon_t=0,s<t \end{cases} \qquad Brief\ notes\ {}^d x_t=\frac{\Theta(B)}{\Phi(\varepsilon_t)} \tag{1}$$

Where, $\nabla^d=(1-B)^d;\Phi(B)=1-\phi_1 B-\phi_2 B^2-\cdots-\phi_P B^P$ is the autoregressive coefficient polynomial of the smooth reversible ARMA(p,q) model; $\Theta(B)=1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q$ is the moving smoothing coefficient polynomial of the smooth reversible ARMA(p,q) model; $\{\varepsilon_t\}$ is the white noise series.

The model is composed of a combination of difference operation and ARMA model, which can forecast future values by using past observations and present observations of time series. It can tap the dynamic and continuous characteristics of the time series, reveal the relationship between the past and present, present and future of the time series, and is adapted to short-term forecasting.

2.2 GM(1,1) model

The GM (1, 1) model, based on the gray system theory proposed by Jurong Deng, is one of the most widely used dynamic forecasting models, which consists of a univariate first-order differential equation. It has significant advantages in predicting complex and uncertain problems.

We know that the original sequence $x^{(0)}=(x^{(10)}(1),x^{(0)}(2),\cdots,x^{(0)}(n))$ is a non-negative sequence, and the new sequence is obtained by accumulating it once $x^{(1)}$

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

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

Let $z^{(1)}$ be the immediately adjacent mean generating series of the series $x^{(1)}$, i.e. $z^{(1)}=(z^{(1)}(2),z^{(1)}(3),\cdots,z^{(n)}(n))$

Among them $z^{(1)}(m)=\delta x^{(1)}(m)+(1-\delta)x^{(1)}(m-1),m=2,3,\cdots,n\text{且}\delta=0.5$

The equation $x^{(0)}(k)+\alpha z^{(1)}(k)=\beta,\ \ k=2,3,\cdots,4$ is the basic form of GM(1,1)

Where, β is the amount of gray effect and $-\alpha$ is the development factor.

The least squares method yields $\hat{\alpha}=(\alpha,\beta)^T=(B^T B)^{-1}\times B^T\times Y_n$.

$$B=\begin{bmatrix} -\frac{1}{2}(\boldsymbol{x}^1(1)+\boldsymbol{x}^1(2)), & 1 \\ -\frac{1}{2}(\boldsymbol{x}^1(2)+\boldsymbol{x}^1(3)), & 1 \\ \vdots & \vdots \\ -\frac{1}{2}(\boldsymbol{x}^1(\boldsymbol{n}-1)+\boldsymbol{x}^1(\boldsymbol{n}))\ , & 1 \end{bmatrix} Y^n=\begin{bmatrix} x^0(2) \\ x^0(3) \\ \vdots \\ x^0(n) \end{bmatrix} \tag{3}$$

2.3 BP neural network model

BP neural network is a multilayer feedforward neural network that can be trained according to the

ISSN 2616-5805 Vol. 3, Issue 2: 35-42, DOI: 10.25236/AJMS.2022.030206

error back propagation algorithm. With good adaptability and self-learning capability, it is particularly suitable for solving problems with clear input and output correspondence and complex internal mechanisms. bp neural network generally includes input layer, hidden layer and output layer, as shown in Figure 1.

