DE GRUYTER OPEN

DOI 10.1515/ptse-2017-0005 PTSE 12 (1): 43-50

Autoregressive Integrated Moving Average Model to Predict Graduate Unemployment in Indonesia

Umi MAHMUDAH

u_mudah@yahoo.com
(State Islamic University of Pekalongan, Pekalongan, Indonesia)

Received: 10.06.2017; Accepted: 31.07.2017

Abstract: *Nowadays it is getting harder for higher education graduates in finding a decent job. This study aims to predict the graduate unemployment in Indonesia by using autoregressive integrated moving average (ARIMA) model. A time series data of the graduate unemployment from 2005 to 2016 is analyzed. The results suggest that ARIMA (1,2,0) is the best model for forecasting analysis, where there is a tendency of increasing number for the next ten periods. Furthermore, the average of point forecast for the next 10 periods is about 1,266,179 while its minimum value is 1,012,861 the maximum values is 1,523,156. Overall, ARIMA (1,2,0) provides an adequate forecasting model so that there is no potential for improvement.*

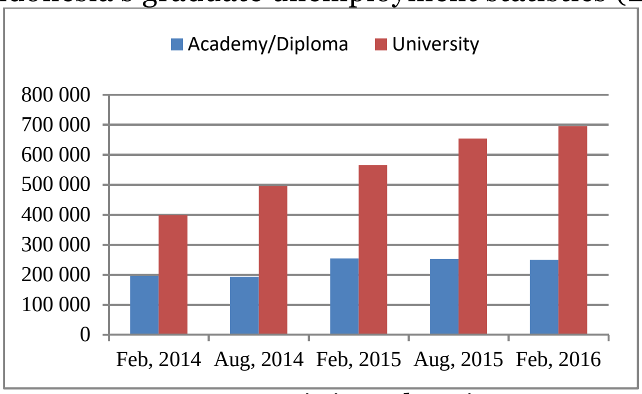
Keywords: forecasting, time series, graduate unemployment, higher education

Introduction

Education is one of the most important aspects in life because it's vital role in promoting all things in this world. It provides both skills and tools the people are needed not only to survive the changing times but also to improve their lives. Well-educated human resources tend to adapt better to any changes in the surrounding environment so that they are expected to have no difficulty in finding suitable job because higher education suppose to give a better chance to everything. Finding a decent job is an assertion for every graduate to improve his standard of living. Unfortunately, recent trends indicate that the higher education graduates struggle to secure their job instead they are forced to take a role that does not match their qualifications. Basically, an educated unemployment or graduate unemployment is unemployment among people with an academic degree. The unemployed graduate is a very serious problem especially for developing countries like Indonesia because of high expectations on them to be able to advance the national competitiveness. With the increasing number of unemployed graduates, it is feared that public sentiment in higher education institutions may decline since it seems that they are not capable to produce graduates who are able to put themselves in the labor market. A data from the Federal Reserve Bank of New York, the United States Census Bureau, and the American Community Survey reports that the unemployment and underemployment rates for recent college graduates between the ages 22 and 27. Geography and anthropology majors provide graduates with the highest unemployment rate (8.8%) followed by major in mass media by 8.6%. Meanwhile education major has the lowest unemployment rate (1.0%) followed by majors in construction services and agriculture by 1.8%. Afterwards, the highest underemployment rate is occupied by criminal justice major (74.4%) while the lowest rate is major in nursing by 13.4%. Whereas in 2008, the unemployment rate of graduates in China is more than 30%. *OECD* (2015) report that Greece, Spain and Portugal are the worst unemployment rates for graduates among OECD countries in 2013. *Financial Times* in 2016 reports that the unemployment rate for graduates in the United Kingdom is 3.1% with 2.3% for workers with a postgraduate qualification and 6.4% for non-graduates.

