International Journal of Mathematics and Computer Research

ISSN: 2320-7167

Volume 12 Issue 11 November 2024, Page no. – 4586-4591

Index Copernicus ICV: 57.55, Impact Factor: 8.316

DOI: 10.47191/ijmcr/v12i11.06

Step Function Intervention Analysis on Indonesia's Consumer Price Index (CPI)

Fitri Jessica Sanjani Harefa[1], Tarno[2], Di Asih I Maruddani[3]

[1,2,3] Department of Statistics, Universitas Diponegoro, Indonesia

ARTICLE INFO

Published Online:
25 November 2024

Corresponding Author:
Tarno

ABSTRACT

The state of the country's economy can be analyzed based on the economic data of some countries, one of the data closely related to the state of the country's economy is the CPI (Consumer Price Index). The Consumer Price Index is the main indicator used to measure a country's inflation. The consumer price index was in a very downward trend from January 2020 to January 2023. The change in the consumer price index was due to the Covid-19 epidemic. Changes in CPI data are responded to by intervention analysis, where forecasting of future conditions is done with intervention analysis. Interference analysis is time series data that is affected by external and internal factors that can cause changes in the model at the same time. The intervention model has 2 functions, namely step and pulse functions. The step function is an intervening phenomenon that produces long-term effect, while the pulse function is an intervening phenomenon that produces short-term (temporary) effect. The research data variable taken is the monthly data of the Indonesian CPI for the period January 2014 to January 2023. The data uses a step function intervention model because there is an extreme decline in the period January 2020 (T = 73) to January 2023 (T = 109) in a long period of time. The resulting intervention model is SARIMA(0,1,1)(0,0,1)12 in order b=0, s=0, and r=1 with a model accuracy using SMAPE of 0,58% which shows the model is classified as very good.

KEYWORDS: Consumer Price Index; SARIMA; Forecasting; Intervention Analysis; Step Function; SMAPE

I. INTRODUCTION

On December 31, 2019, the coronavirus (Covid-19) outbreak spread rapidly worldwide in a relatively short time, originating from Wuhan City, Hubei Province, China [1]. According to the Ministry of Health [2], Covid-19 was detected in Indonesia on March 2, 2020, when two Japanese nationals were infected with the virus.

Indonesia's economy contracted by 2.07% in 2020 due to the Covid-19 pandemic [3]. This economic contraction led to an increase in the prices of goods and services, known as the inflation rate, which is closely related to the Consumer Price Index (CPI). The CPI is an index that shows the price increase of goods and services purchased by consumers over a certain period [4]. The CPI is useful in determining the level of income, price increases, and providing a measure of production costs. Therefore, determining the inflation rate requires CPI data as a reference in calculating changes in the prices of goods and services.

Indonesia's CPI data experienced an extreme decline from January 2020 to January 2023, indicating a drastic drop in the CPI due to the Covid-19 outbreak. The impact of this CPI decline could be felt for quite a long time. This extreme data pattern change was analyzed using intervention analysis. This intervention has two functions: the pulse function is used for short-term interventions, and the step function is used for long-term interventions [5].

The step function in intervention analysis is particularly suited to modeling long-term shifts in a time series that result from a single, sustained event, such as the Covid-19 pandemic. It allows researchers to: (1) identify permanent shifts: capture the lasting impact of a shock or intervention; (2) estimate the duration and intensity: determine how long the impact persists and to what extent.

Previous researchers who used the step function intervention model include [6], who obtained the ARIMA(0,1,1) model with intervention order b=4, s=0, and r=2 to predict the number of postal item deliveries to Semarang. In a study by [7], the ARIMA(0,1,1) model with intervention order b=1, s=1, and r=0 was obtained to predict electricity usage.

Analyzing Indonesia's CPI with a step function intervention model could reveal valuable insights into the pandemic's long-term impact on inflation. The model's parameters would be tuned to reflect the unique characteristics of the CPI's response, capturing both the initial shock and the subsequent economic recovery or adjustment phases.

Through this approach, researchers and policymakers could gain a deeper understanding of inflationary pressures and inform better decisions in stabilizing prices post-crisis.

