Forecast and Analysis of China's CPI Based on SARIMA Model

Xiao Zhang(✉)

School of Economics, Wuhan University of Technology, Wuhan, China
944397835@qq.com

Abstract. The Consumer Price Index (CPI) is an important indicator to measure the level of inflation in our country. It reflects the impact of commodity price changes on the daily lives of residents. It is an important basis for governments at all levels to carry out fiscal policies and the central bank to formulate monetary policies. CPI data reflecting economic phenomena have obvious seasonal time series characteristics. By extracting a total of 59 months from January 2017 to November 2021 in our country, a seasonal differential autoregressive moving average model (SARIMA) is established for empirical analysis and prediction. The results show that SARIMA (0, 1, 0) $(0, 1, 1)_{12}$ has a high degree of fitting and can better reflect the future trend of my country's CPI. Based on this, using this model to predict the trend of my country's CPI in 2022, it is found that my country's CPI will remain stable at about 102% in 2022, which provides a certain reference for the decision-making of the government, enterprises and other market entities.

Keywords: CPI · Time Series Analysis · SARIMA Model · Forecast

1 Introduction

The Consumer Price Index (CPI), as one of the most important social and economic indicators in our country, plays a significant role in the formulation, implementation and adjustment of macroeconomic policies. It is not only closely related to the formulation of government fiscal and monetary policies, but also directly affects the living standards of residents. Especially in the current situation where international relations are complicated and the epidemic situation has not yet been completely stabilized, through research and analysis of the development trend of our country's CPI in the past few years, and the law of change, we can predict the development trend of our country's future CPI in advance, which is of great theoretical and practical significance to ensure the smooth operation of market economy and the steady improvement of residents' quality of life.

For the CPI forecast analysis, scholars have done more research. However, the academic circle has not yet reached a unified conclusion about the optimal model of CPI prediction. Sun Xiaodan [6] used the ARIMA (13, 0, 0) model to predict my country's CPI by selecting CPI data for the past 20 years, and found that my country will have inflation, but it is generally within a controllable range. Hou Tiantian et al. [3] used

D. Qiu et al. (Eds.): ICBBEM 2022, AHIS 5, pp. 1354–1361, 2023.
https://doi.org/10.2991/978-94-6463-030-5_135

principal component analysis to predict the future trend of my country's CPI based on the establishment of an ARIMA model, and analyzed three factors that have a greater impact on CPI, and obtained the difference between the fitted predicted value and the actual value. There is an error of 0.33% between them. Chen Menggen [2] used ordinary least squares regression, ridge regression, LASSO regression, neural network and other methods to compare and analyze the accuracy of its fitting analysis, and found that the prediction results of neural network are more accurate than traditional regression methods. Li Zhichao and Liu Sheng [4] also compared the CPI predictions of the ARIMA model, gray model, and regression model. They found that the gray models GM (1, 1) and ARIMA (3, 1, 7) had higher prediction accuracy.

In summary, although many scholars have used the ARIMA model to analyze and predict and have obtained relatively good results, the author believes that for CPI, an economic cycle indicator with obvious seasonal characteristics, the seasonal differential autoregressive moving average model (SARIMA) is used. It can more accurately fit the CPI trend characteristics. Therefore, this paper extracts 59 months from January 2017 to November 2021 in our country, and establishes a SARIMA model for empirical analysis and forecasting, and then provides a certain reference for the decision-making of various market entities such as our government and enterprises.

2 Theoretical Predictive Model

If the seasonal period of a non-stationary time series $\{X_t\}$ is s, it is used to $\nabla_s X_t$ denote the first-order difference of the season [8], namely:

$$\nabla_s X_t = X_t - X_{t-s} \tag{1}$$

The second-order difference of the season is:

$$\nabla_s^2 X_t = \nabla_s X_t - \nabla_s X_{t-s} = X_t - 2X_{t-s} + X_{t-2s} \tag{2}$$