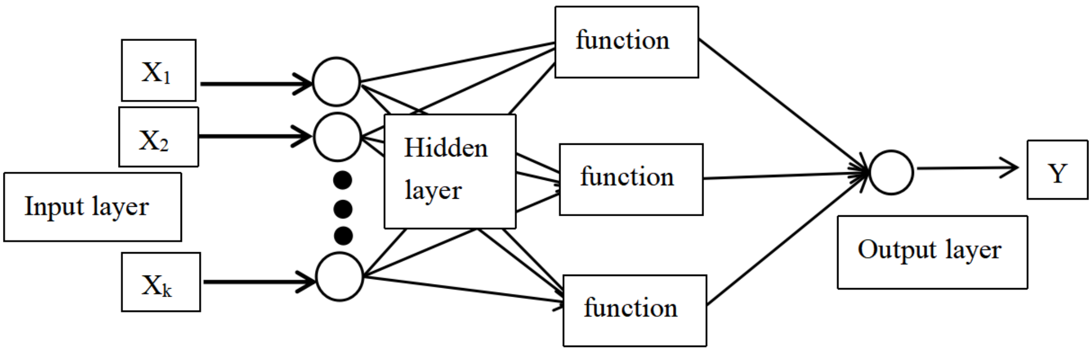

Figure 1: BP neural network structure diagram

2.4 Combined prediction model

Due to the variability and complexity of real data, there is a great deal of randomness and limitation in predicting data using a single model. In order to enhance the flexibility and adaptability of the model, optimize the performance of a single model, and improve the prediction accuracy, Bates proposed the theory of combined prediction models[5] . Today the theory has been widely used in the prediction of many problems. The most long-standing one is the linear combination of the results of a single model prediction given different weights to obtain the results of the combined model prediction. Where the choice of weights is often given by the sum of squared errors.

$$y=\omega_1 y_1+\omega_2 y_2\ \omega_1=\frac{s_1 s_2}{s_1(s_1+s_2)}\ \omega_2=\frac{s_1 s_2}{s_2(s_1+s_2)} \tag{4}$$

Where y is the predicted value of the combined model, y_1, y_2 is the predicted value of the two single models, ω_1, ω_2 is the weight coefficient of the two single models, and s_1, s_2 is the error sum of squares of the two single models.

2.5 Prediction model error indicators

To measure the forecasting effectiveness of a single forecasting model and each combined model, uniform metrics need to be selected. In this paper, two metrics, root mean square error (RMSE) and mean absolute percentage error (MAPE), are selected [6] .

The root mean square error gives a good indication of the accuracy of the forecast and is calculated as follows.

$$RMSE=\sqrt{\frac{\sum_{i=1}^{n}(y_{obs,i}-y_{model,i})^2}{n}} \tag{5}$$

The mean absolute percentage error is used to reflect the prediction error in the form of a percentage. The calculation formula is as follows.

$$MAPE=\frac{1}{n}\sum_{i=1}^{n}|P_{\varepsilon_I}|, P_{\varepsilon_i}=\frac{y_j-y_i}{y_i}\times 100\% \tag{6}$$

The smaller the RMSE and MAPE, the higher the prediction accuracy of the model.

ISSN 2616-5805 Vol. 3, Issue 2: 35-42, DOI: 10.25236/AJMS.2022.030206

3. Data sources

This paper uses monthly CPI data from January 2020 to July 2022 in Beijing for a total of 31 months as the research object. The data are obtained from Beijing Municipal Bureau of Statistics.

4. Analysis of the model

4.1 ARIMA prediction model

Due to the differences in the use of data and the number of forecast periods, two methods of forecasting are used here, one for conventional forecasts and one for ultra-short-term forecasts. The first one divides the monthly CPI data for Beijing from January 2020 to July 2022 into two groups, the first 25 months of data are used to fit the model, and the last 6 months of data are used to test the reliability of the model. In the second, the model is fitted with data from January 2020 to the month before the forecast, and the CPI is forecasted for the next month, and the forecast is repeated six times to obtain the forecast for February to July 2022.

The modeling steps of the ARIMA prediction model are as follows

(1) Smoothing test. Before building the model, the smoothness of the data sample needs to be tested. The time series plot combined with autocorrelation function and ADF test can be used to determine the smoothness of the series. In the test of the above data, the ADF test results show that t is greater than the critical value at 1%, 5% and 10% significance level, and the p-value is greater than 0.5, which means that the CPI series has unit root and therefore the data is not stable. Due to the small amount of data, it is not possible to analyze whether the data are seasonal or not, so the data are directly first-order differenced and then tested, and the results of the series are smooth.

(2) Model pricing. In building the ARIMA model, the model needs to be identified, the lag order selected, and the parameters estimated. Since the sample data are differenced once to eliminate the trend of CPI, d=1. The order of p, q in the ARIMA model can then be determined by the statistical characteristics of the autocorrelation function and the partial correlation function. p, q of the ARIMA model are found to be equal to 0 by the autocorrelation function and partial autocorrelation function of different data. i.e., the ARIMA models using different methods for forecasting have the same parameters, which All of them are ARIMA(0, 1, 0) models.