II. THEORETICAL FRAMEWORK

2.1 Consumer Price Index

The Consumer Price Index (CPI) is closely related to the economic stability of a country. According to BPS [8], the CPI is derived from a collection of prices of products and services consumed by households over a specific period, measuring the mean price changes over time. The formula for calculating the CPI value is as follows:

$$CPI_n = \frac{\sum_{i=1}^{k} \frac{P_{ni}}{P_{(n-1)i}} P_{(n-1)i}.Q_{0i}}{\sum_{i=1}^{k} P_{0i}.Q_{0i}}$$

Description:

CPI_n : Consumer Price Index for period n

P_{ni} : Price of item i in period n

$P_{(n-1)i}$: Price of item i in period n-1

$P_{(n-1)i}.Q_{0i}$: Consumption value of item i in period n-1

$P_{0i}.Q_{0i}$: Consumption value of item i in the base year

k : Number of commodity types in the goods package

Table 1. Plot ACF dan PACF

	Model	ACF Behavior	PACF Behavior
Non Seasonal	AR (p)	Slowly decreases after lag p	Cuts off after lag p
	MA (q)	Cuts off after lag q	Slowly decreases after lag q
	ARMA (p,q)	Slowly decreases after lag (p,q)	Slowly decreases after lag (p,q)
	ARIMA (p,d,q)	Slowly decreases after lag (p,q) after differencing	Slowly decreases after lag (p,q) after differencing
Seasonal	SAR (P)	Slowly decreases after lag P^s	Cuts off after lag P^s
	SMA (Q)	Cuts off after lag Q^s	Slowly decreases after lag Q^s
	SARMA (P,Q)	Slowly decreases after lag $(P,Q)^s$	Slowly decreases after lag $(P,Q)^s$
	SARIMA (P,D,Q)	Slowly decreases after lag $(P,Q)^s$ after *differencing*	Slowly decreases after lag $(P,Q)^s$ after *differencing*

This table outlines the behavior of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) for different seasonal and non-seasonal models, which can guide model selection in time series analysis.

The models obtained from the ACF and PACF plots undergo parameter significance testing and residual testing. The Q-Ljung Box test is used for testing independence, while the Kolmogorov-Smirnov test is applied for normality testing. The lowest AIC value is used to evaluate and determine the best model, and SMAPE is applied to measure the model's accuracy on the data [9].

Table 2. Forecast Accuracy

SMAPE (%)	Forecast Accuracy
SMAPE ≤ 10%	Very good forecasting accuracy
10% < SMAPE ≤ 20%	Good forecasting accuracy
20% < SMAPE ≤ 50%	Moderate forecasting accuracy
SMAPE > 50%	Poor forecasting accuracy

This table categorizes the forecasting accuracy based on SMAPE (Symmetric Mean Absolute Percentage Error) values, helping to evaluate the model's predictive quality. The forecast results obtained from the model will be used to identify the intervention response pattern in the next stages [10].

2.2 Intervention Analysis

The method used to process time series data and explain the impact of an intervention influenced by external or internal factors, which cause a change in the data pattern at a specific time ttt, is known as intervention analysis [11]. The intervention model equation is as follows:

$$Z_t = \frac{\omega_s(B)B^b}{\delta_r(B)} I_t + \frac{\theta_q(B)}{\phi_p(B)(1-B)^d} a_t$$

where:

Z_t : response variable at time t

I_t : intervention variable

b : time delay before the intervention I starts affecting Z

$\omega_s = \omega_0 - \omega_1 B - \cdots - \omega_s B^s$, where s indicates the duration of the intervention's impact on the data after b periods

$\delta_r = 1 - \delta_r B - \delta_r B^r$, where r represents the pattern of the intervention effect after b+s periods from the time of intervention at T.

$\frac{\theta_q(B)}{\phi_p(B)(1-B)^d} a_t$ ARIMA model without the effect of the intervention.

