Proceedings of CONF-MPCS 2026 Symposium: Theoretic Physics and Plasma Physics
DOI: 10.54254/2753-8818/2026.DL33961

Research on CPI Forecasting Based on SARIMA Time Series Model

Yutong Wu

School of Science, North China University of Technology, Beijing, China
wytong2012@126.com

Abstract. CPI, abbreviation of consumer price index, essentially refers to a comprehensive measure of price levels, with a wide range of influence that links micro-level and macro-level factors step by step. It exerts a significant impact on individuals, enterprise production, government policies, financial market investment, and international economic development. For instance, pensions and minimum wages in many countries or regions are linked to the Consumer Price Index and adjusted accordingly. Moreover, fluctuations in CPI lead enterprises to adjust product pricing based on its trend: prices are more likely to rise during inflation and fall during deflation. In terms of monetary policy, taking the central bank as an example, when inflation occurs (i.e., excessively high CPI), the central bank usually raises interest rates, tightens money supply, or increases the deposit reserve ratio to curb excessive demand, while the opposite measures are taken during deflation. In addition, there is a two-way relationship between CPI and seasonality. Seasonal fluctuation is one of the core driving factors of short-term changes in CPI. Therefore, it is of great significance to study the Consumer Price Index by taking seasonal effects into account.

Keywords: Consumer price index, seasonal fluctuation, time series model, random error

1. Introduction

As a core macroeconomic indicator, the Consumer Price Index (CPI) is crucial for measuring a country's inflation level. It is a relative number reflecting changes over time in the price level of consumer goods and services typically purchased by households. In China, the CPI is released monthly by the National Bureau of Statistics, and its fluctuations directly affect the formulation and adjustment of monetary, fiscal, and social policies. Studies have shown that since April 2019, the trends of CPI, core CPI, and PPI have diverged, with CPI rising against the trend amid increasing downward pressure on the macroeconomy. Changes in China's overall price level, whether in the consumption sector or other fields, exhibit obvious structural increases, and this trend is long-term in nature [1]. The structural rise in the overall price level has led to deviations between residents' price perceptions and the trend of the overall price level, and has created pressures and constraints on the adjustment of macroeconomic policies.

There are many commonly used models for CPI forecasting research. For example, Yuan and Zhang constructed a fuzzy AR(p) time series model to forecast CPI. Through actual data simulation, they found that the fuzzy AR(p) time series model achieved better forecasting performance than the

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.DL33961

traditional ARIMA model, providing a better guarantee for policy formulation [2]. Yao and Zou used monthly CPI data to build a combined SVM-ARIMA model based on wavelet decomposition, realizing accurate prediction of CPI, which provides data support and guidance for the formulation of economic policies [3]. Yan et al. adopted the time series method and used web search monitoring data to forecast CPI, providing a reference for grasping the macroeconomic trend [4]. Ma et al. applied the ARIMA time series model to analyze and model monthly consumer price index data of China, the United States and Germany. The results could accurately predict the changing trend of the CPI, and government intervention during the pandemic exerted an important impact on CPI fluctuations [5]. Wu et al. constructed an SVAR model based on nearly 20 years of CPI and market price indices of three agricultural products (japonica rice, corn and soybean), exploring the internal relationship between agricultural product price indices and CPI. This provides a basis for the reasonable regulation of agricultural product prices and the stable development of the market economy [6].

In addition, Hu investigated the CPI of Guangxi Zhuang Autonomous Region and Guangdong Province over the past 20 years, providing references for Guangxi to play its directional role as a financial opening gateway facing ASEAN and for the future economic development of Guangxi [7]. Zhou and Liu pointed out that major economic powers all take price stability as the primary goal of monetary policy, so the price level is a very important reference indicator for monetary policy makers. This paper empirically analyzes the relationship between CPI, retail price index of commodities and broad money supply [8]. Wang used relevant data and materials of the U.S. CPI to illustrate three price statistical methods for insurance products, analyzed the relationship between insurance consumption and related household consumption, enriched empirical literature on the economic impact of insurance activities, and provided a reference for China to handle insurance products in the statistical work of CPI and other consumption indicators [9]. Wu studied the impact of international oil prices and exchange rates on China's consumer price index of transportation and communication. Using a time-varying parameter vector autoregressive model, she empirically analyzed the dynamic relationship between international oil prices and China's CPI for transportation and communication, which is of great significance for formulating effective macroeconomic policies [10].

