ENHANCED KNOWLEDGE IN SCIENCES AND TECHNOLOGY
e-ISSN: 2773-6385

EKST

Vol. 5 No. 1 (2025) 505-514
https://publisher.uthm.edu.my/periodicals/index.php/ekst

Time Series Analysis for Number of Monthly Unemployment in Malaysia

Nurfara Hanim Juhaimi[1], Kamil Khalid[1]*

[1] *Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, UTHM Kampus Cawangan Pagoh, Hub Pendidikan Tinggi Pagoh KM 1, Jalan Panchor, 84600 Pagoh, Muar, Johor, MALAYSIA.*

*Corresponding Author: kamil@uthm.edu.my
DOI: https://doi.org/10.30880/ekst.2025.05.02.058

Article Info

Received: 30 December 2024
Accepted: 17 January 2025
Available online: 19 December 2025

Keywords

Unemployment Rate, Forecasting, Holt's Linear, Naive, Box-Jenkins, Accuracy Measures

Abstract

This study explores challenges in forecasting Malaysia's unemployment rate, focusing on rising trends in recent years. It aims to develop models known as Box-Jenkins (ARIMA), Holt's Linear, and Naïve methods were used, the best model was identified by comparing Mean Absolute Error, Mean Absolute Percentage Error, and Root Mean Square Error values, and unemployment rates were forecasted with the most accurate model. The dataset included 97 observations was collected from 2016 to 2023 based on the Labor Force Survey (LFS). Statistical analysis revealed that the Naïve method performed poorly, Holt's Linear method suited linear trends, and the Box-Jenkins (ARIMA) model was the most reliable. During the training phase, the Box-Jenkins method showed the best performance, with the lowest MAE (6.7405), MAPE (0.1950), and RMSE (8.2968), indicating strong fit to the data. In the testing phase, the Naïve Method achieved the lowest errors, with MAE (0.6750), MAPE (0.0227), and RMSE (0.8026). However, the Box-Jenkins method followed with slightly higher errors, still showing good generalization. Therefore, the Box-Jenkins method will be chosen as the best method because it was more compatible to forecast for the long time period compared to Naïve method that only useful to forecast short time period. The use of larger datasets was recommended to improve forecasting accuracy, providing insights for policymakers and researchers in addressing unemployment challenges.

1. Introduction

Unemployment rates were regarded as more than just numbers; they were considered among the most critical indicators of economic health [1]. This issue had persisted for decades, and this study was conducted to investigate the factors that contributed to unemployment rates in Malaysia. Economic factors such as Gross Domestic Product (GDP) growth, inflation rates, and industrial composition could all have a significant impact [2]. A significant decrease in global economic growth was caused by the recent Coronavirus disease (COVID-19) pandemic. Malaysia's economic growth was significantly affected in 2020 as a result of the Movement Control Order (MCO) that was implemented in March 2020 to prevent the spread of COVID-19 [3]. The issues that occurred during this pandemic were found to have influenced the unemployment rate in Malaysia. As Malaysia's GDP growth decreased, the country's unemployment rate was observed to rise, potentially leading to a crisis.

Furthermore, labour market dynamics such as skill mismatches, wage levels, and job availability were identified as factors that could impact unemployment rates [4]. Social factors, including education levels, demographic changes, and cultural attitudes toward work and employment, were also recognized as significant variables. For instance, inequalities in educational achievement among different regions or demographic groups were found to influence their employment prospects [5]. Additionally, cultural standards regarding gender roles and expectations of employment stability were observed to affect labour force participation rates [6]. In addition, unemployment rates were influenced by policy-related factors such as government interventions, labor market regulations, and the efficiency of employment programs [7]. By examining these various factors, a comprehensive understanding of unemployment in Malaysia was developed, enabling policymakers to create targeted solutions to address this complex problem [8].

In this study, the model for unemployment rates was built by using Autoregressive Integrated Moving Average (ARIMA) model, Holt's Linear method and Naïve method, the best model for Malaysian unemployment rates was determined by analysing Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE) values and the unemployment rates in Malaysia was forecasted by using the best model.