The intervention model has two functions [12]:

1. Step Function; The step function represents an intervention event occurring at time T with a long-term impact.

$$S_t^{(T)} = \begin{cases} 0, t < T \\ 1, t \geq T \end{cases}$$

2. Pulse Function; The pulse function represents an intervention event occurring at time TTT with a short-term impact

$$P_t^{(T)} = \begin{cases} 1, t = T \\ 0, t \neq T \end{cases}$$

According to [11] there are several response patterns following an intervention:

a. If the intervention effect gradually builds up and continues with a tendency to increase, the model is as follows:

$$f(Z_t) = \frac{\omega B^b}{1 - \delta B} S_t^{(T)}$$

b. If the intervention effect is felt only at a single point in time, the model is as follows:

$$f(Z_t) = \omega B^b P_t^{(T)}$$

c. If the intervention effect is felt immediately and continues with a constant effect, the model is as follows: $f(Z_t) = \omega B^b S_t^{(T)}$

d. If the intervention effect is felt immediately and continues with a tendency to decrease over time, the model is as follows:

$$f(Z_t) = \frac{\omega B^b}{1 - \delta B} P_t^{(T)}$$

According to [5], the primary aspect of intervention analysis is determining the orders b, s, and r by observing the intervention response residual plot. Order b is the start time of the intervention effect. Order s is the delay duration of an intervention after b periods, and order r appears after b and s periods. When determining order r, it is essential to observe the pattern in the intervention response residual plot. If the plot shows an exponential pattern, then r=1 ; otherwise, if there is no pattern, then r=0. Once the orders b, s, and r are determined, trial and error testing are conducted to develop potential models.

The model developed based on the intervention response residual plot undergoes parameter significance testing and residual testing. The lowest AIC value is used to evaluate and determine the best model, while SMAPE is applied to measure the model’s accuracy on the data. The best model that passes these tests can then be used to forecast future periods [13].

III. RESEARCH METHOD

3.1. Type and Source of Data

The secondary data used in this research is monthly Consumer Price Index (CPI) data for Indonesia, covering the period from January 2014 to January 2023, obtained from the official website www.bps.go.id.

3.2. Research Variables

This time series dataset consists of 109 data points, divided into two segments: 72 data points before the intervention (January 2014 – December 2019) and 37 data points after the intervention (January 2020 – January 2023).

3.3. Data Analysis Steps

(1) Data Collection: Collect monthly CPI data for Indonesia from January 2014 to January 2023.

(2) Segmentation of CPI Data: The data is divided into two parts:
- Pre-intervention data: 72 data points from January 2014 to December 2019, analyzed using the Box-Jenkins method.
- Post-intervention data: 37 data points from January 2020 to January 2023, analyzed using the intervention model [15].

(3) Forecasting Pre-intervention Data Using the Box-Jenkins Method: Conduct stationarity tests (for variance and mean); Identify the model using the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots; Specify the model, test model parameter significance, and check residual assumptions (independence and normality); Evaluate the model using the Akaike Information Criterion (AIC) and Symmetric Mean Absolute Percentage Error (SMAPE); Forecast the pre-intervention data [14].

(4) Residual Pattern Identification for Intervention Response: Use the forecast results from the pre-intervention model to identify the residual pattern for the intervention response in the subsequent steps.

(5) Forecasting Post-intervention Data Using the Step Function Intervention Model: Calculate the intervention response residuals; Develop the intervention model; Estimate parameters and test parameter significance; Perform residual assumption checks for the intervention model (independence and normality); Evaluate the model using AIC and SMAPE; then forecast the post-intervention data.

IV. RESULTS AND DISCUSSIONS

The Indonesia CPI dataset consists of 109 data points, spanning a period of 9 years and 1 month. The data reveals the following key statistics:

- Lowest CPI Value: 104.33, observed in January 2020.
- Highest CPI Value: 139.07, observed in December 2019.
- Average CPI Value: 120.10.
- Variance: 128.70.
- Standard Deviation: 11.34.

These statistics provide an overview of the variation in Indonesia’s CPI over this period, indicating a relatively high level of stability with only moderate fluctuations around the mean. The peak and trough within the dataset reflect significant economic events that may have influenced the CPI, such as the pre-pandemic conditions in December 2019 and the onset of the pandemic in January 2020.