This article primarily aims to demonstrate the forecasting of the Consumer Price Index based on its seasonal variation characteristics. The second section mainly elaborates on the theoretical foundations of the Seasonal Autoregressive Integrated Moving Average (SARIMA) model and its corresponding mathematical formula. The third section focuses on the predictive analysis and presentation of results using the national Consumer Price Index and its sub-items from 2021 to 2024. The final section summarizes the findings based on the data analysis results from the third part and discusses some identified issues.

2. Theoretical model

SARIMA is specifically designed to address forecasting problems for time series data with seasonal periodic patterns. By introducing P, D, Q, s components, this method improves the ability to model seasonal components when compared to the traditional ARIMA model. To facilitate joint modeling, SARIMA decomposes a time series into three components: trend, seasonality, and random error. The model's good interpretability gives clear meanings to statistical parameters.

An important advantage of the SARIMA model is reflected in the prediction of the Consumer Price Index. Its core lies in the explicit modeling of seasonality and trend. CPI data usually exhibit obvious seasonal patterns and long-term trends. The seasonal differencing and seasonal parameters

DOI: 10.54254/2753-8818/2026.DL33961

in the SARIMA model can directly capture and separate these regularly recurring fluctuations, avoiding their interference with the analysis of long-term trends and random components. In addition, the model has strong interpretability. It decomposes the CPI series into autoregressive (AR), differencing (I), moving average (MA) components and their corresponding seasonal parts. Each component has a clear economic or statistical interpretation, which facilitates the analysis of the internal driving forces of price fluctuations. Furthermore, its forecasting performance is stable after stationarity processing. Through differencing, including regular differencing and seasonal differencing, SARIMA can effectively handle the trend non-stationarity and seasonal non-stationarity commonly found in CPI data, transforming the series into a stationary process for modeling, which is crucial for unbiased forecasting. Just like, a certain study used the Seasonal Autoregressive Integrated Moving Average model to develop a forecasting method for typhoon frequency, providing strong support for policy formulation [11]. Divisekara et al. modeled and forecast the seasonality and volatility of red lentils by calculating seasonal indices and applying the SARIMA model to time series [12]. These studies well illustrate the advantages of this model in forecasting research.

The parameter form of the SARIMA model is $(p,d,p)\times(P,D,Q)_s$. The mathematical expression is

$$\Phi_{\mathrm{P}}(B^s)\phi_p(B)\nabla^d\nabla_s^D y_t = A(t)+\Theta_Q(B^s)\theta_q(B)\varepsilon_t \tag{1}$$

where, $B^k y_t = y_{t-k}$; $\nabla^d=(1-B^s)^D$ represents the non-seasonal difference; $\nabla_s^D=(1-B^s)^D$ represents the seasonal difference; $\phi_p(B)=1-\phi_1B-\phi_2B^2-\ldots-\phi_pB^p$ is the non-seasonal autoregressive polynomial; $\Phi_P(B^s)=1-\Phi_1B-\Phi_2B^{2s}-\ldots-\Phi_PB^{Ps}$ is the seasonal autoregressive polynomial; $\theta_q(B)$ =1+ $\theta_1B+\theta_2B^2+\ldots+\theta_qB^q$ is the non-seasonal moving average polynomial; $\Theta_Q(B^s)=1+\Theta_1B^s+\Theta_2B^{2s}+\ldots+\Theta_QB^{Qs}$ is the seasonal moving average polynomial. It is usually assumed that the white noise series is ε_t~$WN(0,\sigma^2)$; $A(t)=c$ (constant)or $A(t)=c+\beta t$ (linear trend).

3. Data analysis

3.1. Data preprocessing and error analysis

The study has utilized monthly CPI data from 2021 to 2024 obtained from the National Bureau of Statistics, specifically including: the Consumer Price Index, the CPI for Food, Tobacco and Liquor, the CPI for Clothing, the CPI for Housing, the CPI for Household Articles and Services, the CPI for Transportation and Communication, the CPI for Education, Culture and Entertainment, the CPI for Healthcare, the CPI for Other Articles and Services, the CPI for Non-food, the CPI for Consumer Goods, the CPI for Services, and the CPI excluding Food and Energy. By summarizing each dataset individually, the study aims to forecast the CPI value for the following year.

