COMPARATIVE STUDY OF SMOOTHING METHODS AND BOX-JENKINS MODEL IN FORECASTING UNEMPLOYMENT RATE IN MALAYSIA

Norliana Mohd Lip [1*], Nurul Lina Natasha Mohd Rizuan [1], Nur Izzati Iezudin[1], Nur Azzarina Mohamad [1], Nur Rasyida Mohd Rasyid[1] , Fatin Aliyah Hassan [2], Hanafi Ithnin[3]

[1]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Negeri Sembilan, Seremban Campus, 70300 Seremban, Negeri Sembilan, Malaysia.*

[2]*Academy of Language Studies, Universiti Teknologi MARA Negeri Sembilan, Seremban Campus, 70300 Seremban, Negeri Sembilan, Malaysia.*

[3]*Malaysian Nuclear Agency, Bangi, 43000 Kajang, Selangor, Malaysia.*

[*]*Corresponding author: norliana7287@uitm.edu.my*

Abstract

One of the significant problems faced by all countries in the world, especially developing countries is unemployment. It must be taken into account by the government regardless of whether its effect is critical or under control. According to the Department of Statistics Malaysia (DOSM), the unemployment rate refers to the proportion of the unemployed population to the total population in labour force. The high unemployment rate will reflect the negative impact on labour market. This study was conducted to identify the best model among smoothing techniques (double exponential smoothing and Holt's model) and the Box-Jenkins model for forecasting the unemployment rate in Malaysia since there is no appropriate model in estimating the unemployment rate in the future. The performance of the best model was determined by using several error measures such as Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Geometric Root Mean Square Error (GRMSE). The lowest error measure identified which model is appropriate to forecast the unemployment rate in Malaysia. The findings showed that the most appropriate model is ARIMA (2,1,3) in the Box-Jenkins method since it indicated the smallest value of all error measures. In addition, the study demonstrated that there was an increment in the unemployment rate in Malaysia since the forecasted trend line showed low fluctuation which gradually increasing and decreasing from January 2019 until December 2019. This implied that the future value of the unemployment rate in Malaysia is stable and not considered as a critical problem.

Keyword: Box-Jenkins, Double Exponential Smoothing, Holt's method, Unemployment rate

Introduction

The increasing rate of unemployment is one of the issues that recently triggers the world. According to Asif (2013), the results of regression analysis presented a significant impact of unemployment, gross domestic product (GDP), inflation, exchange rate as well as the increasing population rate of three countries which is India, China and Pakistan from the year 1980 to 2009. This study examined the impact of inflation, exchange rates, GDP, and population on the fluctuating rate of unemployment. The revised policies are recommended in order to overcome this situation. Dobre and Alexandru (2008) had modelled the evolution of unemployment rate of Romania using Box-Jenkins methodology during the monthly period of the year 1998 until 2007. Whereas Dumičić et al. (2015) had used several forecasting techniques for forecasting time series with trend component and seasonality, which were double exponential smoothing and Holt-Winters method in order to determine the most precise forecasting method in predicting the future values of the unemployment rate in selected European countries.

In Malaysia, the prediction of the unemployment rate was done by Ramli et al. (2018). The inflation rate and population were the two significant factors of the unemployment based on the results of the regression analysis. ARIMA (2,1,2) was the optimal model used to forecast the unemployment rate in Malaysia. Research done by Nor et al. (2018) was to identify the most accurate forecast method and suitable period to predict the future of Malaysia's unemployment rate in 2016. The methods used in the study were Naïve, Simple Exponential Smoothing (SES) and Holt's method. The findings revealed that Holt's method was the ideal method to be used in forecasting the overall yearly unemployment rate, male yearly unemployment rate and overall quarterly unemployment rate. Whereby SES was found to be the best method in forecasting the female yearly unemployment rate as well as overall monthly unemployment rate.

Forecasting is the entire process through which methods are used to generate future values that could be used for various purposes. Future event prediction was also defined based on previous values of relevant variables (Wheelwright et al., 1998). Several forecasted values are generated with the intention to provide general information on data. For example, the study concerning the future trend of sales and critical planning are very important to the firms to make some improvements. Moreover, present technology such as high-tech computers, with the production of more sophisticated predicting software, has made the prediction models work much smoother. Assis and Amran (2010) had compared the forecasting performances of different time series methods for forecasting the monthly prices of cocoa bean from January 1992 to December 2006. From the study, it is shown that time series forecasting techniques can be applied in various case studies.

