Proceedings of ICFTBA 2024 Workshop: Finance's Role in the Just Transition
DOI: 10.54254/2754-1169/141/2024.GA18853

Application of SARIMA Model: Forecasting CPI and PPI Inflation Rates in the U.S.

Xingjie Yao[1,a,*]

[1]*Department of Actuarial Science and Statistics, University of Toronto, Toronto, Canada*
a. xingjie.yao@mail.utoronto.ca
**corresponding author*

Abstract: This research seeks to improve the precision of future inflation rate forecasts through a comprehensive analysis of the time series data of the U.S. Consumer Price Index (CPI) and Producer Price Index (PPI). This holds significant reference value for the government when formulating economic policies, monetary policies, and fiscal decisions, especially when facing economic fluctuations or external shocks, as it can provide more accurate inflation expectations, thereby enabling the formulation of more forward-looking regulatory measures. The data for this research is sourced from the FRED database, which contains detailed historical CPI and PPI data. This research utilized the traditional SARIMA model for time series analysis and performed thorough data preprocessing and trend analysis to guarantee the stability and predictability of the data. The results indicate that the SARIMA model exhibits different effects in capturing the trend changes of CPI and PPI inflation rates: first, the SARIMA model demonstrates high accuracy in identifying and predicting the long-term trend of CPI; second, it shows stronger adaptability and accuracy in short-term predictions of PPI change rates. This indicates that different economic indicators respond differently to the model. In the future, it may be worth considering the introduction of other advanced forecasting models, such as machine learning algorithms and hybrid models, to further enhance the predictive capability regarding changes in the inflation rate, thereby providing policymakers with more precise data support.

Keywords: CPI, PPI, SARIMA model, Time series analysis.

1. Introduction

Inflation is crucial to the economy of every country because it is closely related to the overall financial markets and policies of the nation. The CPI is a critical macroeconomic indicator that primarily measures fluctuations in the prices of consumer products and services that are relevant to the daily lives of residents. Conversely, the PPI is employed to determine fluctuations in production costs over a specific time frame. It can be said that the CPI and PPI are important indicators for measuring inflation. The government and central bank refer to these indicators to formulate monetary policies and economic regulation measures to address potential inflation risks [1-2]. Therefore, studying the inflation rates of CPI and PPI is of great significance for understanding inflation [3]. In previous studies, the SARIMA model has been widely used in fields such as economics and finance due to its ability to capture seasonal components and trends in time series data. Consequently, the SARIMA model is employed in this research to forecast and analyze the CPI and PPI. First, by using time series analysis tools such as residual plots, autocorrelation function (ACF), and partial autocorrelation function (PACF), select the most suitable SARIMA model. Then, the CPI and PPI datasets are divided into training and testing sets in a 70% to 30% ratio. By comparing the prediction graphs and evaluation metrics, the differences in the prediction performance of the SARIMA model under different metrics are analyzed, providing a reference for future economic forecasting and policymaking.

2. Data

The data for this research comes from the publicly available database of FRED (Federal Reserve Economic Data), specifically the datasets for the Consumer Price Index and the Producer Price Index by Commodity [4]. The data ranges from February 1, 1955, to March 1, 2024, and this dataset reports the CPI and PPI for all items in the United States monthly. First, during the data processing stage, this research converts the observation_date column to date format to ensure the accuracy of subsequent date handling. Since this research only focuses on data from the last 10 years, it has filtered out data from February 2014 to March 2024. In addition, to conduct predictive analysis, a lagged variable of CPI, CPI_t_minus_1, was created, representing the CPI value of the previous period, and missing data rows were deleted to ensure data integrity. For the PPI data, this research also used the same processing method. The basic information of the processed data is as follows in Table 1.

