SICASH 2023
SLIIT INTERNATIONAL CONFERENCE ON ADVANCEMENTS IN SCIENCES & HUMANITIES

Forecasting Consumer Price Index in the United States

[1] Witharana, W. W. S. K., [2] Udugama, U. K. D. T. N., [3] Fernando, P. M. R., [4] Kaumadi, H. M. H., [5] Peiris, T. S. G.

[1,2,3,4,5]Department of Mathematics and Statistics, Faculty of Humanities and Sciences,
Sri Lanka Institute of Information Technology, Malabe, Sri Lanka

Corresponding author - *hs21925368@my.sliit.lk

ARTICLE INFO

Article History:
Received: 10 September 2023
Accepted: 01 November 2023

Keywords:
ARIMA; Consumer price index; Forecasting; Box Jenkins; Time series analysis

Citation:

Witharana, W. W. S. K., Udugama, U. K. D. T. N., Fernando, P. M. R., Kaumadi, H. M. H., Peiris, T. S. G. (2023). Forecasting Consumer Price Index in the United States. Proceedings of SLIIT International Conference on Advancements in Sciences and Humanities, 1-2 December, Colombo, pages 290-295.

ABSTRACT

This report presents the Auto-Regressive Integrated Moving Average (ARIMA) model for forecasting the consumer price index (CPI) in US using monthly data from March 2010 to March 2023. The original series was not stationary, but the first difference series was found to be stationary using the Augmented Dicky Fuller test. The best-fitted model was identified based on the significance of the parameters, volatility (sigma^2), log-likelihood, Akaike, Schwartz, and Hannan-Quinn information criterion. Parameters of the fitted model are significantly deviated from zero. The stability of the model has been checked using the roots of the unit root test. Residuals of the fitted model satisfied the randomness but non-constant variance. The monthly forecasted values of CPI from April 2023 to August 2023 are 301.833, 302.444, 303.038, 303.639, and 304.261. The percentage errors of the forecasted values are less than one percent. This method and results provide useful information to policy and market makers for their planning,

hs21925368@my.sliit.lk

https://doi.org/10.54389/RAQA6627

1. INTRODUCTION

The consumer price index (CPI) measures how prices paid by urban consumers have changed on average, over time for a set market basket of consumer goods and services (Fernando, 2023). Consumers can use the CPI to compare the price of a market basket of goods and services this month to that of a similar market basket last month or a year ago. The CPI reflects the spending habits of two distinct population segments: urban wage earners, clerical workers (CPI-W), and all urban consumers (CPI-U). About 80% of all Americans are represented by the CPI-U. All products and services bought by urban households for consumption are included in the CPI. It tracks price changes over more than 200 categories, divided into major eight groups. The CPI includes a variety of user fees, including those for water and sewage, motor registration, tolls, and so on (Nyoui, 2022). Taxes like sales and excise taxes that are directly linked to the costs of goods and services are also included. However, taxes (including income and Social Security taxes) not directly related to purchasing goods and services for consumption are not included in the CPI. The United States Bureau of Labor Statistics (Statistics, 2023) conducts monthly nationwide retail establishment surveys and collects price data on thousands of products. After placing these products into one of the 200 categories for spending money, it is possible to estimate price changes within each category by weighing the items according to their significance. The importance of these categories is then considered, and additional aggregations are done until an overall CPI number is produced. Each month, data is collected from about 87 urban areas across America. 24,000 retail businesses and around 6,000 residential units are sampled. The price and index incorporate taxes incurred when purchasing goods or services. BLS representatives (Basic Life Support) often visit or call individuals to gather data. The weights for each item in a particular place are aggregated to create the index. The weights indicate how significant they are to each population group's spending. Then, localized data is pooled to calculate an average for U.S. cities. Area indices simply track the average change in prices for each area since the base period, not the level of prices in different cities. The CPI is one of the most widely used indicators of inflation and deflation. In contrast to the producer price index (PPI), which tracks changes in the prices paid to American producers of goods and services, the CPI report employs a different survey methodology, price samples, and index weights. Various authors have attempted different models to forecast CPI (Volodymyr et al., 2021; Nyoni, 2022; Konarasinghe, 2022).

The advanced knowledge of CPI would give insight into the effectiveness of economic policies and their performances. Furthermore, forecast CPI would help for monetary policies, fiscal policies and various policies which help to consumers and the country. This research thus aims to develop a simple time series model to short-term prediction of the US consumer price index .

