Enhanced Knowledge in Sciences and Technology Vol. 2 No. 2 (2022) 011-020
© Universiti Tun Hussein Onn Malaysia Publisher's Office

EKST

Homepage: http://publisher.uthm.edu.my/periodicals/index.php/ekst
e-ISSN : 2773-6385

Forecasting Analysis of Unemployment Rate in Malaysia based on the Naive, Holt's Linear Trend and Box-Jenkins Models

Chia Xin Yi[1], Kamil Khalid[2]*

[1]Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology
Universiti Tun Hussein Onn Malaysia, Pagoh Edu Hub, 84600, Johor, MALAYSIA.

*Corresponding Author Designation

DOI: https://doi.org/10.30880/ekst.2022.02.02.002
Received 02 January 2022; Accepted 03 February 2022; Available online 23 November 2022

Abstract: Unemployment has become one of the most vital challenges to the economy in most of the developed and developing countries, along with the socio-economic problem. Generally, the unemployment rate is the key element to measure whether a country is doing a good job of achieving productive employment or not. The previous studies mainly focused on forecasting quarterly and yearly unemployment rates by using Simple Exponential Smoothing, Holt's Linear Trend and ARIMA model. Yet, there are not many studies that focus on forecasting the monthly unemployment rate in Malaysia. Consequently, this study aims to compare the best model among the Naïve model, Holt's Linear trend model and the Box-Jenkins model for forecasting the monthly unemployment rate in Malaysia. The results revealed that the Naïve model was the best model with the lowest error rates of Mean Square Error (MSE), Mean Absolute Deviation (MAD) and Mean Absolute Percentage Error (MAPE). The forecast value for 3 months ahead unemployment rate was found to be 3.3%.

Keywords: Unemployment Rate, Forecasting, Box-Jenkins, Holt's Linear Trend, Naïve model

1. Introduction

Unemployment has become one of the most vital challenges to the economy in most of the developed and developing countries, along with the socio-economic problem. It is also a main macroeconomic indicator that indicates the state of a country's equilibrium and serves as a barrier to social growth [1]. Generally, the unemployment rate is the key element to measure whether a country is doing a good job of achieving productive employment or not. The unemployment rate is not always constant along the time due to many macroeconomic variables affecting it. It is necessary to keep exploring a new and accurate forecasting model over time and time to forecast the monthly unemployment rate. Forecasting is significant in a variety of fields of concern such as finance and

*Corresponding author: kamil@uthm.edu.my

accounting, economics, marketing and business in predicting the direction of future trends. Time series is one of the quantitative forecasting methods which is based on past data, make use of statistical analysis and forecasting model. Forecasting the unemployment rate is critical for policymakers to plan and strategize before time to avoid the persistent rise in unemployment levels.

The previous study mainly focused on quarterly and yearly data of forecasting the unemployment rate. There are only a few studies that focus on forecasting the monthly unemployment rate. Yet, there are not many studies that focus on using the ARIMA model to forecast the monthly unemployment rate in Malaysia. According to [1], many researchers found a disadvantage of the ARIMA model because they ignore the inclusion of explanatory variables and conduct the forecast solely on previous values of the dependent variable along with the past and present moving average terms.

A study done by [2] implied that Simple Exponential Smoothing was the ideal model to forecast the overall monthly unemployment rate in the year 2016. The Naïve method was one of the forecasting methods applied by [3] in forecasting the unemployment rate in Poland. Naïve method performs better when there is no drastic change in the series of forecast variables. Naïve method is widely used as the benchmark method for measuring forecast accuracy [4]. One of the researchers suggested Holt's linear exponential smoothing method to forecast the monthly unemployment rate in Romania [5]. This was due to the exponential smoothing model reacting more quickly to the change in the patterns of data as compared to others. ARIMA model was studied by [6] in forecasting the yearly unemployment rate in South Sulawesi and the result showed ARIMA (1,2,1) to be a suitable forecasting model due to smaller MSE.

The objective of this study is to build the Naïve model, Holt's Linear Trend model and Box-Jenkins model for modelling the monthly unemployment rate in Malaysia. Besides, this study aims to compare the performance of all the forecasting models based on MSE, MAD and MAPE. This study also targeted to forecast 3 months ahead on unemployment rate in Malaysia using the best model chosen.

2. Methodology

The materials and methods section, otherwise known as methodology, describes all the necessary information that is required to obtain the results of the study.