Indonesia as an emerging economy needs to overcome the complexity issues such as graduate unemployment rate to improve its competitiveness due to the increasingly high demands of times. According to Statistics Indonesia the unemployment rate in Indonesia from 2006 to 2013 steadily goes down while the number of graduate unemployment (Diploma and Bachelor) has fluctuated.

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



Source: Statistics Indonesia

Figure 1 shows the number of unemployed graduates of higher education in Indonesia from February 2014 to February 2016. However, the number shows an increasing trend from 2015 to February 2016. In the period of February in 2014 there are as many as 593,556 people and it is increasing to 688,660 in the period of August in the same year. The number increases again as many as 131,054 people in February 2015. Then in August in the same year it increases by 85,413 to 905,127. Meanwhile in February 2016 the rise is not as much as in previous period, that is as many as 39,539. This figure is alarming because it indicates that the system of higher education in Indonesia has not been able to produce the qualified graduates who have skills and ability to be accepted by labor markets.

The figure above also shows that unemployment of university's graduates is much higher than the graduate unemployment of diploma or academy, where the difference is always above 30 percent. The highest difference is occurred in February 2016 (47 percent), which includes as many as 695, 304 unemployed university's graduates compare to 249,362 unemployed with degree in diploma or academy. The high number of graduate unemployment is a very important issue in a country because it represents institutional ineffectiveness and inefficiency of higher education. Therefore, a better system of higher education is needed so that graduates can develop the skills and knowledge that they need to progress into fulfilling careers. Thus, it is important to predict the graduate unemployment to provide the accuracy picture of higher education quality in a country. Unfortunately, researches on the same topic are very limited, especially on prediction analysis to provide a picture of the future value. This study aims to forecast the number of graduate unemployment in Indonesia by using ARIMA model, which is the most popular model in time series data prediction. Graduate unemployment data in Indonesia from 2005 to 2016 is analyzed by using R version 3.3.1 to provide point forecasts as well as the prediction intervals.

Method

The most well-known forecasting technique for time series data is autoregressive integrated moving average (ARIMA), is also known as Box-Jenkins model. 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. The general form of ARIMA (p,d,q) model is as follows:

$$y_t^{'} = \varphi_1 y_{t-1}^{'} + \cdots + \varphi_p y_{t-p}^{'} + \theta_1 e_{t-1} + \cdots \theta_q e_{t-q} + e_t \quad (1)$$

where y_t' is the differenced series. The equation (1) can be rewritten in the following terms:

$$\left(1 - \varphi_1 B - \varphi_p B^p\right)(1 - B)^d y_t = \left(1 + \theta_1 B + \cdots + \theta_q B^q\right) + e_t \quad (2)$$

where $\varphi_p(B)$ is a stationary AR operator and $\theta_q(B)$ is an invertible MA operator. Therefore, the stationarity condition that is used in AR model and the condition of invertibility from MA model apply are applied to this ARIMA model. The following steps are required in analyzing time series data using ARIMA model.

Stage 1 Stationary Test. Since the ARIMA model applies only on stationary time series data then the first step that needs to be done is to identify whether the original data has met the stationary characteristics. Plotting the actual data is a well-known way to know the stationary data as well as unit root test is also applicable. One of the tests commonly used for unit root test is by using the Augmented Dickey-Fuller test. Statistical test provides a more objective solution than any other methods in determining stationary time series data set.

Stage 2 Identifying Temporary Model. When the first stage states that the actual data is stationary then the next stage is to identify several ARIMA models that probably can be used in prediction analysis. Otherwise, if the original data is non-stationary then a differencing process is performed, which shows the value of d in the ARIMA (p,d,q) model. Nest is to determine the order of both AR(p) model and MA(q) model. Autocorrelation function (ACF) and partial autocorrelation function (PACF) are usually used to determine both orders p and q.

Stage 3 Parameter Significance Test. After obtaining some ARIMA models that may be used, the next step is to calculate the AR parameters and MA parameters in each model. Then, the parameter significance test is performed to determine they are significant or not. It is important to note that a feasible model for prediction analysis using ARIMA is a model whose parameter values are statistically significant. In addition, the eligible model has the smallest σ^2 and the largest log likelihood estimate.