Table 1: Descriptive Statistics of CPI and PPI

Statistic	CPI	PPI
Number of Observations	122	120
Mean	0.2379	0.2443
Standard Deviation	0.362	0.2209
Minimum	-0.6687	-0.0952
25th Percentile	-0.0033	0.1018
Median	0.2206	0.1793
75th Percentile	0.4689	0.313
Maximum	1.3736	1.028

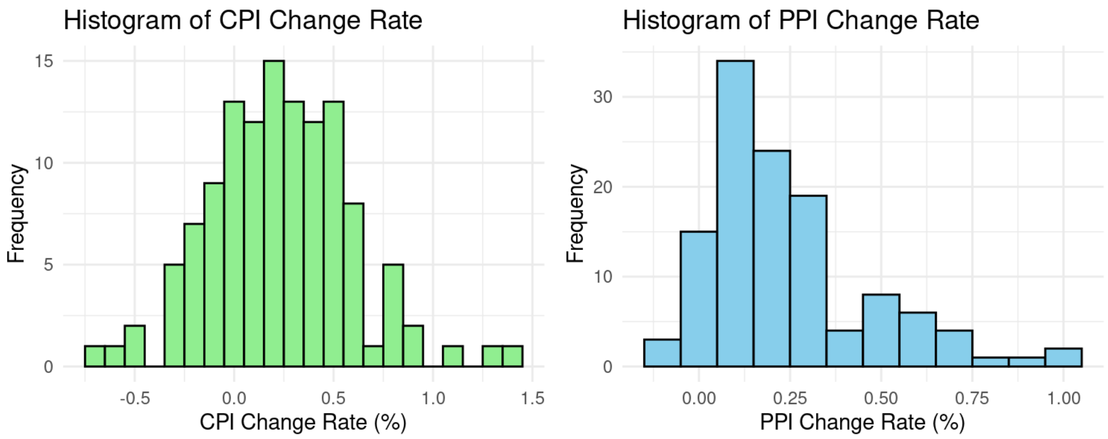

Figure 1: CPI Change Rate Histogram and PPI Change Rate Histogram

Through Table 1, it can be seen from Figures 1 that the overall CPI presents a normal distribution. During the observation period, the CPI fluctuated significantly, reaching a peak of 1.3736 and a

DOI: 10.54254/2754-1169/141/2024.GA18853

trough of -0.6687. However, for most of the time, the range of CPI changes is between -0.0033 and 0.4689, indicating that CPI changes are relatively stable in most months and show a steady upward trend. However, the PPI change rate shows a slight right-skewed distribution overall. During the observation period, the PPI change rate fluctuated significantly, reaching a peak of 1.028 and a trough of -0.0952. Most PPI change rates were concentrated between 0 and 0.25, indicating that the PPI change rate was relatively small most of the time.

3. Methodology

This research uses the SARIMA Model to analyze data on the US CPI and PPI from February 2014 to March 2024. The SARIMA model is a common statistical model, primarily used for future predictions using existing time series data [5]. It is capable of accurately capturing the seasonal and trend components of the time series. Through reasonable parameter selection, the SARIMA model can adapt to various types of time series data, and thus it has been widely employed in fields such as economics, finance, demographics, and public health [6].

In order to use the SARIMA model for forecasting, the following basic steps are usually required:

1. Data visualization: Understanding the overall trends and fluctuations of data by plotting data charts.

2. Data Transformation: After observing the trend of the data, if it is found to be unstable or highly volatile, it is necessary to perform transformations or differencing operations to eliminate trend and seasonal effects, making the data stable.

3. Model order identification: Establish the order of the autoregressive (AR), differencing (I), and moving average (MA) components in the model.

4. Parameter estimation: Determine the specific coefficients of the model through parameter estimation.

5. Model Diagnosis and Selection: The best model is ultimately determined by comparing indicators such as AIC, BIC, and RMSE. Generally, the smaller the values of these indicators, the better.

