BAREKENG: Journal of Mathematics and Its Application
December 2022 Volume 16 Issue 4 Page 1259–1270
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol16iss4pp1259-1270

ARIMA MODEL OF OUTLIER DETECTION FOR FORECASTING CONSUMER PRICE INDEX (CPI)

M. Imron[1], Wika Dianita Utami[2*], Hani Khaulasari[3], Firman Armunanto[4]

[1,2,3]Study Program of Mathematics, Faculty of Science and Technology, UIN Sunan Ampel Surabaya
Jl. Ahmad Yani No.117, Surabaya, 60237, Indonesia
[4]Badan Pusat Statistik (BPS) Indonesia of Probolinggo
Jl. Raya Bromo No. 32 Probolinggo, Indonesia

Corresponding author's e-mail: [2]* wikadianita@uinsby.ac.id

Abstract. *The Consumer Price Index (CPI) is an indicator used by Badan Pusat Statistik (BPS) which describes the average change in the prices paid by urban consumers for a market basket of consumer goods and services in a certain period. In the case of the CPI of Probolinggo City, if the CPI increases then describe inflation occurs and conversely. The CPI of Probolinggo City increase is not fixed. This study is to forecast the Consumer Price Index (CPI) and the result can be used as one of the considerations in carrying out economic development in the future. The research focused on the data of the CPI of Probolinggo City from January 2014 to April 2022. The methodology implemented in this study is Autoregressive Integrated Moving Average (ARIMA). The result shows that ARIMA* $(0,2,1)$ *without an outlier was the best model for predicting the CPI of Probolinggo City for the next 8 months. This model shows the value of MAPE is* 1.69%*. The value of forecasting results in each month has decreased and increased not so significantly where in May 2022 the forecasting value was 108,391 then in June 2022 the forecasting value became 108,411 and so on until December 2022 the forecasting results using ARIMA model* $(0,2,1)$ *of 107,845.*

Keywords: *ARIMA, Consumer Price Index (CPI), Forecasting, Outlier Detection.*

Article info:
Submitted: 28th June 2022 *Accepted: 16th October 2022*

How to cite this article:
M. Imron, W. D. Utami, H. Khaulasari and F. Armunanto, "ARIMA MODEL OF OUTLIER DETECTION FOR FORECASTING CONSUMER PRICE INDEX (CPI)", *BAREKENG: J. Math. & App.*, vol. 16, iss. 4, pp. 1259-1270, Dec., 2022.

1. INTRODUCTION

Inflation is an important aspect of consideration for the government in deciding an economic policy. Knowing the extent of inflation that occurs can give an idea in the future of how much of a country's income is suitable for consumption, storage, or investment [1]. The Consumer Price Index (CPI) is an indicator of governments such as the Badan Pusat Statistik (BPS) in calculating inflation that occurs where data comes from changes in the mean consumer goods and services used by households or consumers in a certain period [2].

The official publication issued by the BPS of Probolinggo City stated that in April 2022, the inflation was 1.08 and the CPI was 108.06. Those conditions lead Probolinggo City to rank 4th out of 8 cities in East Java. Looking at the value of the CPI of Probolinggo City issued monthly by the BPS of Probolinggo City in 2022 from January to April continues to increase [3]. The increase in the Consumer Price Index (CPI) that occurred in Probolinggo City describes the level of inflation. If the CPI of Probolinggo City increases, then the cost of living in a household changed. Low household income with a rising CPI forces people to earn additional income to make ends meet [4]. Responding to this increase requires to analyze the value of the CPI in the future.

The CPI data can be processed and analyzed to determine the value of the CPI in the future so it is expected to be one of the considerations in planning for economic improvement in the future [5]. The time series forecasting is analyzing pre-existing CPI data from Probolinggo City to predict the future [6]. Autoregressive Integrated Moving Average (ARIMA) is a popular method for forecasting univariate time series datasets. The ARIMA model was first popularized by Box and Jenkins [7], where the ARIMA model utilizes data that has occurred in the past to obtain the latest value for the future [8]. Forecasting results can be influenced by outliers in data so that the forecasting value becomes not accurate, then analyze to detect outliers on the data to eliminate the effects of outliers [9].

The findings from previous research showed that ARIMA model in the middle of a pandemic in predicting the stock price of PT Blue Bird Tbk. in an Islamic economic perspective where the best ARIMA model obtained is ARIMA (2,1,2) with an accurate calculation of 83.33% [10]. Another previous research used ARIMA model was forecasting of rice production data from 1993 to 2020 in Aceh City and the accuracy of MSE / MAD of 3.3401×10^{10} with best ARIMA model (3,1,1) [11]. Then another research was the application of the ARIMA Model of Outlier Detection in Palopo City Water Discharge Volume Forecasting which resulted in the ARIMA model (1,1,1) with MSE before being given outlier 9579 and after being given the resulting MSE outlier decreased to 6081,050 [9]. Therefore, steps to minimize the increase in the Consumer Price Index (CPI) of Probolinggo City in the future, this study will be carried out an analysis of the value of the Consumer Price Index (CPI) using the ARIMA model. The results obtained from the best ARIMA model are expected to help economic policy determinants in taking future decision steps.