The objective of this study is to determine the usefulness of smoothing techniques and the Box-Jenkins method in forecasting and estimating the rate of unemployment in Malaysia. Therefore, this study aims to use two different forecasting techniques to evaluate and resolve the researchers' issue where there is no appropiate method already verified in order to predict the future value of unemployment rate in Malaysia. The aims of this study are to determine the best model from these approaches and predict the future value of the unemployment rate based on the selected best models.

Materials and Methods

The Data

The analysis was conducted using secondary data obtained from the Department of Statistics Malaysia (DOSM) from January 2012 until December 2018. The data of unemployment rate in Malaysia was used and several forecasting methods were developed which are Double Exponential Smoothing technique, Holt's method and Box-Jenkins method. Monthly time series data were applied to those models starting from January 2012 to December 2018 with 84 data samples. Based on **Figure 1**, the study observed that the pattern line fluctuated within 84 months.

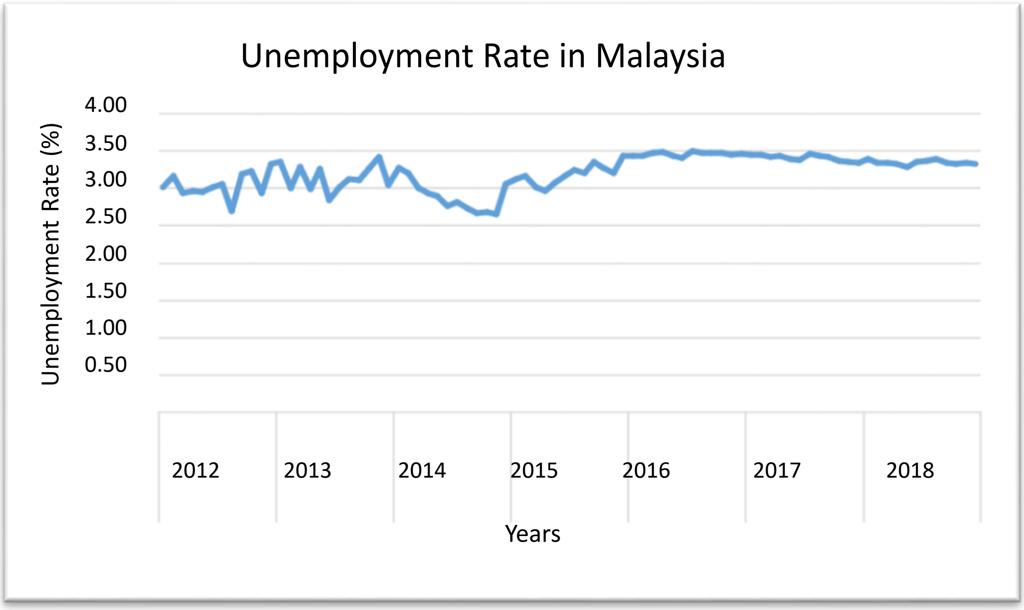

Figure 1 The Graph of Monthly Unemployment Rate

Methods

Double Exponential Smoothing model

Double exponential smoothing is also known as Brown's approach. This model can be used to generate multiple-ahead-forecasts and to update the components at each time with two weights or parameters (Brown, 1956). Consider the equations of double exponential smoothing (Lazim, 2016) as given in the equations:

$$D_t = \alpha Z_t + (1-\alpha)(D_{t-1}) \quad (1)$$

$$D_t' = \alpha D_t + (1-\alpha)(D_{t-1}') \quad (2)$$

$$C_t = 2D_t - D_t' \quad (3)$$

$$K_t = \frac{\alpha}{1-\alpha}(D_t - D'_t) \quad (4)$$

$$\tilde{Z}_{t+h} = C_t + K_t(h) \quad (5)$$

Where;

t : Time
h : Step into future
D_t : Exponentially smoothed value of Z_t, at time t
D_t' : Double exponentially smoothed value of Z_t, at time t
α : Parameter for the level, $0 < \alpha < 1$
C_t : Difference between the exponentially smoothed values, at time t
K_t : Adjustment factor, at time t
Z_t : Actual data value, at time t
$\tilde{Z}_{t+h}$: Fitted value or forecast value, at time $t + h$