2.1 Research Framework

This process started with data collection. The data collected were the monthly unemployment rate from January 2010 to December 2019 which was sourced from the Department of Statistics Malaysia (DOSM) website. Data analysis was firstly started by visualizing the plotted graph of unemployment rate data to observe any consistent pattern. The graph showed there was a trend and no seasonality in the data. Then, non-seasonal models which were the Naïve model, Holt's Linear Trend and Bos-Jenkins model were implemented in this study. The best forecasting model was selected based on smaller error measures such as MAD, MAPE and MSE and it would be used to forecast 3-months ahead on the unemployment rate in Malaysia.

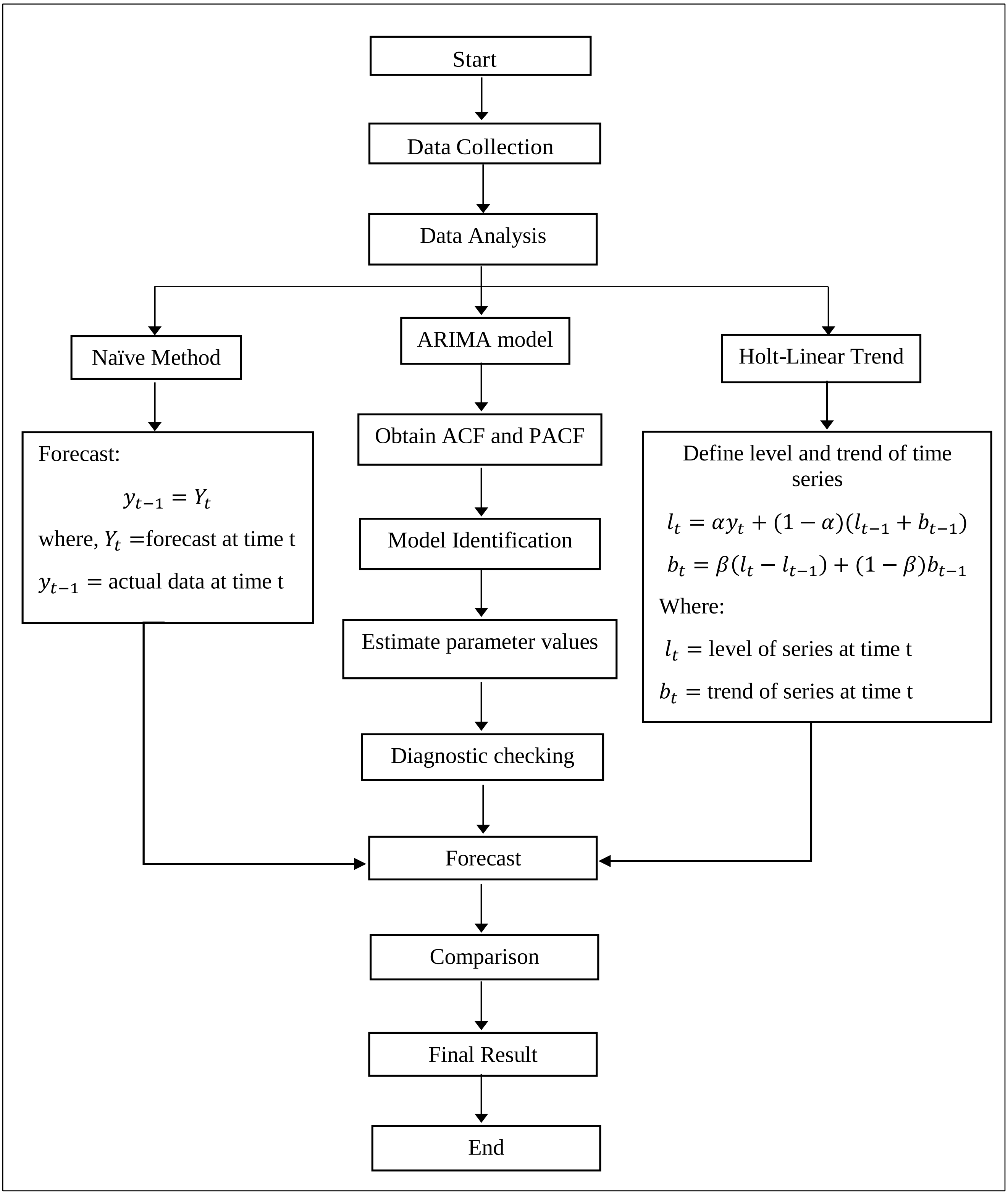

Figure 1: Framework for proposed forecasting time series process

2.2 Mann Kendall Trend test

Mann Kendall Trend test is a non-parametric test, which is also called Kendall's tau test. It is a test that is used to indicate whether the data sets consist of a trend pattern or not. This test is commonly used since it does not need any assumptions about the data being evaluated. The Mann Kendall test is based on the null hypothesis, which states that no trend exists and this is verified against the alternative hypothesis, which assumes that there is a trend pattern [7]. Mann Kendall Trend, S is calculated by using Equation 1.

