EJEM
Econ. J. Emerg. Mark.

Economic Journal of Emerging Markets
Available at http://journal.uii.ac.id/index.php/jep

Predicting unemployment rates in Indonesia

Umi Mahmudah

Universiti Malaysia Terengganu, Malaysia.
E-mail: u_mudah@yahoo.com

Article Info

Article history:
Received : 12 December 2016
Accepted : 14 March 2017
Published : 1 April 2017

Keywords:
forecasting, unemployment rate, ARIMA

JEL Classification:
E24, C22, C53

DOI:
10.20885/ejem.vol9.iss1.art3

Abstract

The main purpose of this study is to predict the unemployment rate in Indonesia by using time series data from 1986 to 2015 using autoregressive integrated moving average (ARIMA). A differencing process is required due to the actual time series of the unemployment rates in Indonesia is non-stationary. The results show that the best model for forecasting the unemployment rate in Indonesia by using the ARIMA (0,2,1) model. The forecasting results reveal that the unemployment rate in Indonesia tends to decrease continuously. The average of the residuals is close to zero which informs a good result of the forecasting analysis.

Abstrak

Tujuan utama dari penelitian ini adalah untuk memprediksi tingkat pengangguran di Indonesia dengan menggunakan data time series selama periode 1986 sampai dengan 2015. Metode analisis yang digunakan adalah autoregressive integrated moving average atau ARIMA. Oleh karena data tingkat pengangguran adalah tidak stastioner maka data tersebut perlu didiferensiasi. Hasil penelitian menunjukkan bahwa model terbaik yang dapat digunakan untuk meramalkan tingkat pengangguran di Indonesia adalah dengan menggunakan model ARIMA (0,2,1). Hasil peramalan mengungkapkan bahwa tingkat pengangguran di Indonesia akan cenderung menurun secara terus menerus. Sementara itu dari hasil rata-rata dari residual mendekati nilai nol dimana hal ini mengindikasikan bahwa analisis peramalan memberikan hasil yang baik.

Introduction

Indonesia as a developing country has the complexity of problems that requires an appropriate policy to minimize its impact. One of the most challenging problems is unemployment rate which influences the country's economic rate of growth. It comes as no surprise that a higher rate of unemployment indicates economic instability as well as an obstacle to reach the expected developments. Significant impacts that can be experienced due to the high unemployment rate among others are as follows. Unemployment may reduce not only potential earnings of individuals but also the feeling of the perceived prosperity. This perhaps forces shrinkage of national revenue. Other than that, this is able to boost the poverty rate. According to International Labor Organization (ILO) definition, unemployed person is someone who does not work but repeatedly looking for a job. Statistics Indonesia (Badan Pusat Statistik/BPS) divides the open unemployment into four criteria; those who do not have jobs but looking for one, those who do not have jobs but preparing business enterprises, those who do not have jobs and neither looking for jobs because they feel impossible to get one, those who have jobs but yet to start (bps.go.id). Meanwhile, the unemployment rate indicates the percentage of the unemployed in the total labor force.

According to Statistics Indonesia, the employment situation in February 2016 compare to February last year is as follows. The number of working people falls by 200 thousand people, which mainly occurs in the Agricultural sectors while the number of unemployed people drops as many as 430 thousand people. Further, the labor force participation rate is decreased by 1.44 percentage point whereas the unemployment rate decreases by 0.31 percentage point. In other words, in February 2016, the unemployment rate of Indonesia falls to 5.5 percent of the nation's labor force, which includes 7.02 million people compare to its rate in 2015 that is 5.81 percent. Whereas in August 2016, the number of labor force as many as 125.44 million people increases to as many as 3.06 million in August 2015. The number of working people increases as many as 3.59 million people while unemployed people decreases as many as 530 thousand. Further, the labor force participation rate is increased by 0.58 percentage point while the unemployment rate is decreased by 0.57 percentage point. Figure 1 shows the statistics of Indonesia's Unemployment from 2006 to 2013.

