Inflation modeling in Indonesia using hybrid autoregressive integrated moving average (ARIMA)-neural network (NN)

D Pratiwi and M Hadijati
Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Mataram, Mataram, Indonesia

*Email: denipratiwi26@gmail.com

Abstract. One of the macroeconomic indicators to see the stability of a country's economy is inflation. This study aims to model the value of monthly inflation in Indonesia from January 2003 to December 2019 using the ARIMA-NN hybrid. The data plot shows a non-linear pattern and trends, so that the differencing process is carried out and the model is built using ARIMA model. The best ARIMA model obtained is SARIMA (1,1,0)(0,1,1)\textsubscript{12} with a Root Mean Square Error (RMSE) of 0.01134. Furthermore, ARIMA residuals that do not satisfy white noise and normality are modeled using NN. The best structure obtained of NN model is (3×2×1) with an RMSE of 0.023984. From the ARIMA and residual NN prediction results, the ARIMA-NN hybrid model is obtained to predict the value of monthly inflation in Indonesia for the next 12 months with the Mean Absolute Percentage Error (MAPE) value is 11.40873%. It means that the model result has high prediction accuracy.

1. Introduction

Inflation can briefly be interpreted as a tendency to increase the prices of goods and services generally and continuously [1]. Based on BPS [2], the month-to-month inflation rate in Indonesia on June 2019 was 0.55 percent with 138.16 of Consumer Price Index (CPI). The highest inflation occurred in Manado at 3.60 percent and the lowest occurred in Singaraja at 0.02 percent. Meanwhile, on August 2019, there was inflation of 0.12 percent. The highest inflation occurred in Kudus with 0.82 percent and the lowest occurred in Tasikmalaya, Madiun, and Pare-Pare with 0.04 percent in each. It can be seen that there has been an increase and decrease in the value of inflation in Indonesia, therefore inflation modeling is necessary. One of the statistical methods that can be used is time series. Time series is a series of observations collected based on the same time interval. One of the time series methods that can be used is Autoregressive Integrated Moving Average (ARIMA). ARIMA is able to model linear pattern. However, in reality the data does not only contain linear elements but still contains non-linear elements [3-5]. To overcome the limitations of ARIMA, the Neural Network (NN) method can be used. The NN method can model data without considering the linearity pattern of the data [6]. Therefore, the ARIMA-NN hybrid method is obtained which will be used to model monthly inflation data in Indonesia for the period January 2003 to December 2019.

Research on the ARIMA-NN hybrid model has been carried out by Zhang [6], who applied data on wolf’s sunspot, Canadian lynx and dollar exchange rates. From that research, it is found that the ARIMA-NN hybrid model produced can be used as an effective way to improve forecasting performance. In the ARIMA-NN hybrid modeling the data is modeled using ARIMA, while the residuals of the ARIMA model are modeled using NN. To model the data with the hybrid ARIMA-
NN, initial testing is carried out using Tersvirta test to determine whether there was a non-linear pattern in the data. The accuracy of forecasting is measured based on the Mean Absolute Percentage Error (MAPE) value [7-8].

The paper is organized as follows. In the next section, it is explained the data set and methodology of research. Empirical results from the real data sets are reported in Section 3. Section 4 contains the concluding remarks.

2. Methodology

Control chart structure and log-normal transformation for CV statistic. The data in this study is secondary data obtained from the official website of Bank Indonesia https://bi.go.id which is the monthly data on inflation from January 2003 to December 2019. Data is divided to in sample data (from January 2003 to December 2018) and out sample data (January 2019 to December 2019).

The steps for modeling hybrid ARIMA-NN are as follows.

1. Collecting inflation data in Indonesia, then dividing the data into in sample data for modeling or training and out sample data to measure the accuracy of forecasting.
2. Testing the linearity data using Tersvirta test.
3. Testing the stationary of the data, then identify the ARIMA model against the stationary data and estimating the ARIMA parameters using Ordinary Least Square (OLS).
4. Testing the best ARIMA model residuals assumptions, including normality and white noise residuals.
5. Next, modeling ARIMA residuals using NN. Determining the NN input combination and conduct trial and error the NN parameters used for training.
6. Calculate the value of out sample RMSE from the trial error results of the NN training. The NN structure with the smallest out sample RMSE value will be used for the NN testing step.
7. Perform the ARIMA residuals testing step and then forecasting using hybrid ARIMA-NN.

3. Result and Discussions

Hybrid is a combination of two or more systems in one function, in this case it is a combinations of ARIMA and NN. In general, the combinations of linear and non linear time series models can be written as follows

\[ y_t = L_t + E_t, \]

where \( L_t \) denotes a linear component and \( E_t \) denotes a non linear component. There are two components that must be estimated from the data, thus are the ARIMA model is used to solve linear cases and the residuals of the linear model still containing non linear relationship information.