(3) Parameter estimation and testing of the model. Using EViews software, seven different sets of data were imported and the model ARIMA (0, 1, 0) was found to have t-values around -1.2, p-values around 0.22, and R2 around 0.7. Since the p-values of Q-statistics are all greater than 0.05, it indicates that the residual series of the model are white noise. This model is reasonably constructed and can be predicted.

(4) Predicted results.

According to the ARIMA(0,1,0) model, the conventional forecast is: using the model fitted with data from January 2020 to January 2022 in Beijing, the data of the last 6 months are forecasted. For the ultra-short-term forecast, the model was fitted with data from January 2020 to the month before the forecast, and the forecast values for the last 6 months were obtained respectively. Finally, the real values were compared with these two sets of forecasts, and the results are shown in Table 1.

Table 1: Prediction results of ARIMA model

Time	Consumer Price Index	Ultra-short-term forecasts	Difference1	Conventional predicted values	Differenc e 2
2022.02	101.2	101.2	0	101.2	0
2022.03	101.8	101.1	-0.7	101	-0.8
2022.04	102	101.7	-0.3	100.9	-1.1
2022.05	102.2	101.9	-0.3	100.8	-1.4
2022.06	102.5	102.1	-0.4	100.6	-1.9
2022.07	102.1	102.4	0.3	100.5	-1.6

From the data in the table, can calculate an RMSE of 0.3916 and a MAPE of 0.28% for the ultra-short-term forecast. The standard RMSE of the conventional forecast is 1.2897 and the MAPE is 0.86%. The accuracy of the ultra-short-term forecast is higher than that of the conventional forecast.

ISSN 2616-5805 Vol. 3, Issue 2: 35-42, DOI: 10.25236/AJMS.2022.030206

4.2 GM(1,1) prediction model

Since the GM (1,1) model does not require much data, ten sets of data are used as samples in this paper for each fitting of the model. According to the difference in the number of forecast periods, two methods of conventional forecasting and ultra-short-term forecasting are used for forecasting. In the former, the CPI data of Beijing from April 2021 to January 2022 are selected for fitting the model, and the data from February 2022 to July 2022 are used for testing. The latter selected ten sets of data before the forecast month to fit the model and forecasted CPI for that month only, and repeated six times to obtain the forecast values from February 2022 to July 2022.

The modeling steps of the GM (1,1) prediction model are as follows

(1) Quasi-exponential test Before building the model, a quasi-exponential test needs to be performed on the data after one accumulation, and only the data that pass the quasi-exponential test to build GM(1,1) are reliable. In this paper, we use matla(2021a) pieces to perform quasi-exponential test on the above data, and they all pass the test.

(2) The GM(1,1) model was developed. Using matlab (2021a) software, the above data were modeled separately.

(3) Model evaluation. When using the GM(1,1) model to forecast the future, the GM(1,1) model needs to be tested for its fit to the original data. The residual test was used, and the results showed that the average relative residuals were all small and the fit was good.

(4) Predicted results

According to the GM(1,1) model, the conventional forecast is: using a model fitted with CPI data from April 2021 to January 2022 in Beijing, the CPI data for the second six months are forecasted. The ultra-short-term forecast: using the model fitted with the first ten sets of data in the forecast month, forecast the next month's data, and repeat six times to obtain the forecast values from February 2022 to July 2022. Finally, all the forecast values are compared with the true values, and the results are shown in Table 2.

Table 2: Prediction results of the GM(1,1) model

Time	Consumer Price Index	Ultra-short-term forecasts	Difference 1	Conventional predicted values	Difference 2
2022.02	101.2	101.981	0.781	101.981	0.781
2022.03	101.8	101.738	-0.062	102.073	0.273
2022.04	102	101.672	-0.328	102.165	0.165
2022.05	102.2	101.803	-0.397	102.257	0.057
2022.06	102.5	101.950	-0.550	102.349	-0.151
2022.07	102.1	102.103	0.003	102.442	0.342

From the data in the table, it can be calculated that the RMSE for the ultra-short-term forecast is 0.444 and the MAPE is 0.35%. The RMSE for the conventional forecast is 0.377 and the MAPE is 0.29%. Both the conventional and short-term forecasts are better. However, the conventional forecast is slightly better than the short-term forecast when compared to both.