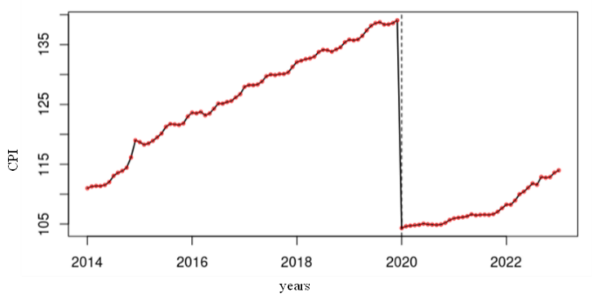

Figure 1. Indonesia CPI Plot

Based on Figure 1, the following conclusions can be drawn: The data exhibits an upward trend from January 2014 (t=1) with a CPI value of 110.99, reaching 139.07 by December 2019 (t=72). However, in January 2020 (t=73), there is an extremely sharp decline to 104.33. This sudden change in trend was heavily influenced by the intervention of the Covid-19 pandemic, which impacted Indonesia starting in early 2020.

The first step in modeling the pre-intervention data involves stationarity testing for both variance and mean. A Box-Cox plot was used to assess variance stationarity, yielding a lambda (λ) value of 3.00. Thus, a transformation was applied using $T(Z_t) = \frac{Z_t^{\lambda} - 1}{\lambda}$ to address non-stationarity in the data. Subsequently, differencing was performed using the ADF test with an order of d=1.

Following this, model identification was carried out by observing the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots to determine potential SARIMA model structures.

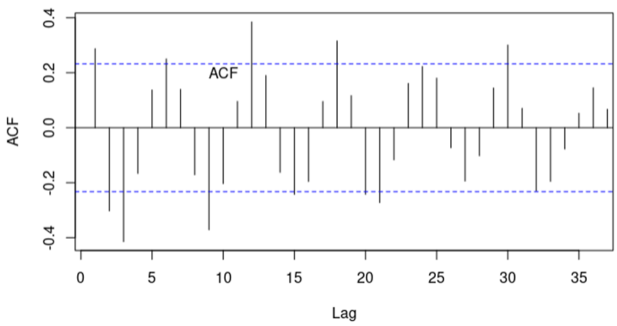

Figure 2. ACF Plot of Differenced Data (Order 1)

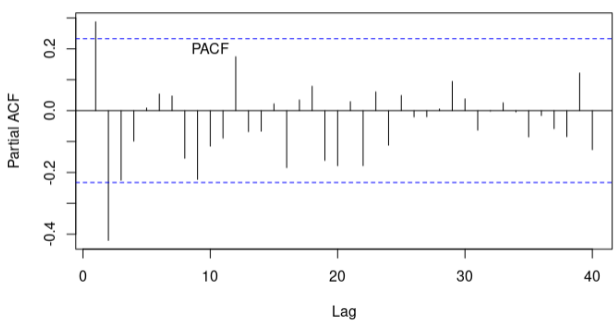

Figure 3. PACF Plot of Differenced Data (Order 1)

Based on Figure 2, it is shown that the lags cut off at the significance line in the ACF plot are lag 1, lag 2, and lag 3, suggesting potential models of MA(1), MA(2), and MA(3) with a seasonal pattern occurring at lag 12. In Figure 3, the PACF plot shows cutoffs at lag 1 and lag 2, indicating possible models of AR(1) and AR(2). The resulting model are:

SARIMA(1,1,0)(0,0,1)^{12}, SARIMA(2,1,0)(0,0,1)^{12}, SARIMA(0,1,1)(0,0,1)^{12}, SARIMA(0,1,2)(0,0,1)^{12}, SARIMA(0,1,3)(0,0,1)^{12}, SARIMA(1,1,1)(0,0,1)^{12}, SARIMA(1,1,2)(0,0,1)^{12}, SARIMA(1,1,3)(0,0,1)^{12}, SARIMA(2,1,1)(0,0,1)^{12}, SARIMA(2,1,2)(0,0,1)^{12}, and SARIMA(2,1,3)(0,0,1)^{12}.

Significance testing of parameters and residual testing are used in the subsequent tests based on the resulting model. The models SARIMA(1,1,0)(0,0,1)^{12} and SARIMA(0,1,1)(0,0,1)^{12} meet these tests. Forecasting can then be conducted on the data by evaluating the model with the smallest AIC value.