2. Methodology

The dataset of Monthly Unemployment by Duration was collected from 2016 to 2023 based on the Labor Force Survey (LFS), which was designed to gather representative data on the labour force at both national and state levels. The dataset included 97 observations in total, with seven variables. The dataset presented the number of unemployed individuals categorized into two groups: active and inactive. In this study, the dataset was analysed using time series analysis and forecasting. Time series analysis was employed to identify seasonality, trends, and patterns within the dataset, while forecasting analysis was used to predict future unemployment levels. Three forecasting methods were utilized in this study: Naïve, Holt's Linear, and the Box-Jenkins method. The statistical software that was used for this analysis included Microsoft Excel and Minitab.

2.1 Naïve method

The Naïve forecast method was described as simple, relying solely on the previous year's actual value data to predict future outcomes [9]. For instance, the forecast for 1996 was based on the observed value from 1995. The Naïve forecast method was found to work best for time series that were stable or followed a random walk pattern. Although the Naïve method was simple and useful as a benchmark, it had limitations. It performed poorly when the time series data contained complex patterns, seasonality, or other underlying structures that more advanced forecasting models, such as ARIMA or exponential smoothing methods, were designed to detect. Furthermore, the method ignored any additional information or external factors that might have influenced future outcomes, concentrating solely on the most recent observation. The formula for the Naïve method as shown in equation (1),

$$Y_{t-1} = Y_t \tag{1}$$

where, Y_t indicates the forecast at time t and Y_{t-1} indicates the actual data at time t-1.

2.2 Box-Jenkins (ARIMA)

The Box-Jenkins methodology was a systematic and widely accepted method used to create Autoregressive Integrated Moving Average (ARIMA) models for time series forecasting. The Box-Jenkins methodology was a step-by-step approach used to create ARIMA models for time series forecasting. It included three steps which was model identification, parameter estimation, and model selection. The process began by examining the data for trends and patterns, checking for stationarity, and transforming the data if needed. Differencing was applied to make the series stationary. Tools like ACF and PACF were used to determine the appropriate values for p and q. Maximum likelihood estimation (MLE) was applied to calculate the best-fitting model. After validation with tests like the Ljung-Box test, the ARIMA model was adjusted if necessary and used to forecast future values.

A combination function of autoregressive (AR) and moving average (MA) models was known as an ARIMA model. ARIMA model was used to show stationary series where its general term is ARIMA *(p,d,q)* [10]. The p and q indicated the order of AR process and MA process while the d indicated the number of times the variable X_t required a differencing approach to achieve stationary results. The ARIMA *(p,d,q)* model could be written as equation (2),

$$\Phi(B)X_t = \Phi(B)(1-B)^d X_t = \theta(B)Z_t \tag{2}$$

where,

Z_t = forecast value period at time t

X_t = time series value at time t

B = backward shift operator

$\theta(z)$ = polynomials of degrees p

$\Phi(z)$ = polynomials of degrees q, where $\Phi(z) \neq 0$ for |z| ≤1

2.3 Holt's Linear Trend

Holt's Linear Trend (HLT) method, also known as Double Exponential Smoothing (DES), was used to obtain forecasts for time series displaying a trend pattern. The Holt's Linear method was proposed by Brown to address the variations between actual data and predicted values. This method was considered more efficient than other methods for modelling trends and levels in a time series, as it required less data and only one parameter, simplifying the process [11]. Holt's Linear Trend was represented by equation (3),

$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + T_{t-1})$$
$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1}$$
$$\hat{Y}_t = L_{t-1} + T_{t-1} \tag{3}$$

where:

L_t = level at time t

α = weight for the level, where $0 \langle \alpha \langle 1$

T_t = trend at time t

γ = weight for the trend, where $0 \langle \gamma \langle 1$

Y_t = data value at time t

$\hat{Y}_t$ = fitted value

2.4 Accuracy Measure

Accuracy measures are utilized as metrics to assess the performance of forecasting models by quantifying the difference between observed and predicted values. These metrics played an important role in evaluating how well a model performed and determining whether it could be trusted for future predictions. Different accuracy measures are deemed appropriate for various types of data and forecasting problems, guiding the selection and refinement of models. Some of the most commonly used accuracy measures included Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE). The equation of MAE, MAPE and RMSE are shown in equations (4), (5) and (6).