Stage 4 Determining the Best ARIMA Model. There are several criteria to determine the best ARIMA model for forecasting process, such as using Akaike's Information Criterion (AIC), AIC-corrected (AICc) and Bayesian Information Criterion (BIC), which is the smaller value of the three criteria is a better model. However, 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).

Stage 5 Diagnostic Checking. Before the forecasting process using the best model has been done, a diagnostic test is needed to prove that the model is adequate. Non-autocorrelation of residuals test can be identified using ACF and PACF plots of residuals. Besides, normality test also needs to be done

Stage 6 forecasting. The last stage is to forecast the actual data to predict the values for the next period using the best ARIMA model based on the above stages.

Results and Discussion

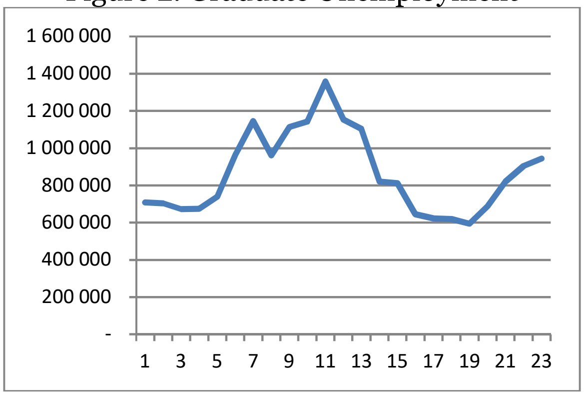
The graduate unemployment data in Indonesia is based on the National Labor Force Survey which is conducted annually. Since 2005 the survey was conducted twice a year in February and August. The data is a part of unemployment rate data in Indonesia that is published by Statistics Indonesia, which is in the category of the unemployment by education attainment. Table 1 shows a general overview of time series data of educated unemployment in Indonesia from 2005 to 2016.

Table 1. Descriptive Statistics

Statistics	Graduate Unemployment
Minimum	593,556
Maximum	1,358,206
Mean	865,755.2609
Std. Deviation	218,145.1494
Skewness	0.6094
Kurtosis	-0.6801

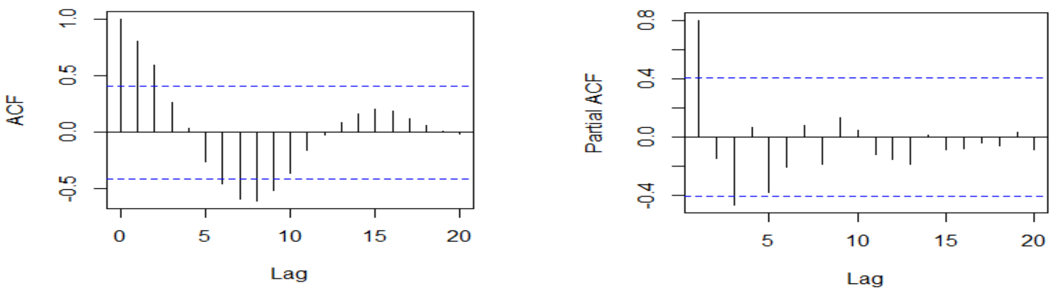
Table 1 shows that the minimum value of the analyzed data is 593,556 which is occurred in the period of February 2014 while the maximum value is 1,358,206 in February 2010. Figure 2 shows plot of graduate unemployment data in Indonesia that is used in this study. From table 1 it can be seen that the skewness value is 0.6094 and it is positive value that indicates the data distribution tends to be on the right side of the normal distribution. Meanwhile the kurtosis is -0.6801 and has negative sign that indicating the distribution of data does not tend to peak.