4.3 BP neural network prediction

4.3.1. Data pre-processing

The Beijing CPI data is one-dimensional data, while the BP neural network has to input high-dimensional data at the input side and its value needs to be between [0,1]. Normalization is performed using $y(t) = x(t)/10^n$. where the number of bits of the highest integer in the n data. Use the first five data as input and the sixth data as output after normalization, i.e. $x = [\, x_1 \;\; x_2 \;\; x_3 \;\; x_4 \;\; x_5 \,]$, $y = x_6$. The CPI data of Beijing from January 2020 to July 2022 are processed to obtain 26 sets of BP neural network sample data.

Here the forecasting is still done using both regular forecasting and ultra-short-term forecasting depending on the number of forecasting periods. The former trains the model using the first 20 sets of data in the sample, and the last six sets are used to test the model. The latter trains the model with all

ISSN 2616-5805 Vol. 3, Issue 2: 35-42, DOI: 10.25236/AJMS.2022.030206

data before the forecast period and the trained model is used to forecast the current CPI only.

4.3.2. BP neural network prediction model building

BP neural network program written using the neural network toolbox in matlab software.

(1) The corresponding data will be imported, the input neuron of this BP neural network is 5, the output neuron is 1, the number of nodes in the middle layer is selected as 10 after testing and comparison, and the network structure is 5-10-1.

(2) The training set is divided into training set, test set, and validation set in the ratio of 8:1:1. The Levenberg-Marquardt learning algorithm is used for training. The final model with very small error of the network system is obtained. The fit is high and the BP neural network is trained successfully.

(3) Using the trained model, the x corresponding to the number of periods to be predicted is brought into the network for simulation to obtain the prediction results. In the conventional forecasting, the first 20 data sets were used as training sets to train the model, and the last six sets were used as test sets to perform the simulation to obtain the normalized forecasts for February 2022 to July 2022. Further inverse normalization is performed to obtain the CPI forecasts for this period. In the ultra-short-term forecast, the model is trained with all the data before the forecast period, and only the current period is used as the test set for simulation to obtain the normalized forecast value for the current period, and then the inverse normalization is performed to obtain the CPI forecast value for the current period. The CPI forecast for February 2022 to July 2022 can be obtained by repeating the training and simulation six times. The results are shown in Table 3.

Table 3: Prediction results of BP neural network model

Time	Consumer Price Index	Ultra-short-term forecasts	Difference1	Conventional predicted values	Difference 2
2022.02	101.2	101.227	0.027	101.227	0.027
2022.03	101.8	101.604	-0.196	101.090	-0.710
2022.04	102	101.377	-0.623	100.708	-1.292
2022.05	102.2	101.922	-0.278	101.224	-0.976
2022.06	102.5	101.926	-0.574	100.828	-1.672
2022.07	102.1	101.764	-0.336	100.641	-1.459

From the data in the table, we can calculate that the RMSE for the ultra-short-term forecast is 0.397 and the MAPE is 0.33%. The RMSE of the conventional forecast is 1.158 and the MAPE is 0.79%. The accuracy of the ultra-short-term forecast is higher than that of the conventional forecast.

4.4 Combined model prediction

Combining ARIMA model, GM(1,1) model, and BP neural network model with each other, four combined models are obtained, namely ARIMA-GM-BP model (abbreviated A-G-B), ARIMA-GM model (abbreviated A-R), ARIMA-BP model (abbreviated A-B), and GM-BP model (abbreviated G-B). In the previous single model forecasts, it can be concluded that the CPI under the infectious diseases is generally better using ultra-short-term forecasts than using conventional forecasts. Therefore, only the combined model's ultrashort-term forecasts are considered here.

Based on the difference between the ultra-short-term forecast value and the true value of each single model from February 2022 to July 2022, the weight coefficients for the combination between the models are calculated and the results are shown in Table 4.

Table 4: Weighting coefficients for each combination model

Weighting factor system Combination Model	A-G-B	A-G	A-B	G-B
ω_1	0.364	0.562	0.508	
ω_2	0.283	0.438		0.445
ω_3	0.343		0.492	0.555

ISSN 2616-5805 Vol. 3, Issue 2: 35-42, DOI: 10.25236/AJMS.2022.030206

The relationship between each single model and the combined model can be reached from the table data. The details are as follows.