The downloaded data was organized into a CSV file, with the data for each of the following indices arranged by month: Consumer Price Index (Overall Index), CPI for Food, Tobacco and Liquor, CPI for Clothing, CPI for Housing, CPI for Household Articles and Services, CPI for Transportation and Communication, CPI for Education, Culture and Entertainment, CPI for

DOI: 10.54254/2753-8818/2026.DL33961

Healthcare, CPI for Other Articles and Services, CPI for Non-food, CPI for Consumer Goods, CPI for Services, and CPI excluding Food and Energy.

According to research findings, among the twelve sub-indices of the Consumer Price Index, the weights of Food, Tobacco and Liquor and Housing are the highest, both exceeding 20%, while the weights of Transportation and Communication and Education, Culture and Entertainment also exceed 10%. The remaining indices have relatively smaller weights. Therefore, the following research primarily analyzes the Consumer Price Index (Overall Index), the CPI for Food, Tobacco and Liquor, the CPI for Housing, the CPI for Transportation and Communication, and the CPI for Education, Culture and Entertainment.

According to the comparison results of multiple SARIMA models. The core idea is to screen the optimal model using different indicators. Among them, AICc stands for the corrected Akaike Information Criterion; the smaller its value, the better the balance between the model's fit and complexity. BoxCox_lambda represents the Box-Cox transformation parameter, which is used to improve data distribution (such as normality and homogeneity of variance). LB_p10 and LB_p20 are p-values from the Ljung-Box test, which check whether residuals are white noise—if $p > 0.05$, it indicates no autocorrelation in residuals and sufficient model fitting. Method, Order, and Seasonal refer to the non-seasonal order (p,d,q) and seasonal order (P,D,Q)_s of the SARIMA model.From the residual tests, most models show that LB_p10 and LB_p20 are close to 1 (e.g., in the row of -128.8227102: 0.999999997 and 1), indicating that residuals are white noise and the model has fully extracted information from the data, with errors being random.Regarding model feasibility: Most models pass the residual white noise test, which means the model specifications are reasonable and feasible overall. However, when selecting a model, priority should be given to the one with the smallest AICc. The model SARIMA(1,0,0)(0,0,0,12) (corresponding to AICc = -3052.745466) fits best and has perfect residual tests. It can be seen from this that the model can control errors within a manageable range, and the model is feasible, which provides a solid foundation for predicting relatively reasonable data in the future.

Next, the author will take the overall CPI as an example for a detailed analysis. Figure 1 shows that from lag 1 onwards, the PACF values for all lag orders fall within the blue-shaded 95% confidence interval and are very close to zero, with no significant non-zero values. The insignificance of all PACF values indicates that there is no significant partial autocorrelation in the residuals, meaning the errors are random. This reveals that the model has successfully captured all dynamic information in the data, leaving no systematic deviations or patterns in the residuals. Combined with the residual PACF plot in Figure 1, it can be further confirmed that the residuals conform to the white noise assumption, indicating that the current model specification is reasonable and can adequately describe the data generating process. Moreover, because the residuals are random, it further validates that when using this model for forecasting, errors will not be systematically amplified, and the reliability of the prediction intervals is relatively high.

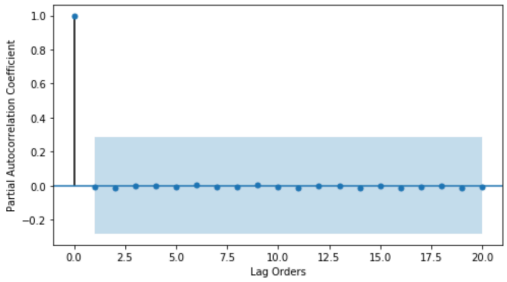


Figure 1. Partial autocorrelation function plot

DOI: 10.54254/2753-8818/2026.DL33961

The horizontal axis in Figure 2 represents the theoretical quantiles, which are the quantiles of the standard normal distribution. The vertical axis represents the ordered sample values, which are the sorted quantiles of the model residuals. The red straight line indicates the perfect linear relationship that theoretical quantiles and residual quantiles should exhibit under an ideal normal distribution. The blue scatter points represent the quantiles of the actual residuals. According to Figure 2, the vast majority of data points (especially in the middle part) closely align with the red straight line, indicating a high degree of fit between the main body of the residuals and the normal distribution. This suggests that the main part of the error distribution is well-behaved, with most residuals conforming to the normality assumption. It demonstrates that the model fits the main body of the data well, and the errors are primarily random.