Figure 2. Graduate Unemployment



The first step to perform prediction analysis using ARIMA model is stationary test. From figure 2 it can be seen clearly that the actual data of educated unemployment in Indonesia from 2005 to 20016 have non-stationary characteristics, which has a considerable fluctuation, which the average of the fluctuation is about 13,594. Then, the number is increased the most in February 2007 to august in the same year, as many as 223,573 people. Meanwhile the most decrease is 284,419 people in the period of February 2011 to August of the same year. Besides, the number tends to rise from February 2006 and it starts to decline in February 2008. Furthermore, the number of unemployed graduate in Indonesia from February 2010 to August 2013 shows a downward trend, but from February 2014 to February 2016 tends to increase. Figure 3 shows plot of autocorrelation function (ACF) and partial autocorrelation function (PACF).

Figure 3. ACF and PACF of Graduate Unemployment



In addition, to the original data plot, in order to know the characteristics of the actual data can also be observed through ACF and PACF coefficients. From figure 3 shows ACF plot indicates that the data is shrinking slowly close to zero after the first lag. Therefore, the differencing process is needed to obtain stationary characteristics. There are several unit root tests which are based on different assumptions. But one of the most popular tests is the Augmented Dickey-Fuller (ADF) test. Then, based on the Augmented Dickey-Fuller test shows that by using 5% threshold, then differencing process is needed because the *p*-value is greater than 0.05, where Dickey-Fuller = -2.3228, Lag order = 2, *p*-value = 0.4494 which means that the null hypothesis is not rejected. Since the first stage does not meet the required stationary data then the second stage of differencing process has to be done. The results of the Augmented Dickey-Fuller test indicate that the data is already stationary (Dickey-Fuller = -3.8122, Lag order = 2, *p*-value = 0.03485). In other words, by using 95% confidence level then the assumption of ARIMA model that requires stationary data is successfully fulfilled. Due to the data is stationary at the second differencing process then ARIMA (1,2,0) can be used in prediction analysis. However, several ARIMA model are considered to determine the best ARIMA model for predicting the number of graduate unemployment in Indonesia. Table 2 shows the alternative ARIMA models and its summary results, which are used as a guide in determining the best model.

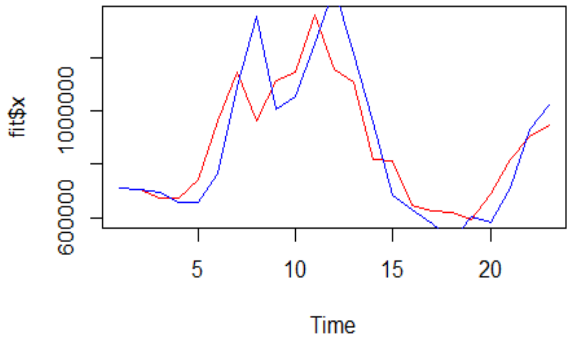
Table 2. Results of ARIMA Models

No	ARIMA	SIGMA	LOG	AIC	AICc	BICc
1	(1,2,0)	2.00E+10	-279.10	562.19	562.86	564.28
2	(2,2,0)	2.00E+10	-279.09	564.17	565.59	567.31
3	(1,2,1)	2.00E+10	-279.14	564.28	565.69	567.41
4	(1,2,2)	1.82E+10	-278.20	564.40	566.90	568.58
5	(1,2,3)	1.37E+10	-276.44	562.88	566.88	568.10
6	(0,2,1)	1.81E+10	-279.34	562.68	563.34	564.76
7	(0,2,2)	1.81E+10	-279.30	564.59	566.00	567.72
8	(2,2,2)	1.83E+10	-277.01	564.03	568.03	569.25

It is important to note that the best ARIMA model needs significant parameter values. From the models in table 2, there are only two models whose parameters are significant, i.e. ARIMA (1,2,0) and

ARIMA (0,2,1). Then the next step is to determine which model is better to be used in forecasting analysis. There are several criteria for comparing quality of fit across multiple models, but Akaike information criteria (AIC) and Bayesian information criteria (BIC) are the most commonly used. When comparing various models for the same data set to determine the best fitting model then the preferred model is the model that has the smallest value of AIC, AICc (AIC-corrected), and BIC. Therefore, from table 2 it can be seen that ARIMA (1,2,0) is the best model because the three criteria have a smaller value than ARIMA (0,2,1). Figure 4 indicates the plot of ARIMA fitted model with the original series of graduate unemployment, which appears that the fitted values always follow the original values.