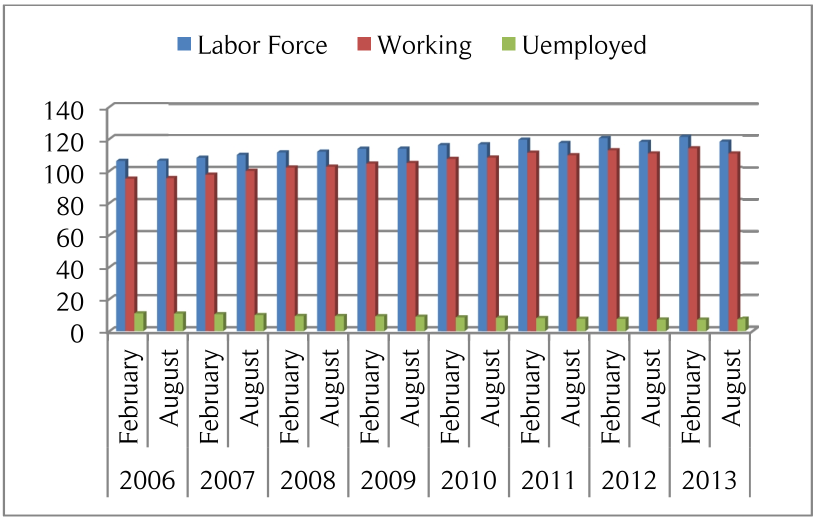

Figure 1. Indonesia's unemployment statistics 2006-2013 (in millions)

Figure 2 shows the unemployment rate in Indonesia from 2006 to 2013. Although the unemployment rate steadily goes down in recent years, Indonesia still has a large amount of people who cannot earn enough to make a decent living due to the definition of an employed person is those who work at least one hour per week. The number looks good although there are millions of people in Indonesia who spent very little time on works.

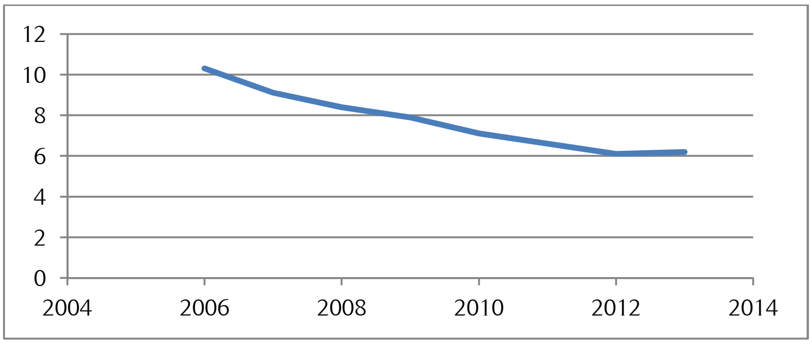

Figure 2. Indonesia's unemployment rate 2006-2013

This study aims to predict unemployment rate in Indonesia by using ARIMA model, which is compatible for a short time period forecasting. Due to the basic and essential properties of a time series, the various steps are needed to be fulfilled in order to use ARIMA model for time series analysis are as follows. Plot the data to see whether the actual data is stationary. Model identification is to determine several possible ARIMA models that can be used for forecasting analysis then the next step is to estimate the parameters of these selected models. Diagnostic checking is to test the assumptions of the residuals to obtain the best model. Finally, forecasting model to predict several values for the future based on the actual series. However, stationary is the most important properties of the use ARIMA method to forecast time series data. Therefore, when the actual data is non-stationary then it is needed a kind of way to make them stationary, one common way is through a differencing process.

A large number of studies are done on the unemployment rate prediction. Montgomery, Zarnowitz, Ruey, & Tiao (1998) predict the United States unemployment rate by using both linear and nonlinear time series models. The results reveal that forecasting accuracy can be improved by using the existing methods. Whereas Floros (2005) uses the United Kingdom unemployment rate in the period of January 1971 to December 2002 to compare the out-of sample forecasting accuracy such as root mean square errors (RMSE), mean absolute errors (MAE) and mean absolute percent errors (MAPE). The results indicate that both MA(1) and AR(4) are the best models for forecasting the unemployment rate in the UK while the mixture MA(4)-ARCH(1) model produces superior forecast whereas MA(4) model produces a well forecasts. Another studies use the UK unemployment rate for forecasting analysis are Johnes (1999), Peel & Speight (2000), Gil-Alana (2001). Kurita (2010) uses ARFIMA model to forecast the Japan's unemployment rate while Nkwatoh (2012) uses several univariate time series models for forecasting unemployment rate in Nigeria, such as trend regression analysis, ARIMA, GARCH, and the mixed ARIMA/ARCH models. The results reveal that ARIMA/GARCH model with an ARIMA (1,1,2)/ARCH (1) outperforms the other models.