Holt's model

Holt (1957) expanded simple exponential smoothing to allow for the prediction of trend data as the model could smoothen the trends and slopes directly. This approach consists of a predictive equation and two smoothing equations (one degree and one trend). The following are the equations of Holt's method (Lazim, 2016):

$$\tilde{Z}_{t+h} = C_t + K_t(h) \tag{6}$$

$$C_t = \alpha Z_t + (1-\alpha)(C_{t-1} + K_{t-1}) \tag{7}$$

$$K_t = \beta(C_t - C_{t-1}) + (1-\beta)(K_{t-1}) \tag{8}$$

Where;

C_t : Exponentially smoothed series, at time t
K_t : Trend of the series, at time t
Z_t : Actual data value, at time t
α : Parameter for the level, $0 < \alpha < 1$
β : Parameter for the trend, $0 < \beta < 1$
$\tilde{Z}_{t+h}$: Fitted value or forecast value, at time $t + h$

ARIMA model

Box-Jenkins methodology is synonymous with the general Autoregressive Integrated Moving Average (ARIMA) modeling. This method was introduced by George E. P. Box and Gwilym M. Jenkins in 1976. ARIMA modeling is commonly used to analyse data series, forecasting and controlling.

The general term of this model is ARIMA (*p*,*d*,*q*), where *p* represented the order of autoregressive (AR) process, *d* represented the number of time the variable X_t needs to use differencing approach in order to achieve stationary and *q* represented the order of moving average (MA) process (Lazim, 2016). The ARIMA model is defined as follows:

$$X_t = \varnothing_1 X_{t-1} + \ldots + \varnothing_p X_{t-p} + \mathrm{a}_t - \theta_1 \mathrm{a}_{t-1} - \ldots - \theta_q \mathrm{a}_{t-q} \tag{9}$$

Where;

$\varnothing'_S$: The autoregressive parameters to be estimated
θ'_S : The moving average parameters to be estimated
X'_S : The original series
a'_S : A series of unknown random errors which are assumed to follow the normal probability distribution

The backshift operator was used by Box-Jenkins to make writing these models easier. Backshift operator, B, has the effect of changing time period t to time period t-1. Therefore, $BX_t = X_{t-1}$ *and* $B^2 X_t = X_{t-2}$. By using this backshift notation, the model may also be written as follows:

$$\left(1 - \varnothing_1 B - \ldots - \varnothing_p B^p\right) X_t = \left(1 - \theta_1 B - \ldots - \theta_q B^q\right) \mathrm{a}_t \tag{10}$$

$$\varnothing_p(B) X_t = \theta_q(B) \mathrm{a}_t \tag{11}$$

Where;

$$\varnothing_p(B) = (1 - \varnothing_1 B - \ldots - \varnothing_p B^p) \tag{12}$$

and

$$\theta_q(B) = (1 - \theta_1 B - \ldots - \theta_q B^q) \tag{13}$$

The above formula shows that the operations of $\varnothing_p(B)$ and θ'_S are polynomials in B of orders p and q respectively. One of the advantages of writing this model in a particular way is that it could be seen the reason why a number of models may be equivalent. Besides, the plotting of ACF and PACF will be used until the data under stationary conditions are met to determine an appropriate form of ARIMA models.

Forecasting Performance of Measures

The accuracy of the models in this research were assessed using the Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Geometric Root Mean Squared Error (GRMSE) as given in equation (14), (15), (16) and (17) respectively.

$$\mathrm{MSE} = \frac{1}{n}\sum_{t}^{n}(x_t - \hat{x}_t)^2 \tag{14}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t}^{n}(x_t - \hat{x}_t)^2} \tag{15}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{t}^{n}\left|\left(\frac{x_t - \hat{x}_t}{x_t}\right)\times 100\right| \tag{16}$$

$$\mathrm{GRMSE} = \left(\prod_{t}^{n} e_{itM}^2\right)^{\frac{1}{2n}} \tag{17}$$

Where;

x_t : Actual observed values

$\hat{x}_t$: Predicted values

n : The number of predicted values

e_{itM}^2 : Error made in estimation period, *t* for series *i* using method M

Result and Discussion

The findings of this study primarily have shown the summary of Portmanteau test which was used to identify the best ARIMA models in forecasting the rate of unemployment. **Table 1** shows that ARIMA (2, 1, 3) was found to be the best model to be utilised as compared to other models since the value of calculated chi-squared was less than the tabulated data at degrees of freedom 6 and 5% of significant level. Besides, ARIMA (2, 1, 3) was shown to have the smallest value of MSE, RMSE and MAPE, whereas the value of GRMSE was the second smallest compared to other models.