$$S = \sum_{k=1}^{n-1} \sum_{j=k=1}^{n} sgn(x_j - x_k) \qquad Eq. 1$$

where x_j and x_k are sequential data value and *j* greater than *k*, *n* is the length of dataset

$$sgn(x_j - x_k) = \begin{cases} +1; & if(x_j - x_k) > 0 \\ 0; & if(x_j - x_k) = 0 \\ -1; & if(x_j - x_k) < 0 \end{cases}$$

2.3 Naïve Method

The naïve method is a technique that used previous periods of data to forecast for the next period, without any adjustments to identify casual variables. It is only used to compare forecasts with more advanced forecasting techniques. The equation of the naïve method is as below.

$$y_{t-1} = Y_t \qquad Eq. 2$$

where Y_t denoted as forecast at time t and y_{t-1} denoted as actual data at time t−1. It is often used as a benchmark for comparing more complicated methods due to its simplicity. This method can be quite well in future forecasting of the financial and stock market. This can be explained when a time series is a random walk, the naïve forecasts are optimal.

2.4 Holt's Linear Trend Method

Holt's Linear Trend is also known as double exponential smoothing, which is the extension of simple exponential smoothing. It allows the forecasting of data with trend patterns. This approach included a forecast equation and two equations for smoothing as follow:

Forecast equation: $\hat{y}_{t+h|t} = \ell_t + hb_t \qquad Eq. 3$

Level equation: $\ell_t = \alpha_{y_t +}(1-\alpha)(\ell_{t-1} + b_{t-1}) \qquad Eq. 4$

Trend equation: $b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \qquad Eq. 5$

where:

l_t denotes as estimate of level of the series at time t
b_t denotes as estimate of trend of the series at time t
α denotes as smoothing parameter for level
β denotes as smoothing parameter for trend

2.5 Box-Jenkins Method

Box-Jenkins are also known as Auto-Regressive Integrated Moving Average (ARIMA) model. ARIMA model is an approach that is the combination of autoregressive (AR) model and moving average (MA) model. The weighted moving average over previous observations is known as autoregressive (AR). Integrated (I) is a linear or polynomial pattern while Moving Average (MA) is a weighted moving average over past errors. It is also a model for non-seasonal series. ARIMA (*p*, *d*, *q*) model is formed with the combination of three model parameters AR(*p*), I(*d*) and MA(*q*) where *p*, *d*, *q* denotes as the order of autocorrelation, degree of differencing involved and order of moving averages. The formula of the ARIMA model can be expressed in Equation 6:

$$\varphi_{\mathrm{p}}(B)(1-B)^d y_t = \theta_q(B)e_t \qquad Eq. 6$$

where:

$\varphi_{\mathrm{p}}(B)$ is a stationary AR operator
$\theta_q(B)$ is an invertible MA operator
$(1-B)^d y_t$ is d^{th} difference and et is residual value at period t.

2.6 Performance Measures

2.6.1 Mean Absolute Deviation (MAD)

MAD is used to evaluate forecasting by adding up the absolute errors. The forecasted accuracy is measured with an average suspected error using this method. Equation 7 shows the formula for calculating the MAD value.

$$MAD = \frac{\sum_{t=1}^{n} |y_t - F_t|}{n} \qquad Eq. 7$$

where:

y_t= Time series value at time t
F_t= Forecast value at time t
n= Total number of periods

The lower the value of MAD relative to the magnitude of data, the more accurate the forecast.

2.6.2 Mean Square Error (MSE)

Mean squared error (MSE) is defined as the square difference between actual and forecast values. Lower MSE value indicated greater forecasting accuracy [8] as shown in Equation 8.

$$MSE = \frac{\sum_{t=1}^{n} (y_t - F_t)^2}{n} \qquad Eq. 8$$

2.6.3 Mean Absolute Percentage Error (MAPE)

MAPE is the most common measure used to compare forecast performances between datasets [9]. It measures the accuracy as a percentage and can be calculated as the average absolute percentage of absolute errors for each period minus actual observation divided by actual observation. The measure of MAPE that was being used stated in Equation 9:

$$MAPE = \frac{\sum_{t=1}^{n} \left|\frac{y_t - F_t}{n}\right| \times 100}{n}, y_i \neq 0 \qquad Eq. 9$$

It is scale-independent as it can use to compare forecasts for series that are on different scales [10]. It works best when there are no extreme values. However, when the original value is small, MAPE had been criticized for its asymmetry and instability.