Basically, the unemployment rate is an important indicator which is often used by foreign exchange market participants to analyze how well the economics in a country. A low unemployment rate indicates a well shape of the national economic. Therefore, forecasting the unemployment rate in Indonesia is important to be done due to its benefit to provide the accurate picture of national economic as well as the valid references for the sakes of investment improvement. This study uses the unemployment rate in Indonesia from 1986 to 2015 to forecast the future unemployment rate by using R version 3.3.1.

Research Method

The most well-known forecasting technique for time series data is autoregressive integrated moving average (ARIMA), is also known as Box-Jenkins model introduced by Box & Jenkins (1976). The most important things to consider in forecasting using this method is the stationary characteristic of the data. Typically, the time series data is non-stationary so the differencing process is required for making the data into stationary, which calculates the difference in observed values. Stationary on the data means that there are no significant fluctuations of the data because this must be horizontally along the time axis. In other words, its fluctuations must be around the constant mean. However, there are three models are usually used in the literatures related to time series; autoregressive (AR) model, moving average (MA) model and ARIMA model. The coefficient of correlation between two values in time series data is known as autocorrelation function (ACF) which is a way to measure the relationship between an observation at time t and the observation at previous times. Whereas partial autocorrelation function (PACF) is the calculated correlation of the transformed time series which is most useful for identifying the order of an autoregressive model.

Autoregressive (AR)

Autoregressive is representation of regression a value from a time series on previous values from similar time series. General model of Autoregressive order p (AR (p)), is also known as ARIMA (p,0,0) is defined as follows (Wei, 2006):

$$Z_t=\varphi_1 Z_{t-1}+\varphi_2 Z_{t-2}+\ldots+\varphi_p Z_{t-p}+e_t \quad (1)$$

Where Z_t represents the value of data at times t, φ_p is autoregressive parameter at p, $Z_{t-1}, \ldots . Z_{t-p}$ represent dependent variables while e_t is error at times t. The equation (1) can be rewritten in the following form:

$$\varphi_p(B) Z_t=e_t \quad (2)$$

where $\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p$. The regression of this model describes the response variable in the previous time period has become the predictor.

Moving average (MA)

This model is applied when the output variable depends linearly on the current and previous values. The general model of moving average order q (MA(q)), is also known as ARIMA (0,0,q) is defined as follows (Wei, 2006):

$$Z_t=e_t-\theta_1 e_{t-1}-\theta_2 e_{t-2}+\ldots+\theta_q e_{t-q} \quad (3)$$

Where θ_q is moving average parameter, e_t is random error at time *t*, e_{t-q} is random error at time (*t-q*). The equation (3) can be rewritten as the following form:

$$Z_t = \theta_q(B)e_t \tag{4}$$

where $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_q B^q$ represents MA operator.

Autoregressive integrated moving average (ARIMA)

ARIMA (*p,d,q*) indicates a stationary time series after differencing process *d*, with AR model of order *p* and MA model of order *q*. The general form of ARIMA (*p,d,q*) is as follows:

$$\varphi_p(B)(1-B)^d Z_t = \theta_0 + \theta_q(B)e_t \tag{5}$$

Where $\varphi_p(B)$ is a stationary AR operator and $\theta_q(B)$ is an invertible MA operator. Then ARIMA (1,1,2) with differencing $1 - B$ can be written as follows (Wei, 2006):

$$Z_t = (1-\varphi_1)Z_{t-1} + \varphi_2 Z_{t-2} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \tag{6}$$

Generally, steps of time series analysis by using ARIMA (*p,d,q*) are plotting the data, model identification, model estimation, diagnostic checking, and forecasting model.

Results and Discussion

This study uses the data unemployment rate in Indonesia from 1986 to 2015 which is taken from the National Labor Force Survey conducted by Statistics Indonesia. However, there are only 29 data where the data of unemployment rate in 1995 is not provided. The data indicates that the lowest rate is 2.5 point in 1990 while the largest rate is 11.2 in 2005. General overview of the unemployment rate is presented in table 1 below. The unemployment rate data is presented in Figure 3 while Figure 4 provides the plot of ACF and PACF.