Similarly, the seasonal D-order difference can be defined $\nabla_S^D X_t$. If the time series $\{X_t\}$ is a seasonal ARMA process after seasonal D-order difference, $\{\nabla_S^N X_t\}$ is called a seasonal ARMA $(P, Q)^s$ process, then $\{X_t\}$ is called a seasonal ARMA $(P, D, Q)^s$ process, and its model is called a seasonal ARIMA $(P, D, Q)^s$ model, namely:

$$\phi(B)\nabla_s^D X_t = \theta(B)\Theta(B)\varepsilon_t \tag{3}$$

Similarly, define a non-stationary multiplicative seasonal ARIMA model. If the time series $\{X_t\}$ is the seasonal D-order difference and the ordinary d-order difference, followed $\{\nabla^d \nabla_S^D X_t\}$ by the multiplicative seasonal ARMA (p, q)(P, Q)$^{\mathrm{s}}$ process, then $\{X_t\}$ is called the multiplicative seasonal ARIMA (p,d,q) (P,D,Q)$^{\mathrm{s}}$ model, namely [5]:

$$\phi(\mathrm{B})\varnothing(\mathrm{B})\nabla^{\mathrm{d}}\nabla_S^D X_t = \theta(B)\Theta(B)\varepsilon_t \tag{4}$$

3 Empirical Analysis

3.1 Data Sources

This article selects our country's monthly CPI index for 589 months from January 2017 to November 2021. The data comes from the National Bureau of Statistics.

3.2 Model Recognition

For stationarity testing, there are two methods that are most widely used: one is graph testing, and the other is to construct statistics for hypothesis testing. Although the former is easy to operate and has a wide range of applications, it often has a strong subjective color, so it often needs to be supplemented by statistical testing methods for further discrimination. In recent years, with the rapid development of econometric software, the unit root test method has been widely used in stationarity statistical test.

According to the characteristics of stationarity, the variance and mean of a stationary time series are constant, and a stationary time series should fluctuate around a central value. It can be seen from the time sequence diagram that although there is no obvious trend, the volatility is large and cannot be determined as a stationary sequence, so further identification is still needed. At the same time, it can be seen from the autocorrelation graph that the autocorrelation coefficient decays to 0 very slowly, and is initially determined as a non-stationary sequence (Fig. 1).

Through the ADF test, it is found that the P values of lags 1, 2, and 3 are 0.4803, 0.4619, and 0.5679 respectively, which are all greater than 0.05, which confirms that the time series is a non-stationary time series. Then the first-order difference operation is performed on the sequence, and the sequence diagram, autocorrelation diagram identification and unit root test are performed on the obtained difference sequence. The results show that the sequence diagram and the autocorrelation diagram conform to the basic characteristics of a stationary time series, and the P values of the first, second, and

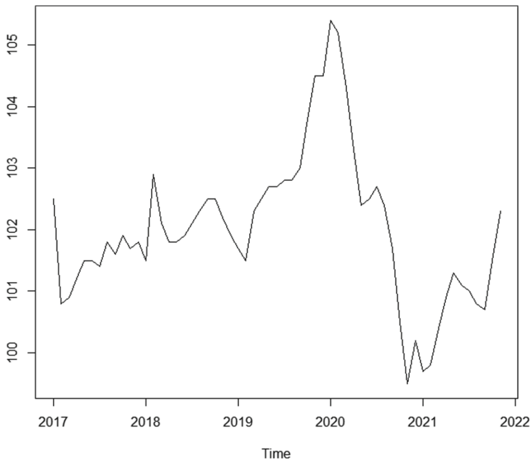

Fig. 1. Timing diagram.

Table 1. ADF inspection results.

Lag order	P value
1	0.4803
2	0.4619
3	0.5679

Table 2. Box.test function test P value after modification.

Delay order	**P value**
6	0.7336
12	0.00211
18	0.009648

third-order lags are 0.01, 0.0195, and 0.0485, respectively, which are all less than 0.05. Therefore, the original hypothesis is rejected and considered There is no unit root, that is, the series is regarded as a stationary series (Table 1).

For this stationary time series, we need to further test whether it is a pure random stationary series, that is, whether there is a connection between the series. If the sequence studied is a purely random sequence and past behavior has no effect on future development, then from a statistical point of view, the analysis of the problem will be meaningless, and the prediction of the future will be of no value. In order to reduce the error of the test results caused by the small data sample size, the modified Box.test function is used to test the results and the results are summarized as shown in Table 2. It can be seen that the delay of the Q statistics under the 12th and 18th orders The values are 0.00211 and 0.009648 respectively, which are both significantly less than 0.05. Therefore, the null hypothesis that the sequence is purely random can be rejected and considered to be of research significance.

3.3 Model Estimation

For CPI data with distinctive seasonal characteristics, it is more accurate to use the SARIMA model to predict [1]. On the basis of considering seasonal factors, a first-order 12 step difference is carried out. From the autocorrelation and partial autocorrelation graphs, we can know that within the 12th order, the partial autocorrelation coefficient and the autocorrelation coefficient have no obvious truncation feature, so try to fit the ARMA(1,1) model. Considering the seasonal sub-correlation characteristics, it can be seen that the autocorrelation coefficients delayed to the 12th order and the 24th order show tailing characteristics, and the partial autocorrelation coefficients have tailing characteristics. Try to use ARMA(0,1)12 for fitting. Therefore, the product-season model SARIMA(1,1,1) $\times$ (0,1,1)12 which is tried to fit is obtained.