In the SARIMA model, the parameters p, d, and q represent the orders of the autoregressive (AR) part, the differencing (I) part, and the moving average (MA) part, respectively. In the seasonal component, P, D, Q, and (m) represent the seasonal autoregressive term, seasonal differencing term, seasonal moving average term, and the length of the seasonal cycle, respectively. By reasonably selecting these parameters, the SARIMA model is capable of producing high-quality prediction results and effectively managing time series data with varying structures.

After comprehensive consideration, this research chose the SARIMA (2,1,2) (1,0,0) (12) model to predict CPI and PPI for the following reasons: Due to the significant fluctuations and instability of this time series, direct modeling may lead to inaccurate results. After applying first differencing, the series became more stable, so d=1 was chosen. However, considering the limitations of the SARIMA model, other forecasting models such as Prophet and LSTM can be introduced in the future to enhance the accuracy and adaptability of predictions [7-9]. After analyzing the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots, it was found that lag 2 exhibits significant autocorrelation. Therefore, p=2 was chosen to capture this autocorrelation, and q=2 was selected to model the dependence of the error terms. Due to the obvious annual seasonal characteristics of the data, a seasonal component (1, 0, 0) (12) was introduced into the model, where (12) indicates a seasonal cycle of 12 months. Moreover, by using a seasonal autoregressive term P=1, the impact of the annual cycle on the data can be better captured.

DOI: 10.54254/2754-1169/141/2024.GA18853

4. Results

4.1. SARIMA for CPI and PPI

This research conducted parameter estimation for the SARIMA (2,1,2) (1,0,0) (12) model to understand the impact of the autoregressive terms on CPI and PPI data. Table 2 and Table 3 respectively show the parameter estimation results for CPI and PPI.

Table 2: The results of the SARIMA (2,1,2) (1,0,0) (12) model's CPI parameter estimation

Parameter	Estimate	Standard Error
AR(1)	-0.204	0.101
AR(2)	0.252	0.1
MA(1)	-0.035	0.043
MA(2)	-0.907	0.044
SAR(1)	0.345	0.093
$sigma^2$	0.0717	NA

Table 3: The results of the SARIMA (2,1,2) (1,0,0) (12) model's PPI parameter estimation

Parameter	Estimate	Standard Error
AR(1)	-0.232	1.467
AR(2)	-0.017	0.146
MA(1)	-0.71	1.46
MA(2)	-0.29	1.46
SAR(1)	0.241	0.123
$sigma^2$	0.0147	NA

According to Tables 2 and 3, it can be seen that there is a difference in the residual variances of the CPI and PPI models. The residual variance of the CPI model is 0.0717, while the residual variance of the PPI model is 0.0147. In contrast, the residual variance of the PPI is smaller, indicating that its model fits better than the CPI model, resulting in lower prediction errors. However, the residual variance is only one aspect of measuring model performance; it is necessary to combine it with other evaluation criteria (such as AIC or BIC) to more comprehensively assess the model's quality. Moreover, both models show a more pronounced seasonal impact.

4.2. Prediction

Next, this research used the SARIMA (2,1,2) (1,0,0) (12) model to predict the CPI and PPI data separately. The following Figure 2 and Figure 3 show the comparison line chart of the training set, predicted values, and actual test values for CPI and PPI.