Table 1. Descriptive statistics

Statistics	Rate
Minimum	2.5
Maximum	11.2
Mean	6.048276
Std. Dev	2.686955
Skewness	0.188896
Kurtosis	-1.0662

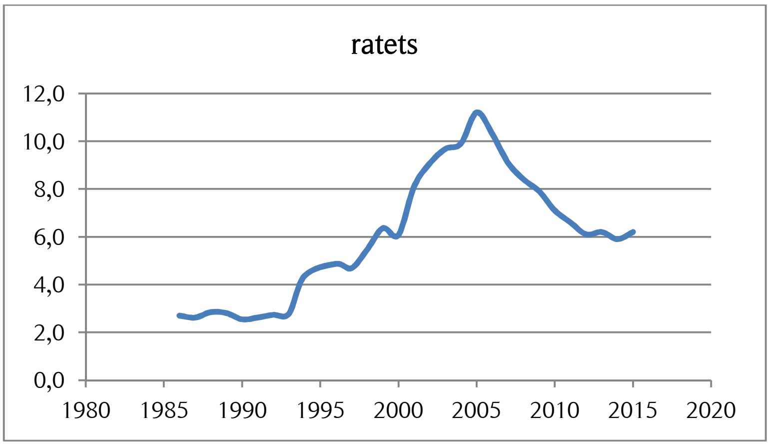

Figure 3. Unemployment rate

Generally, the general statistics of the unemployment rate indicates that the standard deviation, which is a dispersion average of the mean, is 44.43%. The skewness is positive which indicates that the data distribution tends to be on the right of the normal distribution while the negative value of kurtosis indicates that the distribution does not tend to peak. Figure 3 shows the unemployment rate plot while Figure 4 represents the plot of ACF and PACF.