At the same time, in order to avoid the inaccurate model recognition caused by the subjective judgment of the individual, the R language auto. The Arima function function automatically determines the order of the model, and uses the R software to automatically try different order combinations to select the optimal model. The optimal model obtained by running is SARIMA(0,1,1)(1,1,1)12. Taking into account that the automatic order function may have over-fitting [7], this article will compare, test, and optimize the above two models, and then select the optimal model to better predict the trend of my country's future CPI [9].

3.4 Model Checking

For the model SARIMA(1,1,1) × (0,1,1)12, the fitting model is: (1–0.2026B) = (1–0.0476B)(1–0.9999). The significance test of the model is performed, and the results are shown in Table 3. When the delays are 6, 12, and 18, the P values of all statistics statistics are significantly greater than 0.05. Therefore, the residual sequence of the fitted model can be considered It belongs to the white noise sequence, that is, the fitting of the model is significant.

Then continue to test the significance of the model parameters. It can be seen from Table 4 that the coefficients of ar and ma are not significant, so the coefficient is removed and the SARIMA(1,1,1)(0,1,1)$_{12}$ model is re-simulated.

Together, fitting SARIMA(0,1,0)(0,1,1)$_{12}$, the newly fitted model is: $\nabla^2 x_t = (1-0.9999B^{12})\varepsilon_t$, where $\mathrm{Var}(\varepsilon_t) = 0.3823$. Continue to test the significance of the model. From the results in Table 5, when the delay is 6, 12, and 18, the P values of all statistics statistics are significantly greater than 0.05, so the residual sequence of the fitted model can be considered to be white noise The sequence, that is, the fit of the model is significant. Then continue to test the parameter significance of the model. The test result is P = 7.920813e−6, which is much less than 0.05. Therefore, the coefficient is obviously non-zero. It can be considered that the model fits the time series well.

Table 3. ARIMA(1,1,1)(0,1,1)$_{12}$ model significance test results.

Delay order	P value
6	0.9693
12	0.648
18	0.8758

Table 4. ARIMA(1,1,1)(0,1,1)12 parameter significance test result.

parameter	estimated value	P value
Φ1	0.2026	0.6698845
θ1	0.0476	0.4585339
sθ12	−0.9999	1.072501e−5

Table 5. ARIMA(0,1,0)(0,1,1)$_{12}$ model significance test results.

Delay order	P value
6	0.7874
12	0.2138
18	0.412

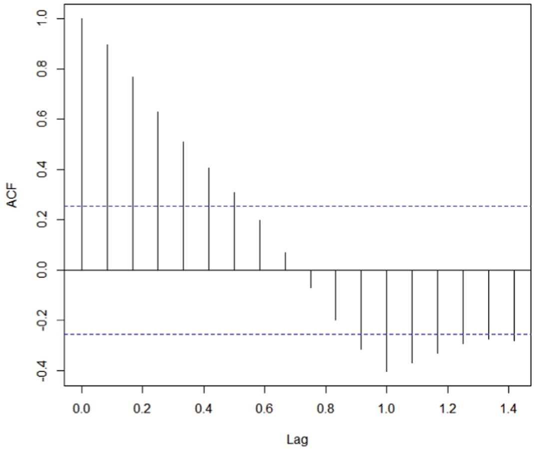

Fig. 2. Autocorrelation diagram.

Repeat the above operations for the model SARIMA(0,1,1)(1,1,1)$_{12}$, and finally, a model and a modified model that can better fit the time series are obtained: SARIMA(0,1,0) (0,1,1)$_{12}$, which is consistent with the results obtained by the first model after modification and optimization. Therefore, this model will be used as the best fitting model for prediction.

3.5 Model Prediction

Through the above analysis, the SARIMA(0,1,0)(0,1,1)$_{12}$ model is used to predict and analyze my country;s CPI in the next 13 months. The specific data of the predictive analysis is shown in Table 6 and the predicted and analyzed CPI trend is shown in Fig. 2. The results show that my country's CPI will fall relatively rapidly from 2020 to 2021 due to the impact of the epidemic, and it will only be about 100% at the beginning of 2021. However, as the situation of the epidemic improves, my country's CPI index has also increased significantly from 2021–2022. It is expected that our country's CPI index will remain at about 102% in 2022, and there will be no significant inflation or deflation. It matches with our country's current prudent fiscal and monetary policies (Fig. 3).