Table 3. AIC Value of SARIMA Model

Model	AIC
SARIMA(1,1,0) (0,0,1)^{12}	102,5654
SARIMA(0,1,1) (0,0,1)^{12}	101,6026

Based on Table 3, it is concluded that the SARIMA(0,1,1)(0,0,1)^{12} model has the smallest AIC value, so this model can be expressed in the following equation:

$$Z_t = Z_{t-1} + a_t - 0,5727a_{t-12} - 0,5199a_{t-1} + 0,2977a_{t-13}$$

The SARIMA(0,1,1)(0,0,1)12 model is the best model because it meets all assumption tests. Therefore, the forecasting results of the SARIMA(0,1,1)(0,0,1)^{12} model are as follows:

Table 4. SARIMA Forecasting

Period	Forecasting
February 2023	139.3138
March 2023	139.2155
April 2023	139.1966
May 2023	139.6001
June 2023	139.4166
July 2023	140.5414

Based on Table 4, it can be concluded that the forecasting results before the intervention using the SARIMA(0,1,1)(0,0,1)^{12} model show a very small monthly increase, thus this model can be used for the post-intervention data through intervention analysis.

In the SARIMA analysis before the intervention, the best model obtained was SARIMA(0,1,1)(0,0,1)^{12}, which will be used to calculate the residual response to the intervention.

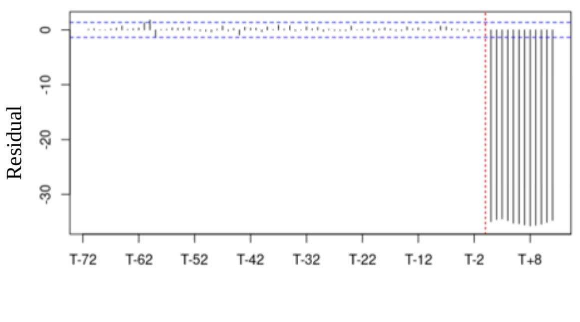

Figure 4. Residual Response Intervention

Figure 4 shows that Indonesia's Consumer Price Index (CPI) data underwent a step function intervention due to the COVID-19 pandemic. At lag 73, in January 2020 (T=73), there was an immediate decline, resulting in an order b=0, as there was no delay in the effect. The order s, representing the delay duration of the intervention's influence after b periods, is determined to be s=0. The order r=1 is chosen because the intervention response residual plot displays an exponential pattern, with fluctuations up and down. Next, a trial and error test can be performed to fine-tune the intervention order and obtain the best model.

Parameter significance test and residual test are used in the next test based on the resulting model. The model SARIMA(0,1,1)(0,0,1)^{12} of order b=0, s=0, and r=1 has fulfilled the test with an AIC value of 196.5822 so that the model can be written in the following equation:

$$Z_t = \frac{-0.0240984}{(1-0.990312878)} S_t^{(73)} + \frac{1-0.47441800(B)}{(1-B)} a_t$$

The level of goodness of a model needs to be shown by knowing the accuracy of the forecasting results. The model SARIMA(0,1,1)(0,0,1)^{12} of order b=0, s=0, and r=1 has a SMAPE value of 0.58% and is classified as a very good model accuracy for forecasting in the next period.

Table 5. SARIMA Forecasting

Period	Forecasting	Actual
February 2023	114.00	114.16
March 2023	114.02	114.36
April 2023	114.04	114.74
May 2023	114.07	114.84
June 2023	114.09	115.00
July 2023	114.11	115.24

Based on Table 5, it shows that Indonesia's CPI data experienced a very small increase every month or even tended to remain constant at 114.

V. CONCLUSION

Based on the results of the intervention analysis on Indonesia's Consumer Price Index (CPI) data, the following conclusions can be drawn:

1. The step function intervention model for Indonesia's CPI data is divided into two models: the best model before the intervention and the best model after the intervention. The best SARIMA model before the intervention is SARIMA(0,1,1)(0,0,1)^{12}, with the following equation:
$$Z_t = Z_{t-1} + a_t - 0{,}5727a_{t-12} - 0{,}5199a_{t-1} + 0{,}2977a_{t-13}$$
2. The best post-intervention model is an intervention SARIMA(0,1,1)(0,0,1)$^{\wedge 12}$ with orders b=0, s=0, and r=1, with the following equation:
$$Z_t = \frac{-0{,}0240984}{(1-0{,}99031287B)} S_t^{(73)} + \frac{1-0{,}47441800(B)}{(1-B)} a_t$$
The SMAPE value of 0.58% indicates that the model has very high accuracy.
3. Forecasting results suggest that Indonesia's CPI in 2023 is expected to increase slightly each month or remain relatively constant at around 114.