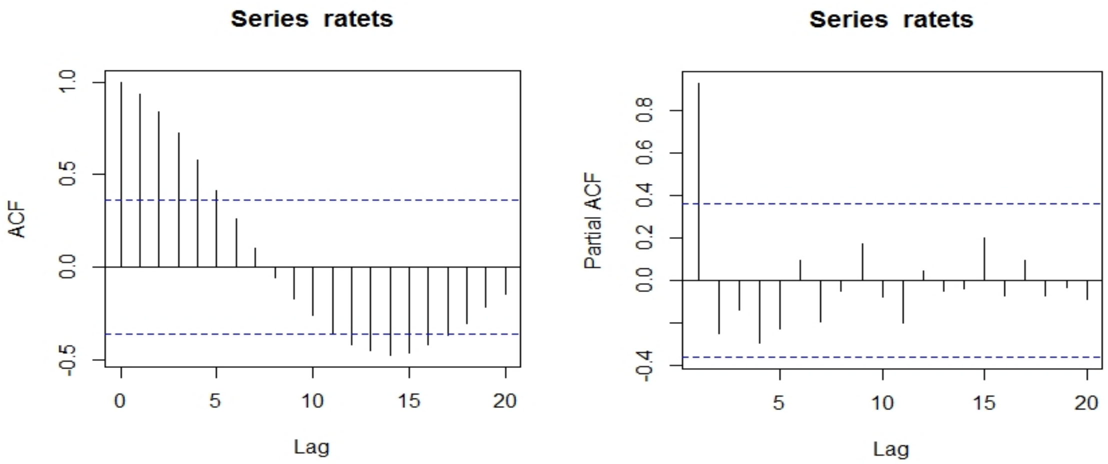

Figure 4. ACF and PACF of the unemployment rate

According to Figure 3, the actual data are clearly non-stationary as the series shows significant fluctuation on the unemployment rate where the highest increase is 2.0 point from 2000 to 2001 meanwhile the highest decrease is 1.2 point from 2005 to 2006. However, since 2005 the unemployment rate in Indonesia indicates consistently goes down except for the period from 2012 to 2013 which is increased by 0.1 point. Further, this rate decreases decisively by 4.1 point in the last ten years. Besides, ACF plot in Figure 4 indicates the data is slowing down to zero which means it is non-stationary. Moreover, the Augmented Dickey-Fuller test informs strongly non-stationary. Due to AR model and MA model are not applicable on non-stationary series so it is required the process of differencing by disposing the trends to provide stationary time series data. Further, based on the Augmented Dickey-Fuller test indicates that a first difference of the data (d=1) is yet to produce stationary characteristic (Dickey-Fuller = -1.7734, Lag order = 3 and p-value 0.6592, which means that the null hypothesis is not rejected). Therefore, the second differencing process (d=2) is needed to be done. The results of the Augmented Dickey-Fuller test indicates that the null hypothesis is rejected (Dickey-Fuller = -3.8691, Lag order = 2, and p-value = 0.03038) which means that by using 95% confidence level then the second differencing process of the unemployment rate produces stationary data that indicates the assumption of ARIMA model is fulfilled. The differenced data are shown in Figure 5. The next step is the identification of acceptable ARIMA models. Figure 5 shows the unemployment rate while figure 6 describes the plot of ACF and PACF from second differencing process.

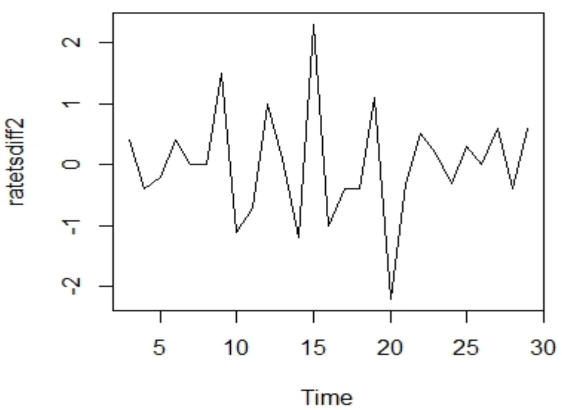

Figure 5. Unemployment rate (d=2)

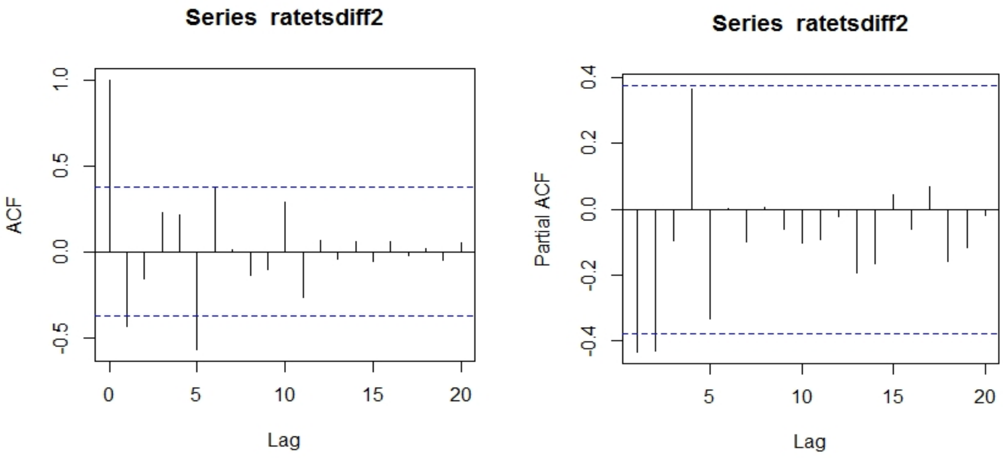

Figure 6. ACF and PACF of the unemployment rate (d=2)

Due to the second difference of the data provides stationary data then the ARIMA (p,2,q) model is a suitable model. The next step determines the order of AR model (p) as well as the order of MA model (q) to obtain the best ARIMA model for forecasting process. Figure 6 indicates that PACF plot then it looks cut off at the second lag so that the AR (2) model can be used in the analysis so an initial candidate model is the ARIMA (2,2,0). Turning to ACF plot looks cut off at the first lag so it is likely the data is generated by the MA (1) model which means that the model can be used in the analysis. It comes as no surprise that the ARIMA (2,2,1) model perhaps suitable to predict the unemployment rate in Indonesia. Although it is possible there are other forms of ARIMA models. All things considered, however, the alternative ARIMA models and its summary results are used in this study are presented in table 2.