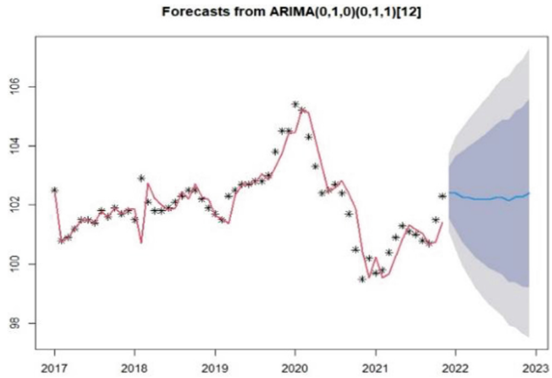

Fig. 3. CPI forecast trend in the next 13 months. (Diagrammed by the author)

Table 6. CPI forecast values in the next 13 months.

	FORECAST	LO 80	HI 80	LO 95	HI 95
DEC 2021	102.4253	101.5393	103.3112	101.07039	103.7802
JAN 2022	102.4007	101.1477	103.6536	100.48453	104.3168
FEB 2022	102.2807	100.7565	103.8049	99.94963	104.6118
MAR 2022	102.2408	100.4867	103.9948	99.55818	104.9234
APR 2022	102.1808	100.2238	104.1379	99.18775	105.1740
MAY 2022	102.1810	100.0400	104.3219	98.90666	105.4552
JUN 2022	102.1811	99.8709	104.4913	98.64789	105.7143
JUL 2022	102.2412	99.7733	104.7091	98.46687	106.0156
AUG 2022	102.2614	99.6453	104.8775	98.26037	106.2624
SEP 2022	102.1416	99.3852	104.8979	97.92607	106.3571
OCT 2022	102.2818	99.3920	105.1716	97.86218	106.7014
NOV 2022	102.2820	99.2646	105.2994	97.66733	106.8967
DEC 2022	102.4073	99.2130	105.6016	97.52210	107.2925

4 Conclusion

This paper selects our country's CPI monthly data from 2017 to 2021, and builds a seasonal product model to select the SARIMA (0, 1, 0) (0, 1, $1)_{12}$ model with higher fitting accuracy for my country's next 13 months The CPI for fitting prediction. The results show that after experiencing the epidemic, China's CPI will slowly rise to around 102% in 2022 and remain at this level. It is recommended that the government continue to adopt a prudent fiscal and monetary policy, and at the same time, on the basis of a prudent policy, adopt a loose monetary policy and fiscal policy to stimulate our country's economy, which will mobilize the vitality and enthusiasm of market entities and ensure steady and upward development of the social economy.

At the same time, it should be pointed out that there is a lag in our investigation and research on CPI data, and there are some statistical errors. With the development of big data and artificial intelligence, it will be a very effective and accurate method to apply big data to CPI prediction. However, the prediction of China's CPI by big data also needs further research.

References

1. Ao Xiqin, Gong Yujie, Wang Jinting, Zheng Yang. CPI prediction based on SARIMA model in Anhui Province[J]. Journal of Bengbu University, 2017, 6(03): 83–86.
2. Chen Menggen, Ren Taoping. CPI prediction model for the new normal economy-construction and empirical comparison [J]. Research World, 2020 (02): 3–8.
3. Hou Tiantian, Yang Cong, Zhan Binghuan. Chinese fiscal revenue forecast based on ARIMA and markov chain model [J]. Journal of pingdingshan university,2020,35(02):6–11+42.
4. Li Zhichao, Liu Sheng. Comparison of predictions based on ARIMA model, grey model and regression model[J]. Statistics and Decision, 2019, 35(23): 38–41.
5. Mao Yujia. Research on combined forecasting model based on time series [D]. Jiangxi University of Finance and Economics, 2020.
6. Sun Xiaodan. Analysis and prediction of my country's CPI index based on ARIMA model[J]. Trade Fair Economy, 2021(15): 15–17.
7. Xia Chun. Calculation, seasonal adjustment and related economic implications of actual economic time series [J]. Economic Research, 2002(03): 36–43+94.
8. Xue Yi; Chen Liping. "Time Series Analysis and R Software" Tsinghua University Press, 2019.12
9. Zhang Ting. Comparative forecast analysis of CPI's SARIMA model and X-12 seasonal adjustment model[J]. Economic Issues, 2014(12): 37–41.