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \tag{4}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{6}$$

where:

$\hat{y}_i$ = actual value

y_i = predicted value

n = number of data points

3. Results and Discussion

The unemployment rate in Malaysia was analysed using three forecasting method which was Naive, Box-Jenkins (ARIMA), and Holt's Linear method. The forecast performance was evaluated to compare the accuracy of these methods for both the training and testing data using MAE, MAPE, and RMSE. The method with the lowest values for these measures was considered the most suitable for forecasting the unemployment rate in Malaysia.

3.1 Time Series Plot

A time series plot is defined as a statistical graph in which data points collected over time are illustrated, with time represented on the horizontal axis and the observed variable on the vertical axis. Trends, seasonal patterns, and variations are highlighted, making it easier for stationarity and autocorrelation to be analysed.

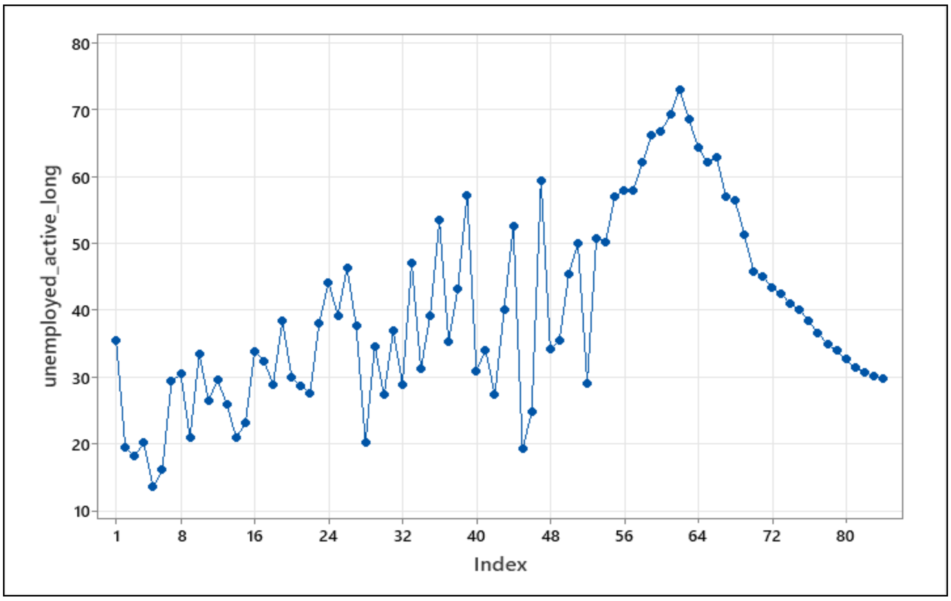

Fig. 1 Time series plot for the unemployment rates

Fig. 1 illustrates the time series plot has the pattern or the trend. Based on the time series plot, we can see that there is an increasing trend of the data and it shows no seasonality.

3.2 Naïve method

The basic principle of this method is straightforward, as it assumes that the forecast for the next period is identical to the most recent actual value from the previous period. Underlying trends, seasonality, or patterns in the data are not considered by the Naïve method. Instead, it is assumed that the future will mirror the past.