Table 1 Portmanteau Test Summary

Statistics	**Model**			
	ARIMA (1, 1, 1)	ARIMA (1, 1, 2)	ARIMA (2, 1, 3)	ARIMA (3, 1, 0)
Calculated Q	20.77	49.02	10.90	18.98
Degree of Freedom	9	8	6	8
Tabulated Q	16.919	15.507	12.592	15.507
Decision (5% sig. level)	Reject H_0	Reject H_0	Accept H_0	Reject H_0
MSE	0.01919	0.02303	0.01857	0.02310
RMSE	0.1385	0.1518	0.1363	0.1520
MAPE	3.0950	3.1515	2.9976	3.2457
GRMSE	0.05456	0.04963	0.05131	0.05610

The comparison of the forecast accuracy between Double Exponential, Holt's and ARIMA (2, 1, 3) were evaluated based on four indicator performances of error measure. **Table 2** shows the value of MSE, RMSE, MAPE and GRMSE for Double Exponential Smoothing, Holt's and ARIMA (2, 1, 3) model in forecasting the unemployment rate in Malaysia. Based on the table shown, ARIMA (2, 1, 3) model has the smallest value in almost all of the error measures. Therefore, ARIMA (2, 1, 3) emerged as the best model in forecasting the rate of unemployment.

Table 2 The MSE, RMSE, MAPE and GRMSE value for models

Model	**MSE**	**RMSE**	**MAPE**	**GRMSE**
Double Exponential Smoothing	0.0195	0.1396	3.1185	0
Holt's	0.0190	0.1380	3.1124	0
ARIMA (2,1,3)	0.01857	0.1363	2.9976	0.05131

The forecast value of unemployment rate in Malaysia in year 2019 by using ARIMA (2,1,3) model are represented in **Table 3** and **Figure 2**. Based on **Figure 2,** the square red line in the graph represented the forecast value of the unemployment rate in year 2019. It shows that the trend line of forecasting is low fluctuation where the values were gradually increasing and decreasing from January 2019 until December 2019.

Table 3 Forecast Value for Unemployment Rate

Year	Months	Forecast Value
2019	January	3.35451
	February	3.32932
	March	3.36028
	April	3.33523
	May	3.36622
	June	3.34127
	July	3.37224
	August	3.34736
	September	3.37831
	October	3.35349
	November	3.38441
	December	3.35964