3.1 Results of the Terasvirta linearity test

To perform the Terasvirta test, it was carried out by regressing \( X_t \) with the predictor variables \( X_{t-1}, X_{t-2}, X_{t-3} \). Furthermore, from the regression results obtained residual (\( \hat{u}_t \)), then regressed with the initial predictor variable, 6 additional predictor variables, and the quadratic and cubic terms of the initial predictor variables, with a significance level of \( \alpha \) 5%. The result is obtained \( p \)-value 0.000. So that it can be concluded to reject \( H_0 \) and it means that \( x_t \) has a non linear pattern.

3.2. ARIMA Modeling

In an autoregressive integrated moving average model, the future value of a variable is assumed to be a linear function of several past observations and random errors. According to Wei [9], the general form of the AR model with the \( p \)-order or the ARIMA (p,0,0) is stated as follows

\[ Z_t = \mu + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} + e_t. \]
The general form of the q-order MA model or ARIMA (0,0,q) can be written as follows

$$Z_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q}.$$ 

Therefore, the ARIMA Box-Jenkins or ARIMA (p,d,q) model is obtained as follows

$$\phi_p(B)(1 - B)^d X_t = \theta_q(B).$$

where $X_t$ and $\epsilon_t$ are the actual value and random error at time period $t$, respectively, $\phi_p$ and $\theta_q$ are model parameters. The symbols p and q are integers and often referred to as orders of the model. Random errors ($\epsilon_t$) are assumed to be independently and identically distributed with a mean of zero and a constant variance of $\sigma^2$. According to Aguilar et al. [10], the ARIMA model which contains seasonal patterns is called Seasonal ARIMA (SARIMA). In general, it can be written as the following equation

$$\varphi_p(L)\varphi_p(B^d)\nabla^d\nabla^s X_t = \theta_q(B^s)a_t.$$ 

The SARIMA model can be used on data that shows a recurring pattern with a certain period and is a combination of non-seasonal factors and seasonal factors. Figure 1 shows time series and Box cox plot $X_t$.

![Figure 1. Time series and Box-cox plot $X_t$](image)

Figure 1 shows the data is not stationary in the mean because it has a trend pattern and Box-cox plot which shows a rounded value of 0 so that the data is not stationary in variance, so it is necessary to transform $\ln(X_t)$ and differencing 1st and 12th orders because the data forms a seasonal pattern every 12 months and then was obtained time series plot in Figure 2.

![Figure 2. Time series plot stationery data](image)
Figure 2 shows stationary time series plot of $X_t$. And then, identify the ARIMA model using ACF and PACF of the stationery data. The result is shown in Figure 3.

![Figure 3. ACF and PACF stationery data](image)

Based on the ACF and PACF in Figure 3, it can be seen that the seasonal pattern of ACF is cut off at lag 12, and PACF dies down, so that in the seasonal pattern it can be concluded that it contains ARIMA $(0,1,1)^{12}$. The ACF regular pattern is cut off pattern at lag 1 and 7, while PACF is cut off at lag 1. In the selection of the ARIMA model order can be done with the parsimony principle, meaning that selection a model with as few parameters as possible. And from all of the possible combinations of the regular and seasonal ARIMA model, it can be found the best model that is SARIMA $(1,1,0)(0,1,1)^{12}$ with RMSE 0.01134. After that the residual diagnostic checking is carried out using Kolomogorov-Smirnov test for testing normality residual and Ljung-Box test for testing white noise of residual ARIMA. With 5 percent significance level, it is concluded that the residuals do not satisfy normality and white noise residuals so the residual pattern still affected the SARIMA$(1,1,0)(0,1,1)^{12}$ models, therefore the residuals can still be modeled using NN and is expected to increase the level of accuracy of the prediction results.

3.3 Neural Network (NN) modeling

Neural Network (NN) is a connected network that mimics the working of nerve cells in the human brain. NN are one of such models that are able to approximate various nonlinearities in the data. NN are flexible computing frameworks for modeling a broad range of nonlinear problems. One significant advantage of the NN models over other classes of nonlinear model is that NN are universal approximators which can approximate a large class of functions with a high degree of accuracy. Relationship of output ($y_t$) with input ($y_{t-1}, ..., y_{t-p}$) can be expressed as [6]

$$y_t = w_0 + \sum_{j=1}^{q} w_j g \left( w_{0j} + \sum_{i=1}^{p} w_{ij} y_{t-i} \right) + e_t,$$

where $w_j (j = 1,2, ..., q)$ and $w_{ij} (i = 0,1,2, ..., p; j = 1,2, ..., q)$ are the model parameters often called the connection weights; $p$ is the number of input nodes and $q$ is the number of hidden nodes. In this study, a combination of $e_{t-12}$, $X_t$, and $X_t$ which were stationary both in variances and mean. Thus, the number of input neurons used in this research is 3 neurons with the parameters and structure of the NN
used for training determined by trial error. Figure 4 shows the NN structure that produces the smallest out sample RMSE 0.023984.