2. RESEARCH METHODS

2.1 Customer Price Index (CPI)

The Consumer Price Index (CPI) is an indicator that describes the change in the mean price of goods and services consumed by the community in a certain period [12]. The CPI will be announced by BPS to the public every month. The value of the CPI that rises illustrates the inflation rate, and the value of the CPI that falls indicates a deflationary situation [4], [12], [13]. The calculation of the CPI is using the Laspeyres equation that has been modified before. The equation of CPI is as follows [14], [12]:

$$\boldsymbol{IHK_t} = \frac{\sum_{j=1}^{k} \frac{P_{tj}}{P_{(t-1)j}} P_{(t-1)j} Q_{oj}}{\sum_{j=1}^{k} P_{oj} Q_{oj}} \tag{1}$$

where:

$\boldsymbol{IHK_t}$ = Consumer Price Index (CPI)-$\boldsymbol{t}$

$\boldsymbol{P_{tj}}$ = The price value of the goods $\boldsymbol{j}$, period $\boldsymbol{t}$

$\boldsymbol{P_{(t-1)j}}$ = The price value of the goods $\boldsymbol{j}$, period $(\boldsymbol{t-1})$
$\boldsymbol{P_{(t-1)j}Q_{oj}}$ = Consumption value of goods $\boldsymbol{j}$, in year $(\boldsymbol{t-1})$
$\boldsymbol{P_{oj}Q_{oj}}$ = Consumption value of goods $\boldsymbol{j}$, in base year
$\boldsymbol{k}$ = Commodities

2.2 Time Series Forecasting

Time series data can be interpreted as a collection of past data in a certain time series [7]. Forecasting using time series data has been applied for business, the natural environment, or the government. Time series forecasting is effective to save decision making, planning, and scheduling of a problem [15]. Time series forecasting is the analysis of existing data within a certain period of time which is then carried out a forecasting process according to existing rules with the aim of obtaining a predictive value for the future [16], [5]. One of the time series forecasting methods that has been widely applied in processing time series data to predict is ARIMA (Autoregressive Integrated Moving Average) [8].

2.3 Stationarity

In time series, stationary is data along a horizontal constant timeline which means that the increase and decrease of data do not fluctuate. In other words, fluctuating data circulates around mean and variance at each point in time [17]. Time series data is not always stationary in means and variances. Non-stationary data in variants can be performed Box-Cox transformations [14], [18]. The Box-Cox transformation equations are as follows [19], [20], [21].

$$\boldsymbol{T(X_t)' = \begin{cases} \frac{X_t^{\lambda}-1}{\lambda}, & \lambda \neq 0 \\ \ln(X_t), & \lambda = 0 \end{cases}} \tag{2}$$

where:

$\boldsymbol{T(X_t)'}$ = Transformation *Box-Cox*
$\boldsymbol{X_t}$ = Actual Data-$\boldsymbol{t}$
$\boldsymbol{\lambda}$ = Transformation Parameter

Non-stationary data in the mean can be stationaries by differencing on the data [19], [17]. Differencing equation is as follows [7].

$$\boldsymbol{X_t' = X_t - X_{t-1}} \tag{3}$$

where:

$\boldsymbol{X_t'}$ = Differencing value
$\boldsymbol{X_t}$ = Actual Data-$\boldsymbol{t}$

Differencing order $\boldsymbol{d}$ follows Equation (4) [22].

$$\boldsymbol{X_t^d = (1-B)^d X_t} \tag{4}$$

where $\boldsymbol{B}$ is backshift operator.

2.4 ARIMA (Autoregressive Integrated Moving Average)

ARIMA is a model that in its application uses time series data where the decrease in the equation from the ARIMA model can predict future values [7]. In general data with non-seasonal patterns instead of random white noise can be modeled using ARIMA [7]. The ARIMA model must be stable with both the mean and the variance at each point [23]. In the case of non-stationary data against the variance can be performed Box-Cox Transformation and if it is not stationary against the mean can be done is differencing [24]. ARIMA consists of 3 parameters $(\boldsymbol{p,d,q})$ i.e. AR represented by $\boldsymbol{p}$ where AR$(\boldsymbol{p})$ is a prediction based on past values from the data, then the MA is symbolized by $\boldsymbol{q}$ where MA$(\boldsymbol{q})$ is the residual value of the calculation used for future forecasting, and in conjunction with differencing symbolized $\boldsymbol{d}$ which is the number of differencing on the data to eliminate trends and or seasonality with the aim of disciplining the data so that the data becomes constant against the mean and variance at each point in time [25], [26]. ARIMA general model $(\boldsymbol{p}, \boldsymbol{d}, \boldsymbol{q})$ can be seen in the following structure [27]:

$$\begin{gathered}\boldsymbol{\Phi}(\boldsymbol{B})\nabla^{\boldsymbol{d}}\boldsymbol{x_t} = \boldsymbol{\Theta}(\boldsymbol{B})\boldsymbol{\varepsilon_t} \\ \boldsymbol{E}(\boldsymbol{\varepsilon_t}) = \boldsymbol{0}, \boldsymbol{Var}(\boldsymbol{\varepsilon_t}) = \boldsymbol{\sigma_\varepsilon^2}, \boldsymbol{E}(\boldsymbol{\varepsilon_t}\boldsymbol{\varepsilon_s}) = \boldsymbol{0}, \boldsymbol{s} \neq \boldsymbol{t} \\ \boldsymbol{E}(\boldsymbol{x_s}\boldsymbol{\varepsilon_t}) = \boldsymbol{0}, \forall \boldsymbol{s} > \boldsymbol{t}\end{gathered} \tag{5}$$

Where $\boldsymbol{\nabla^d = (1 - B)^d B}$; with $\boldsymbol{\Phi(B) = 1 - \varphi_1 B - \cdots - \varphi_1 B^p}$ is the coefficient of AR($\mathbf{p}$) polynomial model ARMA($\mathbf{p}, \mathbf{q}$) reversible stationary [28], [7], [27], and $\boldsymbol{\Theta(B) = 1 - \theta_1 B - \cdots - \theta_1 B^q}$ as the coefficient of MA($\mathbf{q}$) polynomial model ARMA($\mathbf{p}, \mathbf{q}$) reversible stationary [23], [29]. The formation of the alleged ARIMA model is seen from the lag in the ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) plots of the data that have been stationary with the provisions [14], [25]:

Table 1. Plot ACF and PACF

Plot	ACF	PACF
AR (p)	Dies Down	Cut off after lag p
MA (q)	Cut off after lag q	Dies Down
ARMA (p,q)	Dies Down after q-p lag	Dies Down after p-q lag

2.5 Parameter Estimation

The best model is obtained based on the test, it is necessary to estimate the parameters in the model. The parameter estimation equation for ARIMA is as follows [21].