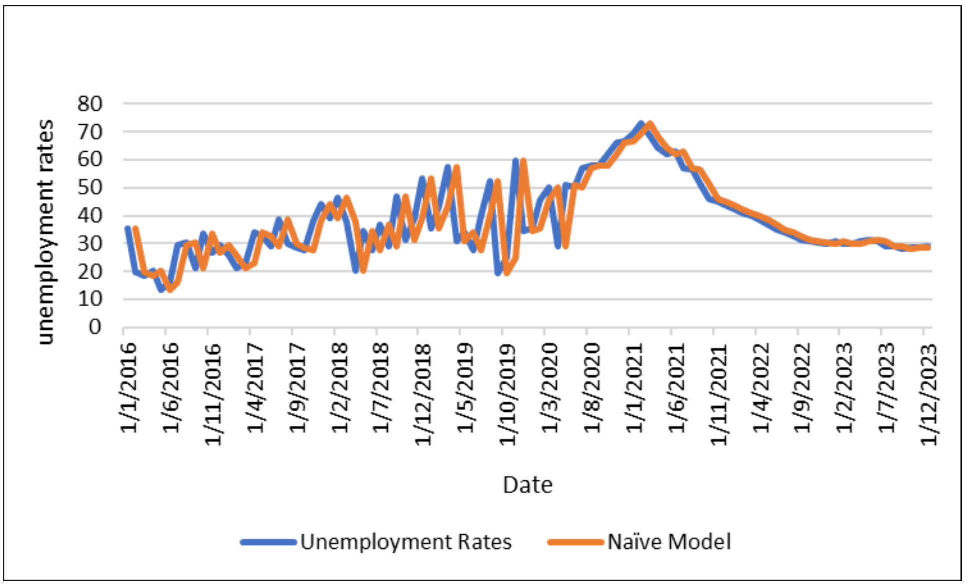

Fig. 2 Naïve model forecast for Unemployment Rates

Fig. 2 shows a time series plot of unemployment rates in Malaysia from January 2016 to December 2023. The blue line represents the actual unemployment rates, while the orange line represents the forecasts generated using the Naïve model. The Naïve model closely follows the fluctuations in unemployment rates, as it assumes the forecast for each period is equal to the actual value of the preceding period. The plot highlights various trends, including periods of stability and significant fluctuations, particularly around 2020 and 2021, which may coincide with external events like the COVID-19 pandemic. Toward the end of the time series, both lines converge, indicating a steady unemployment rate and consistent Naïve forecasts.

3.3 Box-Jenkins (ARIMA)

A combination function of autoregressive (AR) and moving average (MA) models is known as an ARIMA model. An ARIMA model is used to represent stationary series.

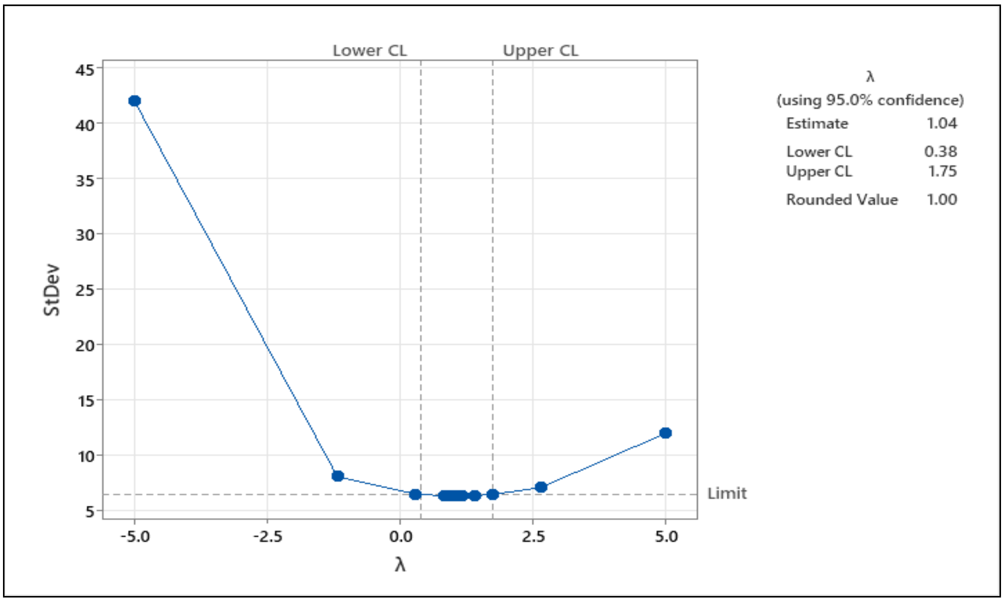

Fig. 3 Box-Cox Plot of Unemployment Rates

Fig. 3 illustrates the Box-Cox plot for forecasting the unemployment rate. The Box-Cox plot is used to assess the variance. Based on the plot, the 95% confidence interval for λ (0.38 to 1.75) includes 1, indicating that a transformation is not appropriate as the variance is constant.