Figure 4. Structure NN (3×2×1) for training

Then performed ARIMA residual testing using NN with best structure NN in Figure 4. The testing result shows that the residual prediction by the best structure NN obtained is not much different with the actual ARIMA residual. So that it can be continued with the prediction of inflation based on the ARIMA-NN hybrid model by adding the ARIMA prediction of inflation and the residual NN prediction. The result of data prediction of inflation based on the ARIMA-NN hybrid model is shown in Table 1.

Table 1. Comparison table of actual and prediction data using hybrid ARIMA-NN

| Period    | Actual | Prediction | Residual |
|-----------|--------|------------|----------|
| Januari   | 0.0282 | 0.0332     | -0.0050  |
| Februari  | 0.0257 | 0.0325     | -0.0068  |
| Maret     | 0.0248 | 0.0322     | -0.0074  |
| April     | 0.0283 | 0.0347     | -0.0064  |
| Mei       | 0.0332 | 0.0353     | -0.0021  |
| Juni      | 0.0328 | 0.0346     | -0.0018  |
| Juli      | 0.0332 | 0.0333     | -0.0001  |
| Agustus   | 0.0349 | 0.0332     | 0.0017   |
| September | 0.0339 | 0.0331     | 0.0008   |
| Oktober   | 0.0313 | 0.0309     | 0.0004   |
| November  | 0.0300 | 0.0303     | -0.0003  |
| Desember  | 0.0272 | 0.0322     | -0.0050  |

Figure 5 shows comparison graph of actual and prediction data of inflation rate in Indonesia period Januari 2019 to December 2019 by using hybrid ARIMA-NN.
Based on Figure 5, it can be seen that the dashed red line shows the prediction data, while the blue line is the actual data. The graph shows that the prediction data pattern is close to the actual data pattern. This is supported by the result of calculation of MAPE value of 11.40873% that it means the prediction of inflation rate using ARIMA-NN hybrid model is relatively good with high prediction accuracy.

4. Conclusion

Based on the result and discussion, it can be concluded that:

a. Monthly inflation data Indonesia from January 2003 to December 2019 still contain a non-linear pattern so it is suitable to be modeled using hybrid ARIMA-NN method. In addition, inflation data in Indonesia also has seasonal pattern every 12 periods, so that the best model is obtained using ARIMA method, SARIMA \((1,1,0)(0,1,1)_{12}\) with RMSE value 0.01134 while residuals SARIMA \((1,1,0)(0,1,1)_{12}\) which still contains non-linear patterns modeled using NN obtained the best structure, namely NN \((3\times2\times1)\) with RMSE of 0.023984.

b. From the ARIMA-NN hybrid model that has been obtained, it can be seen that the accuracy of the forecasting value based on the MAPE value is 11.408730%, meaning that the resulting model has high prediction accuracy.

Acknowledgments

Special thanks to Statistical Research Group (KPBI Statistika) of Universitas Mataram for their support.

References

[1] Suseno, Astiyah S 2009 *Seri Kebanksentralan* Retrieved from Bank Indonesia: http://www.bi.go.id
[2] Badan Pusat Statistik (BPS) 2019 *Berita Resmi Statistik* NTB: Badan Pusat Statistik
[3] Bakar N A, Rosbi S 2017 Autoregressive Integrated Moving Average (ARIMA) Model for Forecasting Cryptocurrency Exchange Rate in High Volatility Environment: A New Insight of
Bitcoin Transaction *International Journal of Advanced Engineering Research and Science* **4** (11) 130-137

[4] Hatta H H M, Daud F M, Mohamad N 2018 An Application of Time Series ARIMA Forecasting Model for Predicting the Ringgit Malaysia-Dollar Exchange Rate *Journal of Data Analysis* **1** (1) 42-48

[5] Unggara I, Musdholifah A, Sari A K 2019 Optimization of ARIMA Forecasting Model using Firefly Algorithm *Indonesian Journal of Computing and Cybernetics Systems* **13** (2) 127-136

[6] Zhang G P 2003 Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. *Neurocomputing* **50** 159-175

[7] Estiko F, Wahyuddin S 2019 Analysis of Indonesia’s Inflation Using ARIMA and Artificial Neural Network *Economics Development Analysis Journal* **8** (2) 151-162

[8] Qadrini L, Asrirawan, Mahmudah N, Fahmuddin M S, Amri I F 2021 Forecasting Bank Indonesia Currency Inflow and Outflow Using ARIMA, Time Series Regression (TSR), ARIMAX, and NN Approaches in Lampung *Jurnal Matematika, Statistika & Komputasi* **17** (2) 166-177

[9] Wei W S 2006 *Time Series Analysis: Univariate and Multivariate Methods* (2nd Edition) Boston: Addison Wesley

[10] Aguilar J J, Turias I J, Come M J 2014 Hybrid Approaches Based on SARIMA and Artificial Neural Networks for Inspection Time Series Forecasting *Transportation Research Part E* **67** 1-13