3. Results and Discussion

The data used in this analysis are the monthly unemployment rate in Malaysia. The data were sourced from Department of Statistics Malaysia (DOSM) website from January 2010 to December 2019. The descriptive statistics of monthly unemployment rate including mean, standard deviation, kurtosis, skewness, coefficient of variation, minimum and maximum values of the unemployment rate are presented in Table 1.

Table 1: Descriptive Statistics

Monthly Unemployment Rate	
Mean	3.2025
Standard Deviation	0.2169
Kurtosis	-0.4500
Skewness	-0.5400
Coefficient of variation	6.7700
Minimum	2.7000
Maximum	3.6000
Count	119.0000

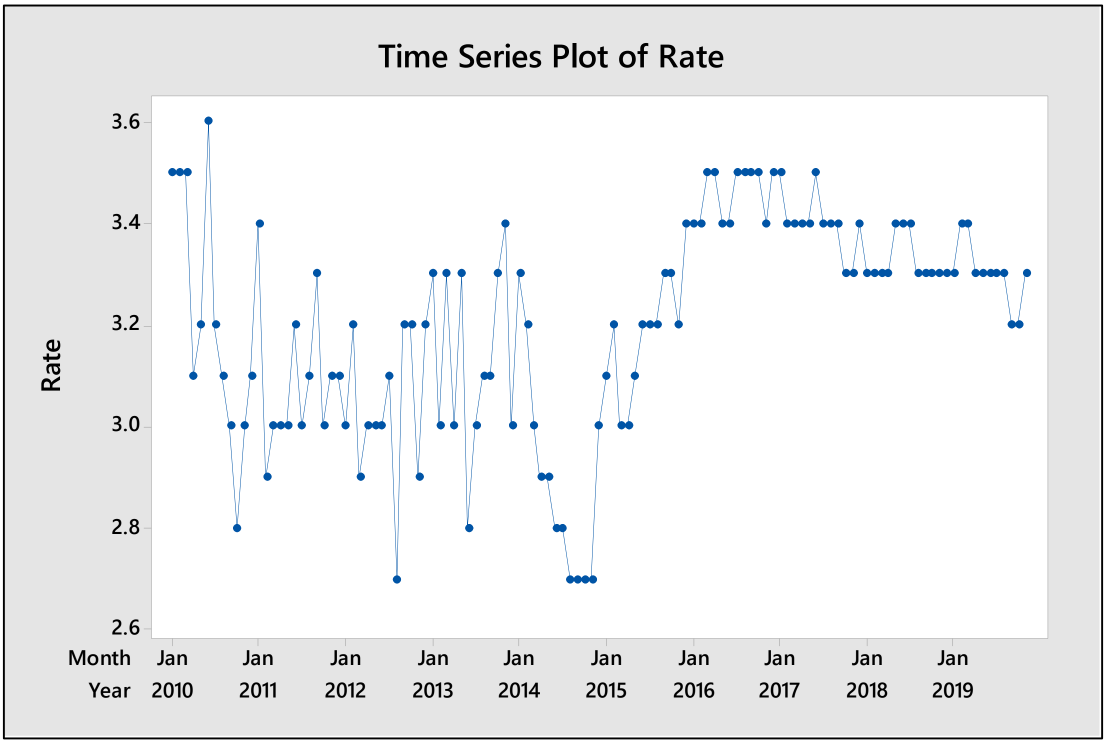

Figure 2: Time Series Plot for Unemployment Rate in Malaysia

Figure 2 shows the monthly unemployment rate in Malaysia from January 2010 to December 2019. A visual examination of the time series plot implies a significant fluctuation in the unemployment rate in which the unemployment rate has increased and decreased over time. This suggests that unemployment rate data are non-stationary.