Table 2. Results of ARIMA models

ARIMA	SIGMA	LOG	AIC	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
(2,2,1)	0.51	-29.46	66.92	0.007126	0.687638	0.47568	0.948244	7.737826	0.81712	-0.03204
(2,2,2)	0.46	-28.35	66.69	0.014363	0.655676	0.498381	0.946595	8.197917	0.856115	-0.09611
(2,2,0)	0.51	-29.51	65.02	0.008093	0.689085	0.482117	0.939228	7.830741	0.828176	-0.06134
(0,2,1)	0.54	-30.34	64.68	-0.00389	0.711738	0.49535	0.761543	8.129948	0.850908	-0.04121
(0,2,2)	0.54	-30.29	66.57	0.002582	0.710604	0.506155	0.832145	8.303136	0.869468	-0.00687
(1,2,1)	0.54	-30.32	66.64	-0.00179	0.711364	0.499107	0.785079	8.192001	0.857362	-0.0275
(1,2,2)	0.54	-30.33	68.67	-0.00446	0.711498	0.49406	0.757889	8.10429	0.848692	-0.05011
(1,2,0)	0.63	-32.13	68.26	0.0094	0.76452	0.57017	0.709518	9.448945	0.979433	-0.19685

This study relies on AIC value to determine the best ARIMA model that can be used in forecasting the unemployment rate in Indonesia. Table 2 reports that the best model is ARIMA (0,2,1) model which has the lowest AIC value (64.68). The validation of the ARIMA model can be done by plotting the fitted model with the original series, which figure 7 shows ARIMA (0,2,1) fitted model with the original series where blue line represents the fitted values while red line represents the actual data. However, it is common to just look at the output of the forecast accuracy to determine the best model. There are accuracy measures for forecast model such as mean error (ME), root mean squared error (RMSE), mean absolute error (MAE), mean percentage error (MPE), mean absolute percentage error (MAPE), mean absolute scaled error (MASE), autocorrelation of errors at lag 1 (ACF1).

Based on Figure 7 it can be seen that the fitted values continually go along with the original values. The assumptions that the residuals are uncorrelated and normally distributed are required to be check due to the forecast confidence intervals for ARIMA models are depended on these assumptions. Therefore, it is important to plot ACF and histogram of the residuals before producing forecast intervals. Figure 8 shows that the forecast errors are normally distributed with mean zero. The Ljung-Box test produces X-squared = 20.56, df = 20, p-value = 0.4234 that means there is little evidence of non-zero autocorrelations in the sample forecast errors. Therefore, due to the assumptions of the 80% and 95% predictions intervals are based on there are no autocorrelations in the forecast errors as well as these values are normally distributed with mean zero and constant variance then it can be said that the assumptions are probably valid. Therefore, it can safely assume that the predictive model cannot be improved. Table 3 shows the forecasting by using ARIMA (0,2,1) for the upcoming ten years while figure 9 and figure 10 indicate the forecasting plot and the plot of ACF and PACF of these results, respectively.