Figure 4. Fitted Models and the Actual Series



The next step is to predict the graduate unemployment for next 10 years by using ARIMA (1,2,0). Forecasting results is presented in table 3 and figure 5 where shows point forecasts as well as the prediction intervals from forecasting analysis. Computing prediction interval for forecasting is usually important to accompany the point forecasts, which is very useful to express the uncertainty in the forecasts because it is difficult to tell how accurate the forecast if only using point 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. Forecasting Results

Year	Period	Point Forecast	Prediction Intervals			
			Lo80	Hi80	Lo95	Hi95
2016	Aug	1,012,861	831,603	1,194,119	735,651	1,290,071
2017	Feb	1,063,155	754,934	1,371,377	591,771	1,534,540
	Aug	1,124,631	629,066	1,620,196	366,730	1,882,533
2018	Feb	1,179,123	487,373	1,870,873	121,183	2,237,063
	Aug	1,237,977	319,113	2,156,842	-167,305	2,643,259
2019	Feb	1,294,106	133,049	2,455,163	-481,577	3,069,790
	Aug	1,351,938	-72,642	2,776,518	-826,769	3,530,645
2020	Feb	1,408,706	-294,995	3,112,407	-1,196,880	4,014,292
	Aug	1,466,138	-533,886	3,466,163	-1,592,636	4,524,912
2021	Feb	1,523,156	-788,006	3,834,317	-2,011,461	5,057,772

Figure 5. Forecasting plot of ARIMA (1,2,0)

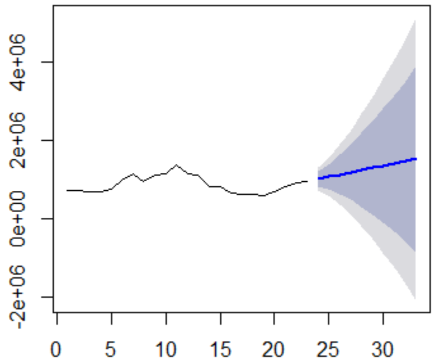


Table 3 indicates that the model is fitted by using 80th and 95th prediction intervals, where the forecasts for the period of August 2016-February 2021 are plotted as a blue line with the 80% prediction interval as a dark shaded area, and the 95% prediction intervals are as a bright shaded area. The results reveal that the number of graduate unemployment in Indonesia tends to increase continuously from 2016 to 2021, where the smallest amount is occurred in the period of August 2016 to February 2016 (50,294) while the largest number is 61,476 and it is occurred in February 2017 to August in the same year. Furthermore, the average of point forecast for the next 10 periods is about 1,266,179 while its minimum value is 1,012,861in August 2016 and the maximum values is 1,523,156 in February 2021. The last step is checking whether the residuals are uncorrelated and normally distributed because the forecast confidence intervals for ARIMA models are depended on these assumptions. A Ljung-Box test is used to determine whether there is significant evidence for non-zero correlations. The test results provide the Ljung-Box test statistic is 14.865 with degree of freedom is 20 and *p*-value is 0.7841, which indicates there is little evidence of non-zero autocorrelations in the forecast errors when lags are set from 1 to 20.

Conclusions

This study uses ARIMA model to predict the graduate unemployment in Indonesia using time series data from 2005 to 2016. After fulfilling several steps that must be met in ARIMA model, this study suggests that the ARIMA (1,2,0) model is the best model for forecasting the graduate unemployment in Indonesia. The forecasting results show a tendency of increasing the number of graduate unemployment in Indonesia for the next 10 periods.