Hypotheses used for AR:
$\boldsymbol{H_0}$: $\boldsymbol{\varphi_i = 0}$ (AR parameter is insignificant)
$\boldsymbol{H_1}$: $\boldsymbol{\varphi_i \neq 0}$ (AR parameter is significant)

$$\boldsymbol{t_{\text{statistics}\ \varphi_i}} = \frac{\boldsymbol{\varphi_i}}{\boldsymbol{SE(\varphi_i)}} \tag{6}$$

Hypotheses used for MA:
$\boldsymbol{H_0}$: $\boldsymbol{\theta_i = 0}$ (MA parameter is insignificant)
$\boldsymbol{H_1}$: $\boldsymbol{\theta_i \neq 0}$ (MA parameter is significant)

$$\boldsymbol{t_{\text{statistics}\ \theta_i}} = \frac{\boldsymbol{\theta_i}}{\boldsymbol{SE(\theta_i)}} \tag{7}$$

where:
$\boldsymbol{\varphi_i}$ = Estimation AR model
$\boldsymbol{\theta_i}$ = Estimation MA model
$\boldsymbol{SE}$ = Standard Error

$\boldsymbol{H_0}$ is rejected if $|\boldsymbol{t_{\text{statistics}}}| > \boldsymbol{t_{\frac{\alpha}{2},(n-n_z)}}$ where $\boldsymbol{n}$ is the amount of data and $\boldsymbol{n_z}$ is number of parameters estimated, or said to be significant if *p-value* $< \boldsymbol{\alpha = 0,05}$.

2.6 Residual Diagnostic Test

Residual diagnostic tests are necessary to determine the feasibility of a model because the calculated data uses the right model and meets the assumptions in the residual diagnostic test will produce values similar to the original data [17], [19]. Models with good feasibility must meet the assumptions in residual diagnostic tests, the assumption is white noise and normal distribution. The white noise assumption test can be performed using the Ljung-Box test with the following Equation [21] :

(8)

$$Q = n(n+2)\sum_{k=1}^{j}\frac{\hat{\omega}_k^2}{(n-k)}$$

where:
Q = Ljung-Box Value
$\hat{\omega}_k$ = Autocorrelation value in lag- k
k = lag- k

As for the assumption of normality to residuals that are not met, outlier detection can be carried out.

2.7 Outlier Detection

In time series analysis, outliers are inconsistent data or far from other values that can make damage the analysis so that the results caused are not reliable or even to the point of making invalid decisions [30], [31]. There are four types of outliers:

a. Additive Outlier (AO) is an outlier that has an effect only on the *th* observation. Assume $\mathbf{X_t}$ following the ARMA model $(\mathbf{p},\mathbf{q})$, and $\mathbf{Z_t}$ is observational data, then $\mathbf{X_t}$ is a free outlier.

$$\begin{aligned}\Phi(B)x_t &= \Theta(B)\varepsilon_t \\ X_t &= \frac{\Phi(B)}{\Theta(B)}\varepsilon_t\end{aligned} \tag{9}$$

Then additive outlier (AO) has formula:

$$\begin{aligned}Z_t &= \frac{\Phi(B)}{\Theta(B)}\varepsilon_t + \omega I_t^{(T)}, \varepsilon_t \sim N(0,\sigma^2) \\ I_t^{(T)} &= \begin{cases}1, & t = T \\ 0, & t \neq T\end{cases}\end{aligned} \tag{10}$$

where $I_t^{(T)}$ is an additive outlier (AO) variable T is numbers of data.

b. Innovative Outlier (IO) is an outlier that has an influence when $\mathbf{T}, \mathbf{T+1}, \cdots$ has formula:

$$Z_t = X_t + \frac{\Phi(B)}{\Theta(B)} + \omega I_t^{(T)} = \frac{\Phi(B)}{\Theta(B)}(\varepsilon_t + \omega I_t^{(T)}), \varepsilon_t \sim N(0,\sigma^2) \tag{11}$$

c. Shift Outlier (SO) has formula:

$$Z_t = X_t + \frac{1}{(1-B)}\omega I_t^{(T)}, \varepsilon_t \sim N(0,\sigma^2) \tag{12}$$

d. Temporary Change (TC) has formula:

$$\begin{aligned}Z_t &= X_t + \frac{1}{(1-\delta B)}\omega I_t^{(T)}, \varepsilon_t \sim N(0,\sigma^2) \\ I_t^{(T)} &= \begin{cases}0, & t < T \\ 1, & t \geq T\end{cases}\end{aligned} \tag{13}$$

2.8 Validation Model

The best model depends on errors. Some of the measures used in calculating errors from the results of forecasting models:

a. MAPE (Mean Absolute Percentage Error) [32], [33].

$$\boldsymbol{MAPE} = \frac{\mathbf{1}}{\boldsymbol{n}}\sum_{\boldsymbol{t}=\mathbf{1}}^{\boldsymbol{n}}\left|\frac{\boldsymbol{X_t} - \boldsymbol{F_t}}{\boldsymbol{X_t}}\right| \times \mathbf{100}\% \tag{14}$$

where:
$\boldsymbol{X_t}$ = Data actual at- $\boldsymbol{t}$
$\boldsymbol{F_t}$ = Forecasting value in at- $\boldsymbol{t}$
$\boldsymbol{n}$ = Total number of data

The MAPE results from the calculations can be explained in the category as following Table 2 [34].