A-G-B model. $y = 0.364y_A + 0.283y_G + 0.343y_B$

A-G model. $y = 0.562y_A + 0.438y_G$

A-B model: $y = 0.508y_A + 0.492y_B$

G-B model. $y = 0.445y_G + 0.555y_B$

The results of the ARIMA model, GM(1,1) model, and BP neural network model for February 2022 to July 2022 ultrashort term forecasts are brought into the relational equation of the above combined model, and the results are shown in Table 5.

Table 5: Combined model prediction results

Time	CPI	A-G-B	Difference1	A-G	Difference 2	A-B	Difference3	G-B	Difference4
2022.02	101.2	101.431	0.231	101.542	0.342	101.213	0.013	101.563	0.363
2022.03	101.8	101.459	0.359	101.380	-0.420	101.348	-0.452	101.664	-0.136
2022.04	102	101.578	-0.122	101.688	-0.312	101.541	-0.459	101.508	-0.492
2022.05	102.2	101.880	-0.020	101.857	-0.343	101.911	-0.289	101.869	-0.331
2022.06	102.5	101.996	-0.104	102.034	-0.466	102.014	-0.486	101.937	-0.563
2022.07	102.1	102.091	-0.309	102.270	0.170	102.087	-0.013	101.915	-0.185

From the data in the table, we can calculate that RMSE of A-G-B combination model is 0.225 and MAPE is 0.19%; A-G combination model RMSE is 0.355 and MAPE is 0.34% ; RMSE of A-B combination model is 0.350 and MAPE is 0.28%; RMSE of G-B combination model is 0.377 and MAPE is 0.34%. The A-G-B combination model has the best prediction effect, and the other three combination models have similar prediction effects.

5. Results and Discussion

Comparing the RMSE and MAPE of the four combined models with the single model of the three, the following points can be made.

(1) The combined A-G-B model is the best model for prediction. RMSE and MAPE are much smaller than other single and combined models.

(2) The combined A-G model, combined A-B model, and combined G-B model showed good stability in forecasting, although they did not demonstrate superiority over each single model. The RMSE and MAPE of these three combined models are very close to each other.

(3) Among the single models, the reason why the effect of ultra-short-term forecasts is generally better than that of conventional forecasts is that an infectious diseases may occur in the region at any time and thus affect the CPI of the region, and ultra-short-term forecasts can make timely adjustments to take into account the infectious diseases situation, while conventional forecasts can only be based on past trends and cannot reflect whether they are affected by the recent infectious diseases.

From the above, it is concluded that the best model for CPI forecasting under the infectious diseases is the combined A-G-B model, and the best method is the ultra-short-term forecast.

Here we will use the example of being Beijing city and use the A-G-B combination model to make an ultra-short-term forecast to get the CPI of Beijing city next month.

The steps are as follows

Using data from January 2020 to July 2022 in Beijing, the ARIMA model, GM(1,1) model, and BP neural network were used to forecast the CPI in August.

The results obtained from the prediction of each single model are substituted into the relational equation of the combined A-G-B model.

Find the result and analyze it.

The ARIMA model predicts the CPI of Beijing for August to be 102, the GM (1,1) model predicts the CPI of Beijing for August to be 102.251, and the BP neural network model predicts the CPI of

ISSN 2616-5805 Vol. 3, Issue 2: 35-42, DOI: 10.25236/AJMS.2022.030206

Beijing for August to be 102.051. The above predicted values are brought into $y = 0.364y_A + 0.283y_G + 0.343y_B$, and the final CPI of Beijing for August is 102.079, down 0.02% from the previous month. The CPI for Beijing in August was 102.079, down 0.02% from the previous month.

References

[1] Zhao Xueyan. Forecast of consumer price index in Shaanxi Province [J]. Modern Business, 2021(29):96-98

[2] Wang, L. M. Consumer price index forecasting based on metabolic GM (1,1) model[J].Price Monthly, 2010(6):23-25,29.

[3] Guo Xiaofeng. Analysis of Chinese CPI trend forecasting based on ARIMA model [J]. Statistics and Decision Making, 2012(11):29-32.

[4] Liu Jinhuan. Methodological discussion of consumer price index forecasting [J]. Statistics and Decision Making, 2014(3):82-84.

[5] Bates J, Granger C. The Combination of Forecasts [J].Operational Research Quar terly, 1969, 20(4):451-468

[6] Zhu Wenyan. Application of combined forecasting model in CPI forecasting [D]. Guangdong: Jinan University, 2017.