DOI: 10.54254/2754-1169/141/2024.GA18853

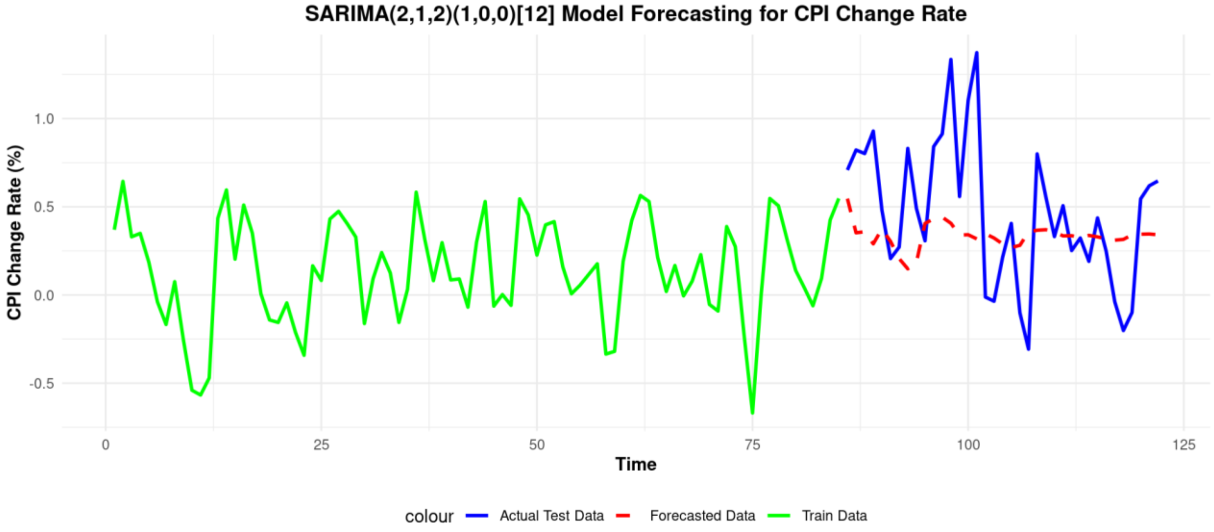

Figure 2: Comparison of Actual and Predicted CPI Values (70% Training Data, 30% Test Data)

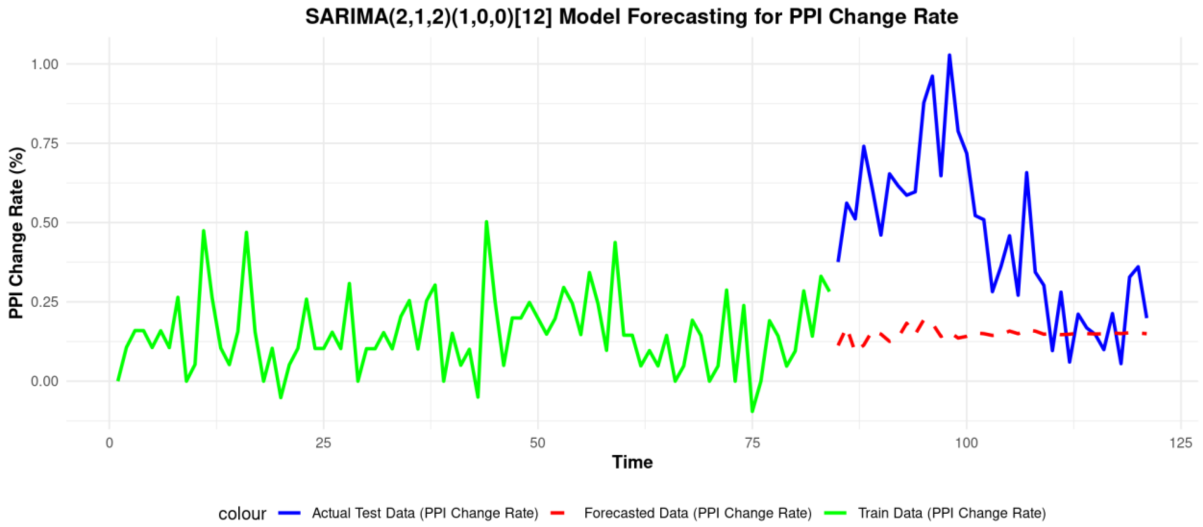

Figure 3: Comparison of Actual and Predicted PPI Values (70% Training Data, 30% Test Data)

It can be seen that the CPI model's predicted values are relatively close to the actual values within the data stable interval, but in cases of significant data fluctuations, the model still struggles to accurately capture the trend. Besides, the predicted values of PPI in certain intervals show significant deviations from the actual test data, indicating that the volatility of PPI data is high, and the selected model struggles to fully capture these changes.