2. MATERIALS AND METHODS

2.1. DATA

Secondary data was obtained from a website called Fred economic data from March 2010 to March 2023(Statistics, 2023).

2.2. METHODOLOGY

In statistics and econometrics, particularly in time series analysis, autoregressive integrated moving average (ARIMA) models are used for forecasting (Chan & Cryer, 2008). An autoregressive-moving average of order p and q, ARMA (p, q) model is given as,

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + \ldots\ldots$$
$$+ a_p Y_{t-p} - b_1 \varepsilon_{t-1} - b_2 \varepsilon_{t-2} - \ldots\ldots - b_q \varepsilon_{t-q} + \varepsilon_t$$

where p is the order of the autoregressive part, q is the order of the moving average part, and is the white noise. The ARMA models are suitable for stationary series. These were generalized for non-stationary series that become stationary by differentiation. The resulting models are called autoregressive integrated-moving average ARIMA (p, d, q) where d is the order of differentiation required for stationary series. To compare models, Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), and Hannan–Quin Information Criterion (HQC) were used (Chan, 202)

The selection of the best-fitted model was based on the approach recommended in the Box-Jenkins methodology with the steps of identification, estimation, diagnostics, and forecasting.

3. RESULTS AND DISCUSSION

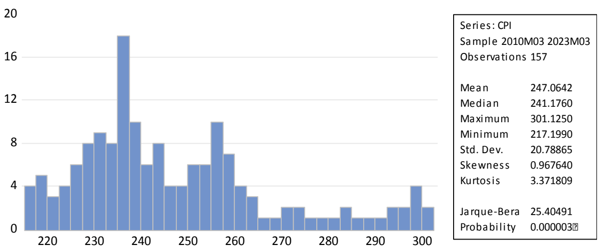


Figure 1: Basic Statistics

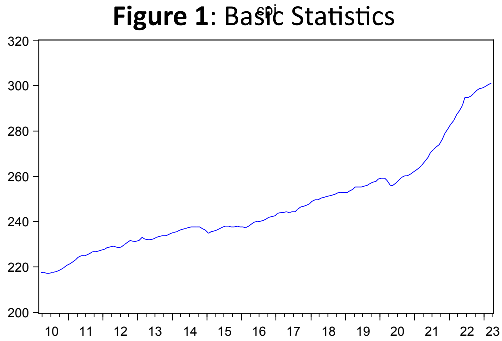


Figure 2: *Original Series*

Figure 1 shows the basic statistics of the series. The CPI of the US varies from a minimum of 217.1990 to a maximum of 301.1250 with a mean of 247.0642 and a standard deviation of 20.78865. The time series plot in Figure 2 shows the CPI of the US has been increasing from March 2010 to March 2023 confirming that the series is not stationary. This was confirmed using the correlogram (Figure 3) and the Augmented Dicky Fuller test ($p > 0.05$) of the original series. In the correlogram of the original series, the ACF declines very slowly, and the indications are outside the 95% CI. The PACF drops immediately after the 1st lag. Therefore, the 1st difference series was taken for further analysis.

However, it was found that the 1st difference series is stationary ($t = -6.146$, $p = 0.000$). In Figure 4, the ACF of the stationary series has an exponential decay pattern and almost about 12 ACFs are significant. In PACF only 1st, 3rd, 5th, and 10th lags are significant. Thus, the following 3 models were identified as parsimonious models. The comparison between the 3 estimated models is shown in Table 1.