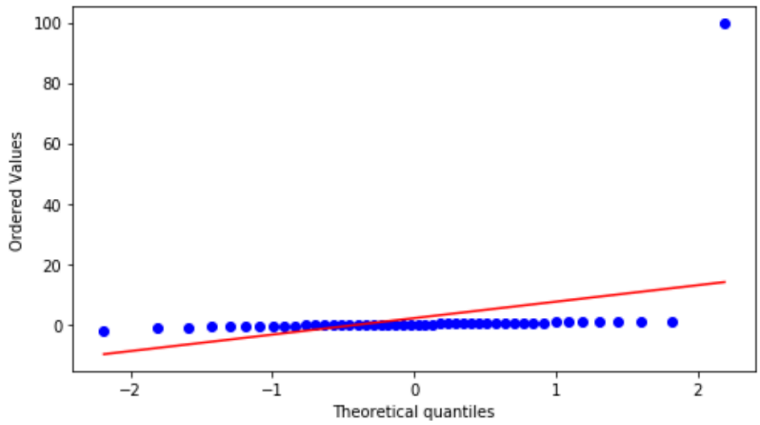


Figure 2. Residual QQ plot

3.2. Numerical presentation

Figures 3 illustrate the month-on-month changes in China's CPI and the predicted values of its five major sub-categories from January to December 2025. The annual average of the overall CPI is approximately 99.9, with more than half of the months below 100, indicating that the overall price level shows a mild downward or low-level operation trend, and inflation pressure is not significant. Moreover, the standard deviation of the overall CPI is about 0.37, reflecting that price remain generally stable without severe inflation or deflation, which shows that the CPI is stable throughout the year with a small fluctuation range. For various sub-categories, the CPI for housing shows the most obvious changes. The annual average is about 112.85, and the standard deviation is about 9.8, with a significant fluctuation range, especially the obvious surge at the end of the year. This is highly related to seasonal factors such as the peak rental season, heating season, and decoration demand before the Spring Festival at the end of the year. It also reflects that the prices of the real estate and related consumer markets are significantly affected by seasonal and policy factors, and the year-end price peak driven by seasonality is the core of the annual trend.

The CPI for transportation and communication also fluctuates obviously, with a standard deviation of about 3.7, which is directly related to factors such as returning home at the end of the year, peak tourism season, and oil price adjustments. The CPI for food, tobacco, and alcohol fluctuates slightly, with a standard deviation of about 0.4. Its characteristics of low average value and low fluctuation mean that prices do not fluctuate sharply even during peak consumption seasons such as the Spring Festival, reflecting the effect of policies to ensure supply and stabilize prices, which provides a solid foundation for overall price stability. The CPI for education, culture, and entertainment has the smallest fluctuation, with a standard deviation of about 0.2. The demand for

DOI: 10.54254/2753-8818/2026.DL33961

education, culture, and entertainment consumption is less sensitive to prices, and prices are less affected by seasonal factors. Therefore, it can be concluded that the end of the year is the core node of price fluctuations throughout the year. Seasonal demand for housing, transportation, and communication is the main driving force, while food, tobacco, alcohol, education, culture, and entertainment are the least affected by seasonality and can control price stability. From a policy perspective, it can be concluded that for the seasonal fluctuations in housing, transportation, and communication, methods such as advance reserves and price regulation can be used to stabilize the year-end price peak and stabilize residents' inflation expectations.

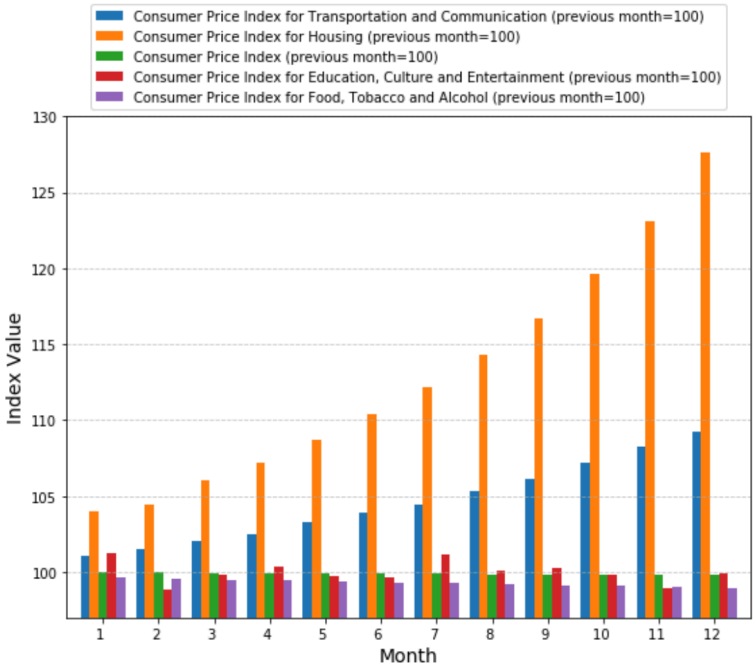


Figure 3. Display chart of important CPI impact indicators and overall value prediction results