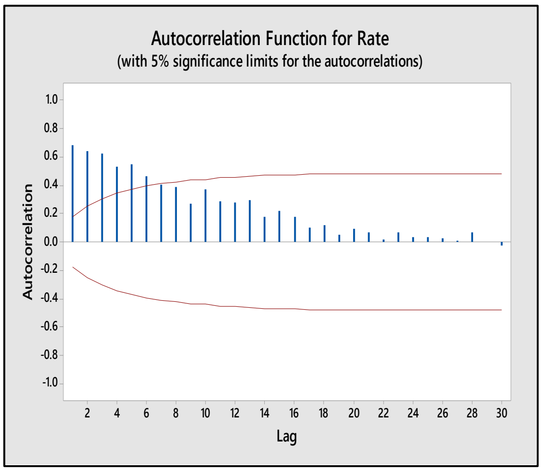

Figure 3a: ACF plot for Unemployment Rate in Malaysia

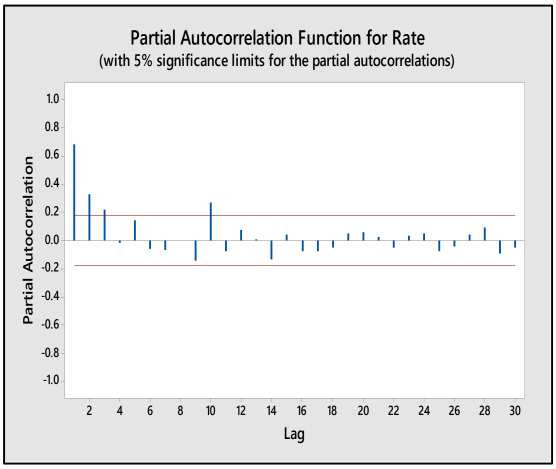

Figure 3b: PACF plot for Unemployment Rate in Malaysia

Moreover, the non-stationary behaviour series can be proved by autocorrelation function (ACF) and partial autocorrelation function (PACF) plots in Figure 3a and Figure 3b. The ACF of non-stationary data decreases slowly, showing a non-stationary time series. Hence, differencing must be taken in developing a stationary ARIMA model.

```
> TS = ts(Data$Rate,frequency=12,start=c(2010,1))
> library(Kendall)
> MK=MannKendall(TS)
> summary(MK)
Score =  1623 , Var(Score) = 134975
denominator =  5298.785
tau = 0.306, 2-sided pvalue =1.0133e-05
```

Figure 4: Mann Kendall Trend test

H_0: There is no trend in the data.

H_1: There is a trend in the data.

Mann-Kendall Trend test is performed to whether there is a trend in data. Figure 4 shows the test statistic is 0.306 and the two-sided *p*-value is 1.0133e-05. Since the p-value is less than 0.05, there is sufficient evidence to reject the null hypothesis and conclude that the trend is present in this data.

Table 2: Comparison of forecast accuracy for different models

Method	Training			Testing		
	MSE	MAD	MAPE	MSE	MAD	MAPE
Naive	0.031509 (3)	0.11887 (2)	3.8589% (2)	0.00333 (1)	0.03333 (1)	0.4976% (1)
Holt's Linear Trend	0.02935 (1)	0.12008 (3)	3.8376% (3)	0.004274 (3)	0.047078 (3)	1.4332% (3)
Box-Jenkins	0.031179 (2)	0.11537 (1)	3.6984% (1)	0.00341 (2)	0.03916 (2)	1.1903% (2)

The ideal forecasting method was selected based on the accuracy of the testing data set whereby the method that had the smallest forecasting error was the best forecasting method. In this study, the accuracy of the forecasting was evaluated by using mean absolute deviation (MAD), mean square error (MSE) and mean absolute percentage error (MAPE) in the testing part. Three forecasting methods were being compared by separating the data into training and testing data. The calculation of MSE, MAD and MAPE for those methods were shown in Table 2. From the table, the Naive method had the lowest MSE, MAD and MAPE values which were 0.00333, 0.0333 and 0.4976% respectively for the testing set. It represented the best forecasting method as high accuracy in forecasting the monthly unemployment rate. Box-Jenkins also performed very well in the training part since it had the highest values of MSE, MAD and MAPE which were 0.0312, 0.1154 and 3.6984% compared to other methods. Although the MSE value of Holt's Linear Trend is smaller than the other two methods in the training set, it shows the highest values in the rest of the performance error values in both the training and testing sets. This method is inaccurate in forecasting the monthly unemployment rate.

Table 3: Forecast values and actual values for 3 months ahead

Period	Forecast (%)	Actual Value (%)
January 2020	3.3	3.2
February 2020	3.3	3.3
March 2020	3.3	3.9

The results indicate that the Naïve method is the best method to forecast the unemployment rate for the next 3 months. Forecast values for 3-months ahead on unemployment rate are 3.3% as shown in Table 3. The forecasted values in January 2020 and February 2020 do not show significantly different from actual values. However, in March 2020, there is an increase in the unemployment rate which is 3.9%. The negative impact of MCO due to the Covid-19 pandemic has caused the high unemployment rate in March 2020. Further investigation can be done to identify the forecasting model for the unemployment rate in Malaysia during the MCO period.