Table 2. MAPE Category

MAPE	Interpretation
$< \mathbf{10}\%$	Excellent
$\mathbf{10} - \mathbf{20}\%$	Good
$\mathbf{20} - \mathbf{50}\%$	Enough
$> \mathbf{50}\%$	Bad

b. MSE (Mean Square Error) [25].

$$\boldsymbol{MSE} = \frac{\mathbf{1}}{\boldsymbol{n}}\sum_{\boldsymbol{t}=\mathbf{1}}^{\boldsymbol{n}}(\boldsymbol{X_t} - \boldsymbol{F_t})^{\mathbf{2}} \tag{15}$$

where:
$\boldsymbol{X_t}$ = Data actual at- $\boldsymbol{t}$
$\boldsymbol{F_t}$ = Forecasting value in at- $\boldsymbol{t}$
$\boldsymbol{n}$ = Total number of data

2.9 Data Sources

The type of this research is quantitative with secondary data obtained from the official website of the BPS of Probolinggo City. The data taken as the object of research is the CPI of Probolinggo City from January 2014 to April 2022 with a total of 100 data divided into two parts, there are in-sample data of 88 data and out-sample of 12 data. The following sample data obtained:

Table 3. CPI of Probolinggo City January 2014 to April 2022

Year/month	CPI
January 2014	112.23
⋮	⋮
April 2022	108.60

ARIMA forecasting with outlier detection in the process should follow the step below:

a. Plotting data.
b. Stationarity test (mean and variance stationarity).
c. Make a model conjecture by looking at the ACF and PACF plots.
d. Test the estimation of the parameters of the ARIMA model.
e. Model residual diagnostic test (white noise assumption test and normality test).
f. Checking the outlier on the data if the data is not normally distributed, conducting a parameter estimation test of the detected outlier, then adding parameters from the detected outlier on the model and re-checking residual tests on the model to which the outlier has been added.
g. Validate models without outliers and models after the addition of outliers.
h. Forecasting the CPI of Probolinggo City for the next 8 months.

3. RESULTS AND DISCUSSION

3.1. ARIMA Model

The plot of data can be identified whether there is a seasonal pattern or not. Figure 1 is a plotting of the CPI data from Probolinggo City from January 2014 to April 2021.

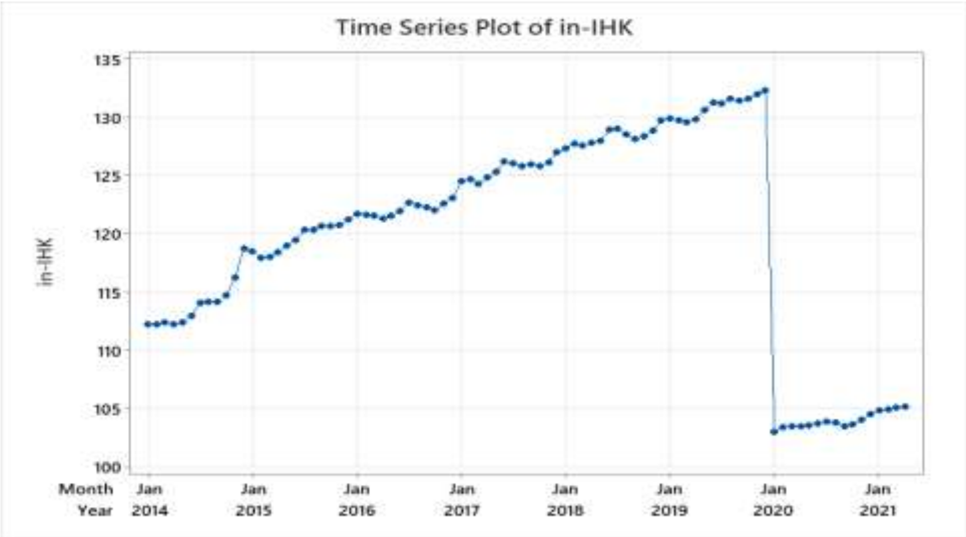
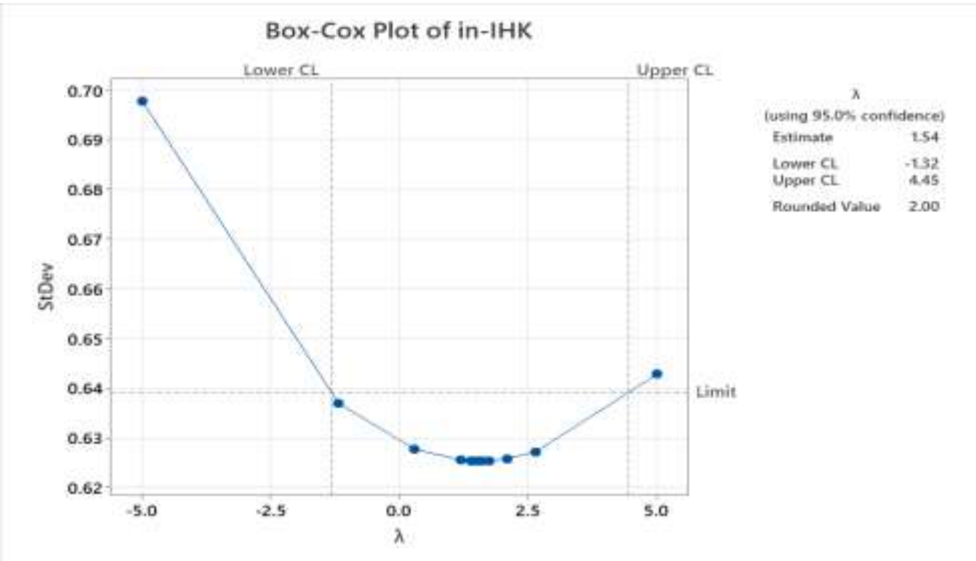
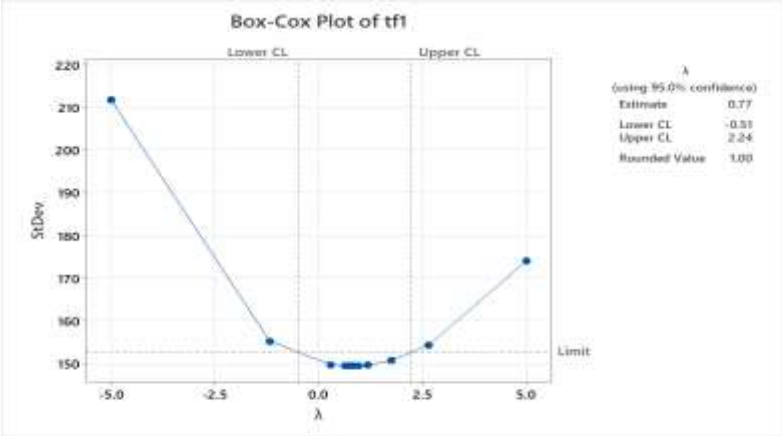