4. Conclusion

This experiment uses the SARIMA time series model to predict the resident index. First, error analysis and feasibility evaluation of the model are conducted. Based on the results of AICc, Boxcox_lambda, LB_p10, LB_p20, PACF plots, and QQ plots, it is concluded that the model has good feasibility for CPI prediction. Then, the predicted overall CPI values and several important CPI values are studied in combination with seasonal influencing factors, basic economic principles, policy guidelines, etc., and a relatively reasonable analysis result for the predicted values is obtained. At the same time, the model also has certain limitations. Although the main part of the residuals is close to the red straight line and approximately follows a normal distribution, a point in the upper right corner obviously deviates from the straight line and is much higher than the predicted value. This indicates that the SARIMA model has a strong dependence on data distribution and stationarity, and can only describe the linear autocorrelation and seasonal patterns of CPI, but cannot capture the nonlinear and external driving factors it is subject to. For capturing nonlinear factors, on the basis of using the SARIMA model, dummy variables can be introduced to represent specific periods with outliers. Alternatively, more complex models can be used, such as the GARCH model to handle heteroscedasticity, or nonlinear models to fit the data. This makes the CPI forecast more accurate and reliable.

DOI: 10.54254/2753-8818/2026.DL33961

References

[1] Xu Kun, Lu Qianqian & Xu Guangjian. (2019). CPI Weights, Structural Price Increases, and Price Perception. Price: Theory & Practice, (09), 4-10.
[2] Yuan Lanlan & Zhang Aiwu. (2023). Research on CPI Forecasting Methods Based on Fuzzy Time Series Models. Journal of Mathematics in Practice and Theory, 53(01), 76-82.
[3] Yao Jinhai & Zou Jiajun. (2022). Construction and Numerical Simulation of an SVM-ARIMA Model for CPI Forecasting. Statistics & Decision, 38(21), 48-52.
[4] Yan Yan, Zhang Zhixin & Zhang Hongyao. (2023). CPI Forecasting Based on Internet Search Data Monitoring. Journal of Shenyang University of Technology (Social Science Edition), 16(04), 365-370.
[5] Ma Yao, Sun Zhongyu & Zou Yiling. (2021). Analysis and Forecast of CPI in China, the United States, and Germany Based on the ARIMA Model. Economic Research Guide, (08), 1-4+14.
[6] Wu Xindi, Fang Yiming & Hu Yanrong. (2020). Dynamic Relationship Analysis Between Agricultural Product Price Fluctuations in China and CPI Based on the SVAR Model: A Case Study of Japonica Rice, Corn, and Soybeans. Journal of Southern Agriculture, 51(06), 1485-1492.
[7] Hu Sirong. (2020). Discussion on the Economic Models of Guangdong and Guangxi Based on CPI Index Analysis. Marketing Circle, (25), 121-122.
[8] Zhou Yi & Liu Shujun. (2020). Analysis of Factors Influencing CPI Based on Multiple Regression Model. Modern Economic Information, (08), 3-4+6.
[9] Wang Xiangnan. (2025). Insurance in Household Consumption and Its Impact: Based on US CPI Data. Insurance Studies, (12), 6-24.
[10] Wu Yue. (2024). The Impact of International Oil Prices on China's Transportation and Communication CPI: An Empirical Analysis Based on the TVP-VAR Model. Shanghai Business, (06), 22-25.
[11] Wang Yi, & Huang Peihuang. (2023). Prediction of Typhoon Frequency Based on SARIMA Time Series Model. Computer Science and Application, 13, 2464.
[12] Divisekara, R. W., Jayasinghe, G. J. M. S. R., & Kumari, K. W. S. N. (2020). Forecasting the red lentils commodity market price using SARIMA models. SN Business & Economics, 1(1), 20.