4. Conclusion

The study aims to select the best model in forecasting the monthly unemployment rate in Malaysia. This study depicts the unemployment rate in Malaysia from the year 2010 to year 2019. Visual inspection of data patterns of unemployment rate showed that there was a trend pattern in the data but significant seasonal or horizontal patterns were not found in the data. Several approaches have been conducted to analyze the time series data. Since the data was identified to have a trend pattern, Naïve method, Holt-Linear Trend and ARIMA forecasting methods were selected to compute the time series analysis of unemployment rate in Malaysia. Some performance metrics such as Mean Square Error (MSE), Mean Absolute Deviation (MAD) and Mean Absolute Percentage Error (MAPE) were used to access the forecast accuracy of models for both training data and testing data. Several conclusions can be drawn from the dataset. Based on the findings, the Naïve method was chosen as the most desirable forecasting method as it achieved the lowest valued for those three performance measures (MSE, MAD and MAPE). The unemployment rate for the next 3 months was found to be 3.3%. Compared to the actual value, the unemployment rate in Malaysia is stable in January 2020 and February 2020, but due to the unexpected condition which is Covid-19, the unemployment rate in March 2020 becomes higher. Further study should be done to examine an appropriate model used to forecast future values of the unemployment rate in MCO period. Future studies could focus on describing the seasonal nature of unemployment rates in a nonlinear framework and modelling the unemployment rate using a structural VAR model.

Acknowledgement

I would like to express my thanks to my supervisor, En. Kamil Bin Khalid for his continuous support throughout my project. I would also like to thank the Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia for its support. Lastly, I would like to express my gratitude to my friends and family for their patience, encouragement and wholehearted support as I complete my project.

References

[1] Nikolaos, D. and Paraskevi, K., “Forecasting Unemployment Rates in USA Using Box-Jenkins Methodology,” International Journal of Economics and Financial Issues, vol. 8, no. 1, pp. 9-20, Jan. 2018

[2] Maria, E. et al., “Forecasting of Unemployment Rate in Malaysia Using Exponential Smoothing Methods,” International Journal of Engineering & Technology, vol. 7, no. 4.30, pp. 451-453, 2018, doi: 10.14419/ijet.v7i4.30.22365

[3] Michal, G. & Tomasz, R. (2021), “Forecasting the Unemployment Rate: Application of Selected Prediction Methods,” European Research Studies Journal, 24(3), pp. 985-1000, doi:10.35808/ersj/2396

[4] Thabo, M. et al., “Using Feature Engineering Techniques to Improve Multivariate Model Forecasting of the South African Unemployment Rate,” Proc. of the International Conference on Electrical, Computer and Energy Technologies (ICECET), December 9-10, 2021, Cape Town-South Africa

[5] Voineagu V. et al., “Forecasting Monthly Unemployment by Econometric Smoothing Techniques,” Economic computation and economic cybernetics studies and research, vol. 3, no. 3, pp. 255-267, 2012

[6] Didiharyono, D. & Syukri, M. (2020). Forecasting with ARIMA model in Anticipating Open Unemployment Rates in South Sulawesi. *International Journal of Scientific & Technology Research, 9(3),* pp. 3838-3841

[7] Gadedjisso-Tossou, A. et al., “Rainfall and Temperature Trend Analysis by Mann–Kendall Test and Significance for Rainfed Cereal Yields in Northern Togo,” Sci 2021, 3, 17. https://doi.org/10.3390/sci3010017

[8] Ayundyah, K. and Arum, H. P., Forecasting Bank Indonesia Rate based on Fuzzy Time Series with Higher Forecast Accuracy Rate. Proceedings International Conference on Mathematics, Sciences and Education, University of Mataram 2015, Lombok Island, Indonesia, November 4-5, 2015

[9] Sungil, K. & Heeyoung, K. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting, 32*, pp. 669-679, doi: 10.1016/J.IJFORECAST.2015.12.003

[10] Estrada, M., R. et al. (2020). Evaluation of Several Error Measures Applied to the Sales Forecast System of Chemicals Supply Enterprises. *International Journal of Business Administration*, *11(4)*, pp.39-51, doi:10.5430/ijba.v11n4p39