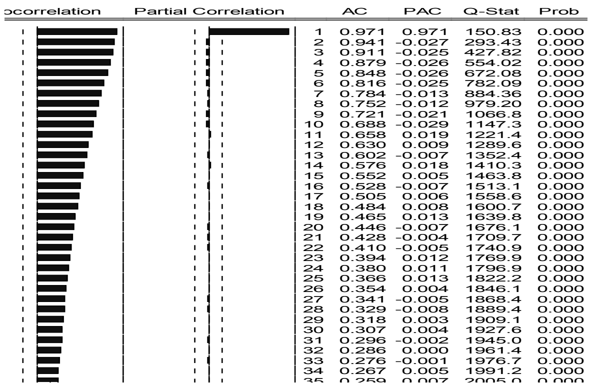


Figure 3: Correlogram of the Original Series

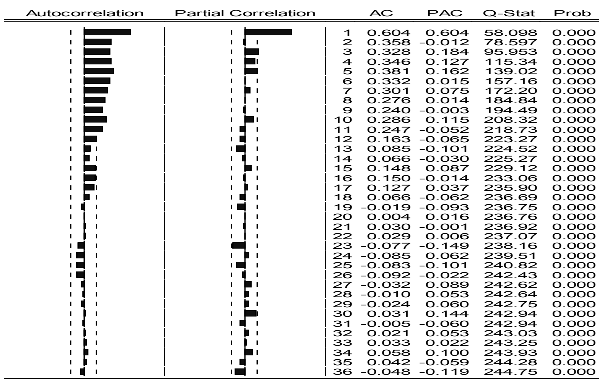


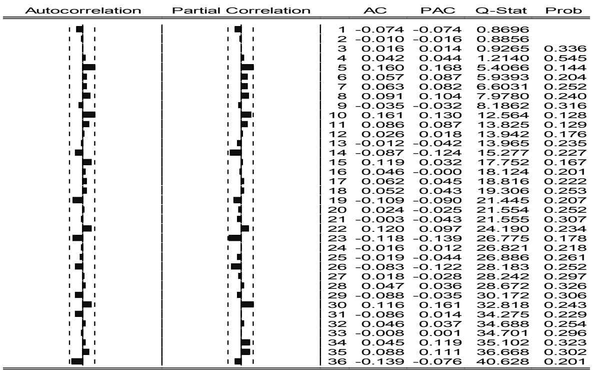
Figure 4: Correlogram of the first difference series (Stationary series)

Table 1: Summary of the four estimated models

	ARIMA (1,1,1)	ARIMA (1,1,2)	ARIMA (1,1,3)
Significance of AR (1)	Significant	Significant	Significant
Significance of MA (1)	Not Significant	-	-

Significance of MA (2)	-	Significant	-
Significance of MA (3)	-	-	Significant
Volatility	0.326767	0.320395	0.326352
Log-likelihood	-134.3372	-132.8370	-134.2368
Akaike	1.773554	1.754321	1.772267
Schwartz	1.851756	1.832522	1.850468
Han-nan-Quinn	1.805316	1.786083	1.804029

According to Table 1, ARIMA (1,1,2) was found to be the best-fitted model since two of the coefficients are significant, volatility is the lowest, and Akaike, Schwartz, and Hannan-Quinn information criteria are lowest in ARIMA (1,1,2). The third and final step of the Box-Jenkins method is diagnostics of errors and forecasting. The correlogram of the residuals of the ARIMA (1,1,2) model is shown in Figure 5.



Autocorrelation / Partial Correlation (lag)	AC	PAC	Q-Stat	Prob
1	-0.074	-0.074	0.8696	
2	-0.010	-0.016	0.8856	
3	0.016	0.014	0.9265	0.336
4	0.042	0.044	1.2140	0.545
5	0.160	0.168	5.4066	0.144
6	0.057	0.087	5.9393	0.204
7	0.063	0.082	6.6031	0.252
8	0.091	0.104	7.9780	0.240
9	-0.035	-0.032	8.1862	0.316
10	0.161	0.130	12.564	0.128
11	0.086	0.087	13.825	0.129
12	0.026	0.018	13.942	0.176
13	-0.012	-0.042	13.965	0.235
14	-0.087	-0.124	15.277	0.227
15	0.119	0.032	17.752	0.167
16	0.046	-0.000	18.124	0.201
17	0.062	0.045	18.816	0.222
18	0.052	0.043	19.306	0.253
19	-0.109	-0.090	21.445	0.207
20	0.024	-0.025	21.554	0.252
21	-0.003	-0.043	21.555	0.307
22	0.120	0.097	24.190	0.234
23	-0.118	-0.139	26.775	0.178
24	-0.016	0.012	26.821	0.218
25	-0.019	-0.044	26.886	0.261
26	-0.083	-0.122	28.183	0.252
27	0.018	-0.028	28.242	0.297
28	0.047	0.036	28.672	0.326
29	-0.088	-0.035	30.172	0.306
30	0.116	0.161	32.818	0.243
31	-0.086	0.014	34.275	0.229
32	0.046	0.037	34.688	0.254
33	-0.008	0.001	34.701	0.296
34	0.045	0.119	35.102	0.323
35	0.088	0.111	36.668	0.302
36	-0.139	-0.076	40.628	0.201