Figure 1. Plot data of the CPI

From Figure 1, there is a cyclic pattern and there is no seasonal pattern so that the data can be processed using the ARIMA model. Forecasting using ARIMA should be data used stationary in variance and mean.



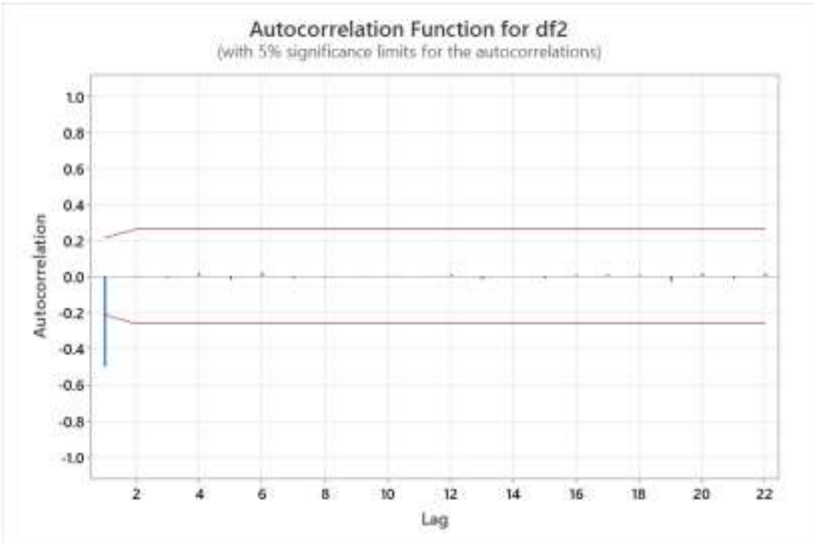
(a)



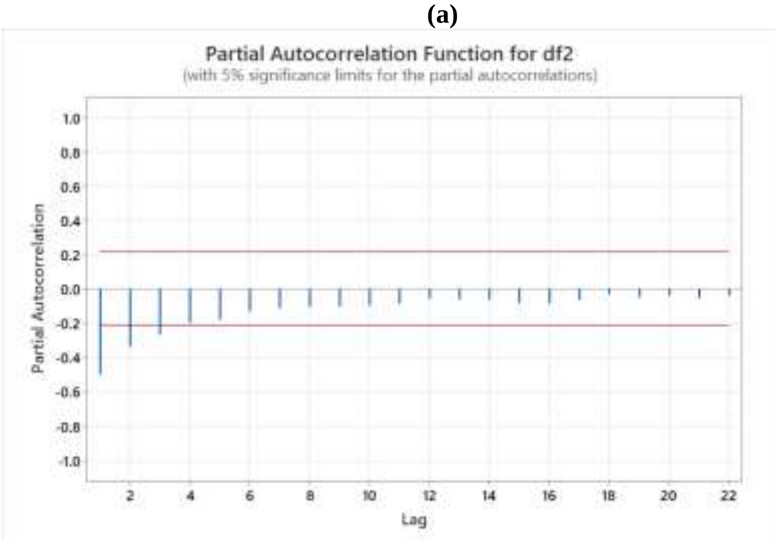
(b)

Figure 2. Plot *Box-Cox* of CPI

Based on the Box-Cox plot Figure 2 (a) the value of the CPI rounded value data is not stationary in the variant because the data is said to be stationary in variance when the value of rounded value data is equal to 1 and if it is not equal to 1 then a Box-Cox transformation is carried out until the value of the rounded value is equal to 1. In Figure 2 (b) the data has been transformed Box-Cox where it can be seen that the value of the rounded value has met the standard said to be stationary in variance, equal to 1. Then proceed with looking at the stability of the data against the mean through the ACF and PACF plots. The following is the plot of ACF and PACF data on the CPI of Probolinggo City after differencing two times.



(a)



(b)

Figure 3. Plot (a) ACF and (b) PACF after Differencing

Looking at the ACF and PACF plot lags in Figure 3, the alleged temporary model can be taken. The model combinations can be seen in Table 4.