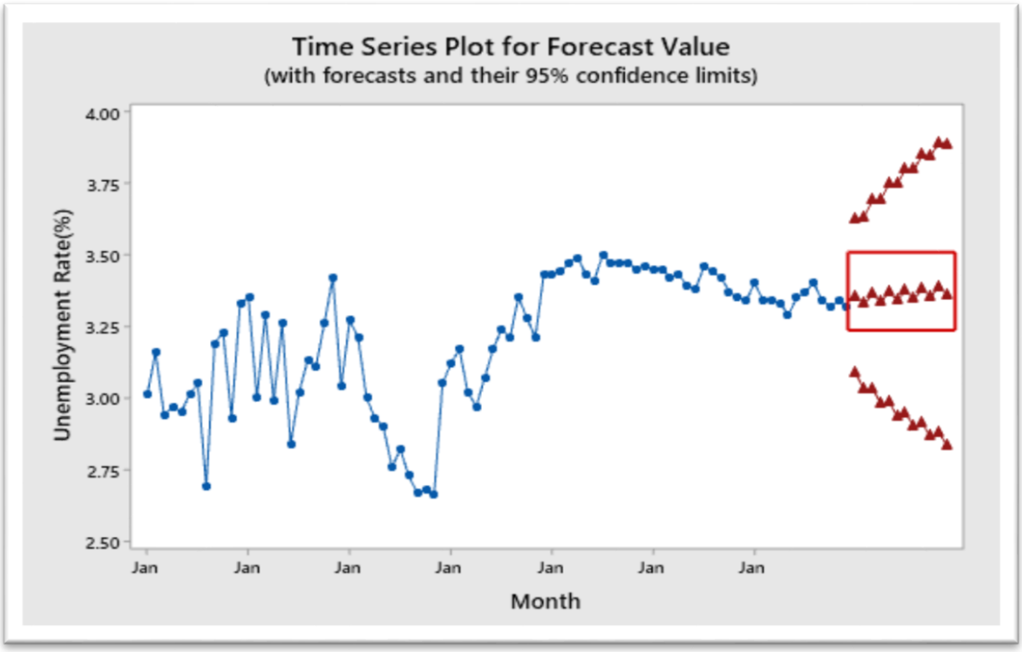

Figure 2 Graph of Forecast Values of Unemployment Rate in Malaysia using ARIMA (2, 1, 3)

Conclusion

The focus of this study is to determine the best and suitable model in forecasting the unemployment rate in Malaysia. Only two types of quantitative forecasting techniques which are smoothing methods (Double Exponential Smoothing and Holt's) and Box-Jenkins model had been used. It is found that ARIMA (2, 1, 3) in Box-Jenkins was the best model to forecast the unemployment rate in Malaysia as compared to smoothing methods. It was based on four indicator performance of error measures which are Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Geometric Root Mean Squared Error (GRMSE). The smallest value of error measures determined the best model. The last purpose of this study is to forecast the unemployment rate in Malaysia for a year ahead which is from January 2019 to December 2019 based on the selected best model. The forecasting was done by using ARIMA (2, 1, 3) and the results showed that the trend line of forecast values were gradually increasing and decreasing throughout the year of 2019. It is proven that there was an increment in the unemployment rate in 2019. However, the future value of the unemployment rate in Malaysia is still stable and not considered as a critical problem. Therefore, the signal of future unemployment rate is necessary for policy and decision maker to strategise the government expenditure and income. For future research, it is recommended that the seasonal nature of the unemployment rate in Malaysia can be modelled based on the nonlinear framework.

Acknowledgement

The author would like to thank Universiti Teknologi MARA Negeri Sembilan for the facilities involved in making this research a success.

Conflict of Interests

Author hereby declares that there is no conflict of interests with any organisation or financial body for supporting this research.

References

Asif, K. (2013). Factors Effecting Unemployment: A Cross Country Analysis. *International Journal of Academic Research in Business and Social Sciences, 3*(1).

Assis, K., Amran, A., & Remali, Y. (2010). Forecasting Cocoa Bean Prices Using Univariate Time Series Models. *Journal of Arts Science & Commerce, 1*(1), 71-80.

Department of Statistics Malaysia. Malaysia Unemployment Rate 1985-2017. Retrieved from https://tradingeconomics.com/malaysia/unemployment-rate. (5th March 2019).

Dobre, I., & Alexandru, A. A. (2008). Modelling Unemployment Rate Using Box-Jenkins Procedure. *3*(2).

Dumičić, K., Časni, A. Č., & Žmuk, B. (2015). Forecasting Unemployment Rate in Selected European Countries Using Smoothing Methods. *World Academy of Science, Engineering and Technology International Journal of Economics and Management Engineering, 9*(4).

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting, 20*(1), 5-10.

Lazim, M. A. (2016). *Introductory Business Forecasting a Practical Approach 3rd Edition. UiTM Press*.

Nor, M. E., Saharan, S., Lin, L. S., Salleh, R. M., & Asrah, N. M. (2018). Forecasting of Unemployment Rate in Malaysia Using Exponential Smoothing Methods. *International Journal of Engineering & Technology, 7*(4.30), 451-453.

Ramli, S. F., Firdaus, M., Uzair, H., Khairi, M., & Zharif, A. (2018). Prediction of the Unemployment Rate in Malaysia. *International Journal of Modern Trends in Social Sciences, 1*(4), 38-44.

Wheelwright, S., Makridakis, S., & Hyndman, R. J. (1998). *Forecasting: methods and applications*: John Wiley & Sons.