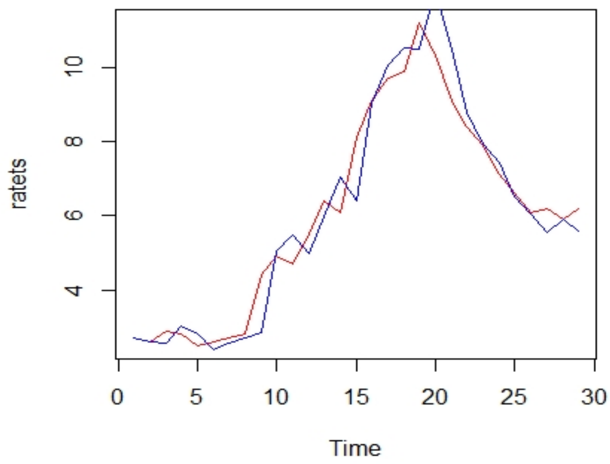

Figure 7. Fitted values and the original series

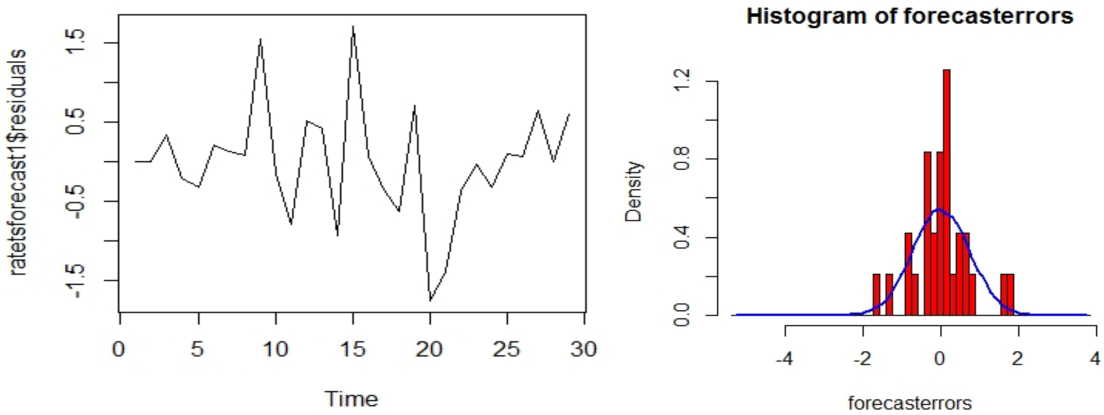

Figure 8. Residuals plot and distribution of forecast errors

Table 3 informs the forecasting results for the next ten years and the prediction intervals, which is an interval associated with a random variable yet to be observed, with a specified probability of the random variable lying within the interval. As a matter of fact, the prediction intervals play a key role in forecasting analysis due to these values express how much the uncertainty is associated with each forecast. The prediction intervals tend to grow wider when the forecasting provides higher uncertainty. However, the lower and upper limits of these intervals are also presented. Table 3 also indicates that the model is fitted by using 80^{th} and 95^{th} prediction intervals. The general description of the point forecast for the next ten years is as follows. In 2016, the unemployment rate in Indonesia is expected to be the highest rate (6.1 percent) while the minimum rate is expected to be 5.5 percent in 2025. The average of the point forecast is 5.8 percent with standard deviation as many as 0.2 percent. Furthermore, according to table 3, the point forecast indicates that unemployment rate in Indonesia is expected to be 6.1 percent at the end of 2016 and it is expected to decrease by 0.1 point to be 6.0 percent at the end of 2017. In addition, the point forecast indicating a linear line which decreases continuously.

Table 3. Forecasting results

Year	Point Forecast	Forecast Interval			
		Lo 80	Hi 80	Lo 95	Hi 95
2016	6.127414	5.182107	7.072722	4.6817	7.5731
2017	6.054829	4.445533	7.664125	3.5936	8.5160
2018	5.982243	3.670599	8.293888	2.4469	9.5176
2019	5.909658	2.842048	8.977267	1.2182	10.6012
2020	5.837072	1.958357	9.715788	-0.0949	11.7691
2021	5.764487	1.020929	10.50805	-1.4902	13.0191
2022	5.691901	0.031868	11.35193	-2.9644	14.3482
2023	5.619316	-1.00664	12.24528	-4.5142	15.7528
2024	5.54673	-2.09254	13.186	-6.1365	17.2300
2025	5.474145	-3.22392	14.17221	-7.8284	18.7767

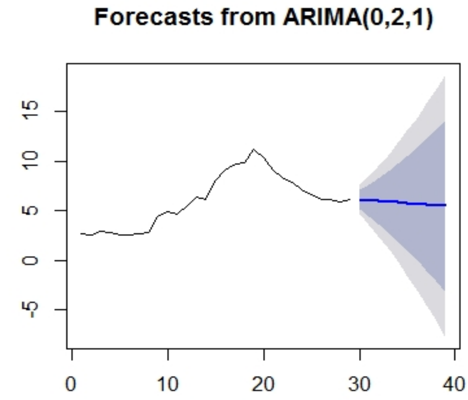

Figure 9. Forecasting plot from ARIMA (0,2,1)