Table 4. Combinations ARIMA Model

Model	Parameter	p-value	Description	MSE
ARIMA (1,2,0)	AR(1)	0.000	significant	16.0079
ARIMA (2,2,0)	AR(1) AR(2)	0.000 0.002	significant	14.3665
ARIMA (3,2,0)	AR(1) AR(2) AR(3)	0.000 0.000 0.015	significant	13.5210
ARIMA (0,2,1)	MA(1)	0.000	significant	10.4782
ARIMA (1,2,1)	AR(1) MA(1)	0.745 0.000	insignificant	10.5995
ARIMA (2,2,1)	AR(1) AR(2) MA(1)	0.000 0.000 0.000	significant	16.4212
ARIMA (3,2,1)	AR(1) AR(2) AR(3) MA(1)	0.000 0.092 0.487 0.000	insignificant	17.5466

Based on Table 4, it is found that only the ARIMA model (0,2,1) meets the parameter estimation test due to the value of the ARIMA model (0,2,1) having a *p-value* less than α and it can be seen that the model with the best feasibility is based on the deviation of calculations caused by the model in which the ARIMA model is selected (0,2,1) with MSE 10.4782. Further ARIMA model (0,2,1) white noise test is carried out using Ljung-Box where the decision is taken if the *p-value* is greater than α then the model is white noise and normally distributed. The Ljung-Box results are shown in Table 5.

Table 5. Residual Diagnostic Test ARIMA Model (0, 2, 1)

Lag	Chi-Square	DF	p-value White Noise	p-value Normality Test
12	1.24	10	1.00	< 0.010
24	2.14	22	1.00	
36	2.69	34	1.00	
48	2.98	46	1.00	

The results shown in Table 5 that the ARIMA model (0,2,1) was white noise because of the *p-value* $> \alpha$, but it is not normality distributed because of the *p-value* $< \alpha$ which indicates that in the data there is an outlier. It is also seen in Figure 1 that there is a fluctuating spike in decline identifying the presence of an outlier in the data. The following step is to detect outlier with the results can be seen in Table 6.

Table 6. List of Outliers on the CPI

Observation	Outlier Type	p-value
73	Shift	0.0001
12	Shift	0.0001
11	Shift	0.0004

Observation	Outlier Type	p-value
37	Shift	0.001
54	Shift	0.0165
60	Shift	0.0136
42	Shift	0.015
48	Shift	0.0159

Table 6 shows that there are 10 outliers in the CPI data of Probolinggo City where all outliers are shift outliers. Of the 10 outliers identified all parameters have a *p-value* $< \alpha$ so it can be concluded that the existing outlier affects the data. Outlier observations are then added one by one to the ARIMA model $(0,2,1)$ so that the value for the residual assumption, there is white noise, and normal distribution is shown in Table 7.

Table 7. Residual Diagnostic Test ARIMA Model $(0,2,1)$ Outlier

Lag	Chi-Square	DF	p-value White Noise	p-value Normality Test
6	447.6	5	<.0001	<0.0100
12	872.23	11	<.0001	
18	1273.05	17	<.0001	
24	1638.77	23	<.0001	

Table 7 shows the detection of 10 obeservations added to the ARIMA model $(0,2,1)$ one by one to satisty all assumptions. However, the residual model not normal distributed (p-value normality test $< \alpha$) and also does not meet white noise assumption (p-value white noise $< \alpha$). Because the outlier detection in the CPI data resulted in a value that could not meet the assumptions violated, the ARIMA model was decided $(0,2,1)$ without outliers that will be used as a forecasting model for the CPI of Probolinggo City in the future and strengthened by looking at the validation of the model, the MAPE value of the ARIMA model $(0,2,1)$ without outliers and ARIMA model $(0,2,1)$ outliers shown in Table 8.

Table 8. Validation ARIMA Model (0,2,1)

ARIMA Model	MAPE
ARIMA (0,2,1) without outliers	1.69%
ARIMA (0,2,1) with outlier	1.77%

The results in Table 8 state that ARIMA $(0,2,1)$ without outlier with the results of weighing the protection or error given by 1.69% is included in the very good category so it is worth using to predict the CPI of Probolinggo City in the future.

3.2. Forecasting the CPI of Probolinggo City

Forecasting of the CPI of Probolinggo City for the months of May 2022 to December 2022 using the ARIMA model $(0,2,1)$ without outliers are shown in Table 9.

Table 9. Forecasting the CPI of Probolinggo City May 2022 to December 2022

Year/month	Forecasting
May 2022	108.391
June 2022	108.411
July 2022	108.203
August 2022	108.223
September 2022	108.014
October 2022	108.034

Year/month	Forecasting
November 2022	107.825
December 2022	107.845

Results of the ARIMA model forecasting process $(0,2,1)$ without outliers in Table 9 to the value of the CPI of Probolinggo City for the next 8 months show a subduction and not so significant increase, whereas in May 2022 the forecasting value of 108,391, then in June 2022 the forecasting value will be 108,411, and so on until in December 2022 the forecasting results using ARIMA $(0,2,1)$ of 107,845.

The BPS of Probolinggo City stated that the increase and decrease in the value of the CPI of Probolinggo City were caused by 8 expenditure groups, there are Food, Beverages, and Tobacco; Clothing and Footwear; Housing, Water, Electricity, and Household Fuel; Household Supplies, Appliances, and Routine Maintenance; Health; Transportation; Information, Communications, and Financial Services; Recreation, Sports, and Culture; Education; Provision of Food and Beverages/Restaurants; Personal Care and Other Services.

The CPI has decreased (deflation) at first glance will benefit consumers because the prices of goods and services will become cheaper and more affordable but in regions with moderate economies, especially producers, this is not good, on the contrary, an increase in the CPI or Inflation is highly expected by producers. So, with this research in the future, it is hoped that it will provide benefits and can be used as one of the input providers in achieving a better economy.