4.3. Evaluation Indicators

To further evaluate the model's performance, this research calculated commonly used evaluation metrics such as MSE and RMSE, with specific results shown in Table 4.

DOI: 10.54254/2754-1169/141/2024.GA18853

Table 4: Performance Evaluation of CPI and PPI Forecasts: MSE and RMSE Metrics

Metrics	CPI	PPI
MSE	0.173	0.153
RMSE	0.416	0.392

From Table 4, it can be seen that the RMSE and MSE of the PPI model are relatively low, suggesting that the model effectively captures the trends in the data and fits the data well. But the RMSE and MSE of the CPI model are slightly higher, indicating that its prediction errors are relatively large, especially when dealing with fluctuations in the test data, resulting in significant deviations [10].

5. Conclusion

This research conducted time series analysis and forecasting of CPI and PPI using the SARIMA (2,1,2) (1,0,0) (12) model. The results indicate that the SARIMA (2,1,2) (1,0,0) (12) model is better at accurately capturing the overall trend in CPI forecasting. According to Figure 2, it can be seen that the predictions made by the model using 70% of the training data are relatively close to the overall trend of the actual test data, but the prediction performance is somewhat lacking in the intervals with significant data fluctuations. According to Table 4, this model shows better fit and smaller residual variance when predicting PPI, indicating higher accuracy in capturing the rate of change of PPI. However, its predicted values still have significant deviations during certain periods. From Figure 3, It is apparent that the prediction line is excessively flat, suggesting that the model is subject to certain constraints in its ability to accommodate the substantial fluctuations in PPI data.

Overall, the model has its strengths and weaknesses in predicting the changes in CPI and PPI rates, but it still falls short in capturing long-term trends. This may be due to the limitations of the model itself. Although this research has addressed the issue of non-stationarity to some extent through differencing, the SARIMA model's ability to capture highly nonlinear data remains insufficient. Moreover, the SARIMA model may perform poorly when dealing with complex or multi-period seasonal patterns. Therefore, in the future, other forecasting methods such as Prophet and LSTM models can be considered, or more external factors (such as economic and policy factors) can be introduced for a more comprehensive analysis to improve the model's accuracy.

References

[1] *Fama, E. F., & Schwert, G. William. (1977). Asset Returns and Inflation. Journal of Financial Economics, 5(2), 115–146.*

[2] *Stock, J. H., & Watson, M. W. (2003). Forecasting Output and Inflation: The Role of Asset Prices. Journal of Economic Literature, 41(3), 788–829.*

[3] *Ang, A., Bekaert, G., & Wei, M. (2007). Do macro variables, asset markets, or surveys forecast inflation better? Journal of Monetary Economics, 54(4), 1163–1212.*

[4] *Federal Reserve Bank of St. Louis. (2023). FRED Economic Data. Stlouisfed.org; Federal Reserve Bank of St. Louis. Retrie from https://fred.stlouisfed.org/*

[5] *Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: the Forecast Package for R. Journal of Statistical Software, 27(3).*

[6] *Janacek, G. (2003). Non-linear Time Series Models in Empirical Finance. Journal of the Royal Statistical Society: Series D (The Statistician), 52(4), 696–696.*

[7] *Qiu, J., Wang, B., & Zhou, C. (2020). Forecasting stock prices with long-short term memory neural network based on attention mechanism. PLOS ONE, 15(1), e0227222.*

[8] *Moghar, A., & Hamiche, M. (2020). Stock Market Prediction Using LSTM Recurrent Neural Network. Procedia Computer Science, 170, 1168–1173.*

DOI: 10.54254/2754-1169/141/2024.GA18853

[9] Zhang, G. Peter. (2003). Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. Neurocomputing, 50, 159–175.
[10] Young, P., & Shellswell, S. (1972). Time series analysis, forecasting and control. IEEE Transactions on Automatic Control, 17(2), 281–283.