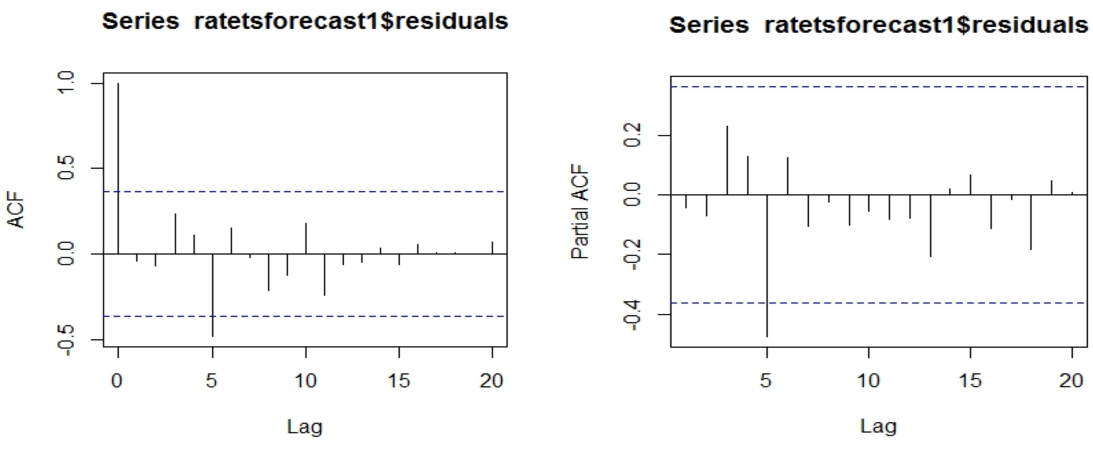

Figure 10. ACF and PACF plot of forecasting plot ARIMA (0,2,1)

Figure 9 indicates that the forecasts are shown as a blue line, with the 80% prediction interval as a dark shaded area, and the 95% prediction intervals are as a bright shaded area. The forecast of the unemployment rate in Indonesia with an 80% prediction interval indicates that the actual unemployment rate should lie within the interval with probability 0.8. The prediction intervals for both the 80% and 95% continuously increase when forecast the further time of years. Moreover, both of the ACF plot and the PACF plot show that the sample autocorrelation for forecast errors at lag 5 is exceed the significance bounds. Basically, a good forecasting method yields the residuals with zero mean to obtain unbiased results. However, the forecasting results inform the average of the residuals is -0.0038 which is closer to zero. In other words, the ARIMA (0,2,1) model produces a good results of forecasting the unemployment rate in Indonesia.

Conclusion

This study aims to predict the unemployment rate in Indonesia using time series data from 1986 to 2015. The ARIMA model is one of the popular methods for predicting the future values based on time series data. However, the model requires some properties that have to be fulfilled in order to produce empirical results accurately. This study suggests that the ARIMA (0,2,1) model is the best method for forecasting the unemployment rate in Indonesia. The results reveal that the future unemployment rate tends to decrease continuously while the point forecast indicates that unemployment rate in Indonesia is expected to be 6.1 percent at the end of 2016 and it is expected to decrease by 0.1 point to be 6.0 percent at the end of 2017. However, the lower and upper limits of prediction levels are presented due to these values express how much the uncertainty is associated with each forecast. The prediction intervals tend to go along with the forecast horizon, which indicates that they increase in length when the forecast horizon increases.

References

Floros, C. (2005). Forecasting the UK unemployment rate: Model comparisons. *International Journal of Applied Econometrics and Quantitative Studies*, *2*(4), 57-72.

Gil-Alana, L.A. (2001). A fractionally integrated exponential model for UK unemployment. *Journal of Forecasting*, *20*(5), 329-340.

Johnes, G. (1999). Forecasting unemployment. *Applied Financial Economics*, *6*(9), 605-607.

Kurita, T. (2010). A forecasting model for Japan's unemployment rate. *Eurasian Journal of Business and Economics*, *3*(5), 127-134.

Montgomery, A.L., Zarnowitz, V., Ruey. S.T., & Tiao, G.C. (1998). Forecasting the US unemployment rate. *Journal of the American Statistical Association, 442*(6). 478-493.

National Labour Force Survey, Statistics Indonesia (Badan Pusat Statistik/BPS), Republic of Indonesia.

Nkwatoh, L.S. (2012). Forecasting unemployment rates in Nigeria using univariate time series models. *International Journal of Business and Commerce*, *1*(12), 33-46.

Peel, D.A., & Speight, A.E.H. (2000). Threshold nonlinearities in unemployment rates: Further evidence for the UK and G3Economies. *Applied Economics*, *32*(6), 705-715.

Wei, W.W.S., (2006). *Time series analysis unvariate and multivariate methods* 2nd edition. United State of America: Addision-Wesley Publishing Company.