4. CONCLUSIONS

The CPI is an important factor in economic policymaking. A significant increase in the value of the CPI led to a state of inflation. Forecasting is a way to determine the value of the CPI of Probolinggo City, the results of the prediction can be taken into consideration in making future policies. The results obtained forecasting using the ARIMA model $(0, 2, 1)$ without outliers, MAPE 1.69% with a predicted value in May 2022 of 108,391, and the following month experienced a not too significant increase and decrease.

AKNOWLEDGEMENT

The researcher's gratitude to the Badan Pusat Statistik (BPS) of Probolinggo City for giving researchers the opportunity to study and get to know more about the world of work within the BPS of Probolinggo City so that researchers can compile this journal. Thank you to the supervisors, field lecturers, and all parties who have helped in compiling this journal. Researchers hope that this journal can be useful in the future.

REFERENCES

[1] N. Hauzenberger, F. Huber, and K. Klieber, "Real-time Inflation Forecasting Using Non-linear Dimension Reduction Techniques," *Int. J. Forecast.*, no. xxxx, 2020, doi: 10.1016/j.ijforecast.2022.03.002.

[2] S. N. Afiyah and D. K. Wijaya, "Sistem Peramalan Indeks Harga Konsumen (IHK) Menggunakan Metode Double Exponential smoothing," *J. Ilm. Teknol. Inf. Asia*, vol. 12, no. 1, p. 56, 2018, doi: 10.32815/jitika.v12i1.243.

[3] B. P. S. (BPS) Kota Probolinggo, *Perkembangan Indeks Harga Konsumen Kota Probolinggo April 2022*, no. No. 06/05/3574/Th. XXII, 9 Mei 2022 Perkemb. 2022.

[4] D. S. S. Indah Lely Cristanti, Bambang Ismanto, "Pengaruh Indeks Harga Konsumen (Ihk) Dan Inflasi Terhadap Suku Bunga Tahun 2008-2018 Indonesia Indah," *J. Progr. Stud. Pendidik. Ekon. FKIP Univ. Kristen Satya Wacana*, vol. 34, no. 17, pp. 2826–2829, 2018.

[5] A. Wanto and A. P. Windarto, "Analisis Prediksi Indeks Harga Konsumen Berdasarkan Kelompok Kesehatan Dengan Menggunakan Metode Backpropagation," *J. Penelit. Tek. Inform. Sink.*, vol. 2, no. 2, pp. 37–43, 2017, [Online]. Available: https://zenodo.org/record/1009223#.Wd7norlTbhQ.

[6] Y. Wigati, R. Rais, and I. T. Utami, "Pemodelan Time Series Dengan Proses ARIMA Untuk Prediksi Indeks Harga Konsumen (Ihk) Di Palu – Sulawesi Tengah," *J. Ilm. Mat. Dan Terap.*, vol. 12, no. 2, pp. 149–159, 2017, doi:

10.22487/2540766x.2015.v12.i2.7908.
[7] J. Sun, "Forecasting COVID-19 pandemic in Alberta, Canada using modified ARIMA models," *Comput. Methods Programs Biomed. Updat.*, vol. 1, no. September, p. 100029, 2021, doi: 10.1016/j.cmpbup.2021.100029.
[8] J. Kiarie, S. Mwalili, and R. Mbogo, "Forecasting the spread of the COVID-19 pandemic in Kenya using SEIR and ARIMA Models," *Infect. Dis. Model.*, 2022, doi: 10.1016/j.idm.2022.05.001.
[9] D. Risky Arifanti and A. Asrirawan, "Peramalan Volume Debit Air Kota Palopo Menggunakan Model ARIMA Deteksi Pencilan," *Infin. J. Mat. dan Apl.*, vol. 1, no. 2, pp. 31–39, 2021, doi: 10.30605/27458326-50.
[10] A. F. AFI, *Analisis Prediksi Harga Saham Pt Blue Bird Tbk Ditengah Pandemi Covid-19 Dengan Metode ARIMA Dalam Perspektif Ekonomi Islam*. 2022.
[11] N. Nurviana, A. Amelia, R. P. S. Riezky Purnama Sari, U. N. Ulya Nabilla, and T. Talib, "Forecasting Rice Paddy Production in Aceh Using ARIMA and Exponential Smoothing Models," *Cauchy*, vol. 7, no. 2, pp. 281–292, 2022, doi: 10.18860/ca.v7i2.13701.
[12] H. S. Noor and C. Komala, "Analisis Indeks Harga Konsumen (IHK) Menurut Kelompok Pengeluaran Nasional Tahun 2018," *J. Perspekt.*, vol. 3, no. 2, p. 110, 2019, doi: 10.15575/jp.v3i2.48.
[13] Y. Yolanda, "Indeks Harga Konsumen Kota Pekanbaru Berdasarkan Pendekatan ARIMA dan Error Correction Model," 2020.
[14] F. Eugeni, N. Nainggolan, and Y. A. R. Langi, "d ' CartesiaN Penerapan Metode Autoregressive Integrated Moving Average (ARIMA) dalam Model Intervensi Fungsi Step terhadap Indeks Harga Konsumen di Kota Manado," 2020.
[15] F. Martínez, F. Charte, M. P. Frías, and A. M. Martínez-Rodríguez, "Strategies for time series forecasting with generalized regression neural networks," *Neurocomputing*, vol. 491, pp. 509–521, 2022, doi: 10.1016/j.neucom.2021.12.028.
[16] A. Keramik, "Peramalan Penjualan Keramik Menggunakan Metode Moving Mean Dan Exponential Smoothing Pada Usaha Agus Keramik Kusyanto 1 , Dadang Suhardi 2 , Robi Awaluddin 3."
[17] A. Machmudin and B. S. S. Ulama, "Peramalan Temperatur Udara di Kota Surabaya dengan Menggunakan ARIMA dan Artificial Neural Network," *J. Sains dan Seni ITS*, vol. 1, no. 1, pp. 118–123, 2012.
[18] O. Trisnawati and A. A. Box-jenkins, "Peramalan Curah Hujan di Stasiun Juanda Menggunakan Metode ARIMA Box-Jenkins dan Radial Basis Function Neural Network," vol. 11, no. 2, 2021.
[19] D. P. Anggraeni, D. Rosadi, H. Hermansah, and A. A. Rizal, "Prediksi Harga Emas Dunia di Masa Pandemi Covid-19 Menggunakan Model ARIMA," *J. Apl. Stat. Komputasi Stat.*, vol. 12, no. 1, p. 71, 2020, doi: 10.34123/jurnalasks.v12i1.264.
[20] A. Fructuoso da Costa and A. Fernando Crepaldi, "The bias in reversing the Box-Cox transformation in time series forecasting: An empirical study based on neural networks," *Neurocomputing*, vol. 136, pp. 281–288, 2014, doi: 10.1016/j.neucom.2014.01.004.
[21] R. Prameswari, *Peramalan Ekspor Kendaraan Bermotor Di Indonesia Menggunakan Metode ARIMA (Autoregressive Integrated Moving Average) Box-Jenkins*. 2020.
[22] N. AZIZAH, *Peramalan Jumlan Bencana Bnajir Di Indonesia Menggunakan Metode Seasonal Autoregressive Integrated Moving Average (SARIMA)*. 2022.
[23] K. E. ArunKumar, D. V. Kalaga, C. Mohan Sai Kumar, M. Kawaji, and T. M. Brenza, "Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving mean (ARIMA), seasonal autoregressive Integrated moving mean (SARIMA) for forecasting COVID-19 trends," *Alexandria Eng. J.*, vol. 61, no. 10, pp. 7585–7603, 2022, doi: 10.1016/j.aej.2022.01.011.
[24] H. Panjaitan, A. Prahutama, and S. Sudarno, "Peramalan Jumlah Penumpang Kereta Api Menggunakan Metode Arima, Intervensi Dan Arfima (Studi Kasus : Penumpang Kereta Api Kelas Lokal EkonomiDAOP IV Semarang)," *J. Gaussian*, vol. 7, no. 1, pp. 96–109, 2018, doi: 10.14710/j.gauss.v7i1.26639.
[25] A. Swaraj, K. Verma, A. Kaur, G. Singh, A. Kumar, and L. Melo de Sales, "Implementation of stacking based ARIMA model for prediction of Covid-19 cases in India," *J. Biomed. Inform.*, vol. 121, no. August, p. 103887, 2021, doi: 10.1016/j.jbi.2021.103887.
[26] H. Alabdulrazzaq, M. N. Alenezi, Y. Rawajfih, B. A. Alghannam, A. A. Al-Hassan, and F. S. Al-Anzi, "On the accuracy of ARIMA based prediction of COVID-19 spread," *Results Phys.*, vol. 27, p. 104509, 2021, doi: 10.1016/j.rinp.2021.104509.
[27] X. Wang, "Research on the prediction of per capita coal consumption based on the ARIMA–BP combined model," *Energy Reports*, vol. 8, pp. 285–294, 2022, doi: 10.1016/j.egyr.2022.01.131.
[28] K. Gencer and F. Başçiftçi, "Time series forecast modeling of vulnerabilities in the android operating system using ARIMA and deep learning methods," *Sustain. Comput. Informatics Syst.*, vol. 30, no. September 2020, 2021, doi: 10.1016/j.suscom.2021.100515.
[29] D. Yonggang, W. Huan, W. Mingqiang, T. Linjiang, and Y. Tao, "Application of ARIMA-RTS optimal smoothing algorithm in gas well production prediction," *Petroleum*, no. 2020, 2022, doi: 10.1016/j.petlm.2021.09.001.
[30] R. Agustianto, I. Purnamasari, and Suyitno, "Analisis Data Ketinggian Permukaan Air Sungai Mahakam Daerah Kutai Kartanegara Tahun 2010-2016 Menggunakan Model Autoregressive Integrated Moving Average (ARIMA) Dengan Efek *Outlier* (Studi Kasus : Data Mean Ketinggian Tiap Bulan Permukaan Air Sungai ," *J. EKSPONENSIAL*, vol. 11, no. 1, pp. 39–46, 2020.
[31] Mutia Kanza, "Peramalan Harga Saham PTPP Pada Jakarta Islamic Index Menggunakan Model ARIMA Dengan Deteksi *Outlier*," 2015.
[32] M. Matyjaszek, P. Riesgo Fernández, A. Krzemień, K. Wodarski, and G. Fidalgo Valverde, "Forecasting coking coal prices by means of ARIMA models and neural networks, considering the transgenic time series theory," *Resour. Policy*, vol. 61, no. February, pp. 283–292, 2019, doi: 10.1016/j.resourpol.2019.02.017.
[33] D. A. Lubis, M. B. Johra, and G. Darmawan, "Peramalan Indeks Harga Konsumen dengan Metode Singular Spectral Analysis (SSA) dan Seasonal Autoregressive Integrated Moving Average (SARIMA)," *J. Mat. "MANTIK,"* vol. 3, no. 2, pp. 74–82, 2017, doi: 10.15642/mantik.2017.3.2.74-82.
[34] D. Sulistiyawati, "Analysis of Population Prediction in Pasuruan City Using ARIMA," vol. 15, no. 3, pp. 525–534, 2021.