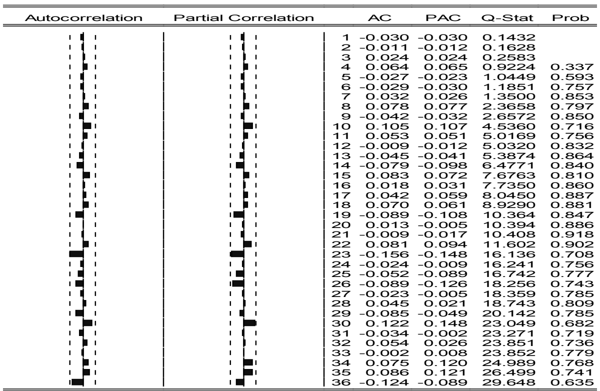
Figure 5: Correlogram of the residuals

All the autocorrelations of the residuals at different lags are not significantly different from zero except at lag 5 and 8. However, in the PACF of the residuals, only the partial autocorrelation at lag 5 is not significantly different from zero. Therefore, an additional parameter of AR (5) and MA (5) were included separately for the identified best-fitted model ARIMA (1,1,2). Accordingly, the following two models, ARIMA (1,1,2) + AR (5) and ARIMA (1,1,2) + MA (5) were identified as the most appropriate models for the stationary series achieved by taking the first difference of the original series. Therefore, the selected model was re-estimated, and the summary results are shown in Table 2:

Parameters and Indicators	ARIMA (1,1,2) + AR (5)	ARIMA (1,1,2) + MA (5)
Significance of AR (1)	Significant	Significant
Significance of MA (2)	Not Significant	Not Significant
Significance of AR (5)	Significant	Not applicable
Significance of MA (5)	Not applicable	Not Significant
Volatility	0.303211	0.312313
Log-likelihood	-128.6482	-130.8744
Akaike	1.713439	1.741980
Schwartz	1.811190	1.839732
Hannan-Quinn	1.753141	1.781683

Table 2: The Summary of the re-estimated models

According to table 2, **ARIMA (1,1,2) + AR (5)** was considered the best-fitted model since two of the coefficients are significant, volatility is lowest and Akaike, Schwartz, and Hannan-Quinn information criteria are lowest in that model.



Autocorrelation / Partial Correlation (lag)	AC	PAC	Q-Stat	Prob
1	-0.030	-0.030	0.1432	
2	-0.011	-0.012	0.1628	
3	0.024	0.024	0.2583	
4	0.064	0.065	0.9224	0.337
5	-0.027	-0.023	1.0449	0.593
6	-0.029	-0.030	1.1851	0.757
7	0.032	0.026	1.3500	0.853
8	0.078	0.077	2.3658	0.797
9	-0.042	-0.032	2.6572	0.850
10	0.105	0.107	4.5360	0.716
11	0.053	0.051	5.0169	0.756
12	-0.009	-0.012	5.0320	0.832
13	-0.045	-0.041	5.3874	0.864
14	-0.079	-0.098	6.4771	0.840
15	0.083	0.072	7.6763	0.810
16	0.018	0.031	7.7350	0.860
17	0.042	0.059	8.0450	0.887
18	0.070	0.061	8.9290	0.881
19	-0.089	-0.108	10.364	0.847
20	0.013	-0.005	10.394	0.886
21	-0.009	-0.017	10.408	0.918
22	0.081	0.094	11.602	0.902
23	-0.156	-0.148	16.136	0.708
24	-0.024	-0.009	16.241	0.756
25	-0.052	-0.089	16.742	0.777
26	-0.089	-0.126	18.256	0.743
27	-0.023	-0.005	18.359	0.785
28	0.045	0.021	18.743	0.809
29	-0.085	-0.049	20.142	0.785
30	0.122	0.148	23.049	0.682
31	-0.034	-0.002	23.271	0.719
32	0.054	0.026	23.851	0.736
33	-0.002	0.008	23.852	0.779
34	0.075	0.120	24.989	0.768
35	0.086	0.121	26.499	0.741
36	-0.124	-0.089	29.648	0.635

Figure 6: Correlogram of residuals of the best fitted model

Figure 6 represents the correlogram of residuals of the re-estimates model; there are no values that are crossing the lines. That means the correlogram of residuals is flat. P-values for the Q-statistics are all over 0.05. Therefore, we must accept the null hypothesis, which means residuals are white noise. However, the Breusch-Pagan-Godfrey test for heteroscedasticity (p value = 0.0000 < 0.05) shows that the residuals are significantly deviated from constant variance. Figure 7 shows AR roots and MA roots lie inside the circle. This means the ARIMA process is (covariance) stationary and invertible.