Based on these conclusions, the following recommendations can be made:

1. This study only utilized the step function intervention model on Indonesia's CPI data. It is recommended that future research apply the pulse function as well to provide a clearer comparison of the step and pulse functions in analyzing intervention effects in Indonesia.
2. Model identification for intervention through trial and error with various orders significantly impacts the forecasting results. Therefore, it is recommended to carefully select the best intervention model for more accurate outcomes.

REFERENCES

1. Singh, N., Tang Y., Zhang, Z., and Zheng, C. 2020. COVID-19 Waste Management: Effective and Successful Measures in Wuhan, China. *Journal Resources, Conservation and Recycling*, Vol. 163, No.5, 1-2.
2. The Ministry of Health Republik Indonesia. 2020. Pedoman Pencegahan Dan Pengendalian Coronavirus Disesase (Covid-19). Jakarta: The Ministry of Health Republik Indonesia
3. Badan Pusat Satistik. 2021. Pendapatan Nasional Indonesia 2016-2020. Available at https://www.bps.go.id/publication/2021/06/08/bcb06430a707226bff9f8d99/pendapatannasional-indonesia-2016-2020 (accessed at 9 July 2022).
4. Lubis, D. A., Johra, M. B., & Darmawan, G. 2017. Consumer Price Index Forecasting using Singular Spectral Analysis (SSA) Method and Seasonal Autoregressive Integrated Moving Average (SARIMA). *Jurnal Matematika*, Vol. 3, No. 2, 74 - 82.
5. Wei, W.W. S. 2006. Time Series Analysis Univariate and Multivariate Methods. America: Pearson Education, Inc.

6. Crystine, A., Hoyyi, A., & Safitri, D. 2014. Step Function Intervention Analysis (Case Study on the Number of Postal Items Delivered to Semarang in 2006-2011). *Jurnal Gaussian*, Vol. 3, No. 3, 293-302.
7. Aritara, R. 2011. Analysis of Step Function Intervention on the Increase in Basic Electricity Tariff (TDL) on the Amount of Electricity Usage. *Postgraduate Thesis*. Universitas Negeri Yogyakarta.
8. Badan Pusat Statistik. 2017. Inflasi dan Indeks Harga Konsumen. Available in https://www.bps.go.id/statictable/2009/06/15/907/indeks-harga-konsumen-dan-inflasi-bulanan-indonesia-2006-2022.html (accessed at 12 Juli 2022).
9. Estrada, M. R. C., Camarillo, M. E. G., Parraguirre, M. E. S., Castillo, M. E. G., Jua´rez, E. M., & Go´mez, M. J. C. 2020. Evaluation of Several Error Measures Applied to the Sales Forecast System of Chemicals Supply Enterprises. *International Journal of Business Administration*, Vol. 11, No. 4, 39-50.
10. Lawrence, S. M., King, A. D., Newbrander, W., Lawrence, K. D., & Klimberg, R. K. 2009. Fundamentals of forecasting using excel. *Australian Health Service Alliance* (Issue 1). Industrial Press Inc.
11. Tsay, R. S. 2010. Analysis of Financial Time Series (3rd ed.). Hoboken, New Jersey: John Wiley & Sons.
12. de Beer, J.A.A., van de Pol, F.J.R. (1988). Time Series Analysis with Intervention Effects: Method and Application. In: Saris, W.E., Gallhofer, I.N. (eds) *Sociometric Research*. Palgrave Macmillan, London.
13. Shumway, R. H. & Stoffer, D. S. 2017. Time Series Analysis and Its Applications. New York: Springer.
14. Aljandali, A. 2017. The Box-Jenkins Methodology. In: *Multivariate Methods and Forecasting with IBM® SPSS® Statistics. Statistics and Econometrics for Finance*. New York: Springer.
15. Matarise, F. 2011. *Intervention Analysis in Time Series.* In: Lovric, M. (eds) International Encyclopedia of Statistical Science. Berlin: Springer.