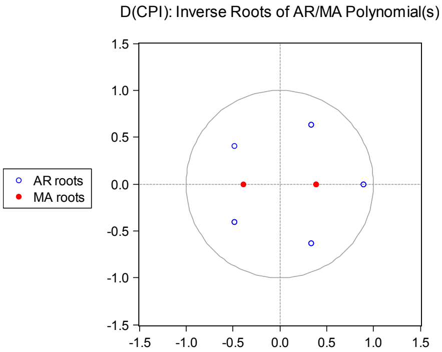


Figure 7: Inverse roots of the model

Using the static forecasting method, the forecasted CPI of the US for the next 5 months in 2023 is shown in second column of Table 3. The actual values are shown in the third column (*Statistics, 2003). It was found that there is hardly any difference between values predicted using the best fitted model and the actuals. The percentage error is much less than 1%.

Table 3:

Month	Forecast-ed	Actual	Percentage Error
April2023	301.833	303.363	0.504
May 2023	302.444	304.127	0.553
June 2023	303.038	305.109	0.679
July 2023	303.639	305.691	0.671
August 2023	304.261	307.026	0.901

Table 3: Forecasted Vs Actual CPI

4. CONCLUSIONS

The original series of the consumer price index was not stationary, but the 1st difference series was found to be stationary. Between the three models that were estimated, was found to be a good model and it was further modified and was the best-fitted model for forecasting the consumer price index for the next 5 months. The actuals and the forecasted values for the training data set as well as for the independent data set are not significantly deviated. Thus, this model is recommended to forecast future CPI in the United States.

REFERENCES

Chan, K.-S., & Cryer, J. (2008). *Time Series Analysis.* New York : Springer Science+Business Media.

Department of Statistics & Actuarial Science, University of Iowa. (2008). *Time Series Analysis With Applications in R* (Second ed.). 8 Springer Science+Business Media. https://mybiostats.files.wordpress.com/2015/03/time-series-analysis-with-applications-in-r-cryer-and-chan.pdf

Fernando, J. (2023, August 11). *Consumer Price Index (CPI) Explained: What It Is and How It's Used*. investopedia: https://www.investopedia.com/terms/c/consumerpriceindex.asp

Kenton, W. (2023, June 28). *Goodness-of-Fit*. investopedia: https://www.investopedia.com/terms/g/goodness-of-fit.asp#:~:text=There%20are%20multiple%20types%20of,specific%20distribution%20of%20a%20population.

Konarasinghe, K. M. U. B. (2022). Modeling Consumer Price Index of Thailand. *Journal of New Frontiers in Education and Social Sciences,2* (1) 1-15

Nyoni, T. (2022). *Time Series Modeling and Forecasting of the Consumer Price Index in Belgium.* https://mpra.ub.uni-muenchen.de/92414/

Oren, B. , Jonathan, B., Itamar, C. Eliya, C., Allon, H. and Noam , K. (2023). Forecasting CPI inflation components with Hierarchical Recurrent Neural Networks. *International Journal of*

Forecasting, 39 (3), 1145-1162.

States, G. (2023). *U.S. Bureau of Labour Statistics*. https://www.bls.gov/opub/mlr/2016/

Statistics, U. B. (2023, June). *Consumer Price Index for All Urban Consumers: All Items in U.S. City Average (CPIAUCSL)*. Fred Economic Data: https://fred.stlouisfed.org/series/CPIAUCSL

Statistics, U. B. (2023, June). *United States Consumer Price Index (CPI)*. Trading Economics: https://tradingeconomics.com/united-states/consumer-price-index-cpi

Volodymyr, S., Alexey, H., Larysa, S., and Leonid, D, (2021). A forecasting the consumer price index using time series model. *SHS Web of Conferences, 107*. 1002.