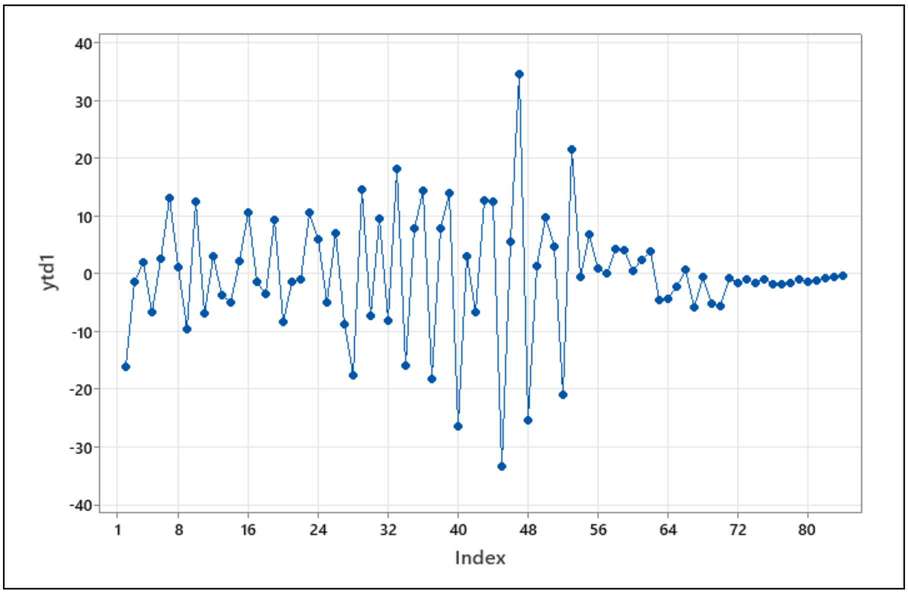

Fig. 4 Time series plot of unemployment rates after differencing

Based on the Fig. 4, the time series plot has a constant variance and mean. This indicates that the time series data has achieved stationarity, allowing us to move forward with estimating model parameters.

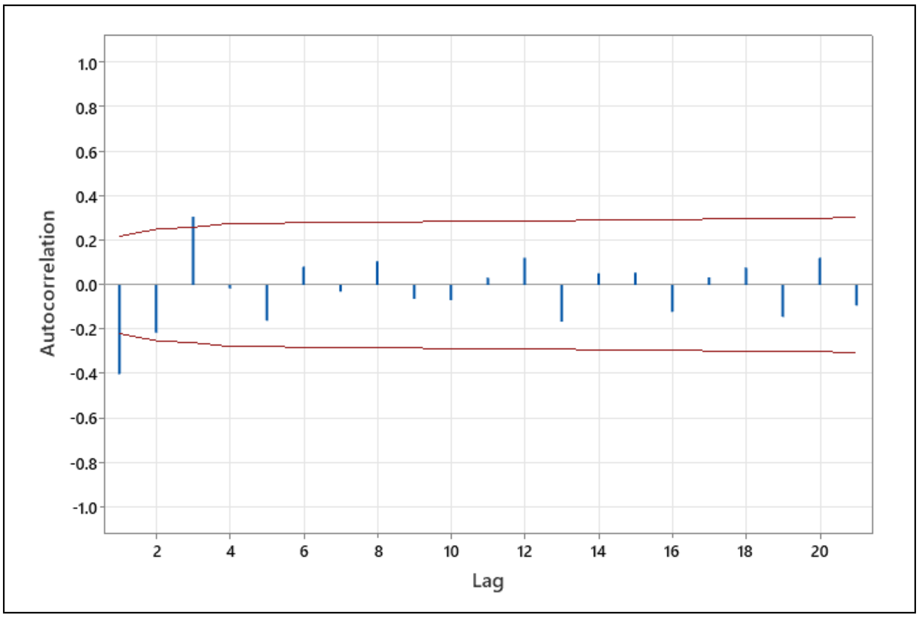

Fig. 5 ACF plot of unemployment rates after differencing

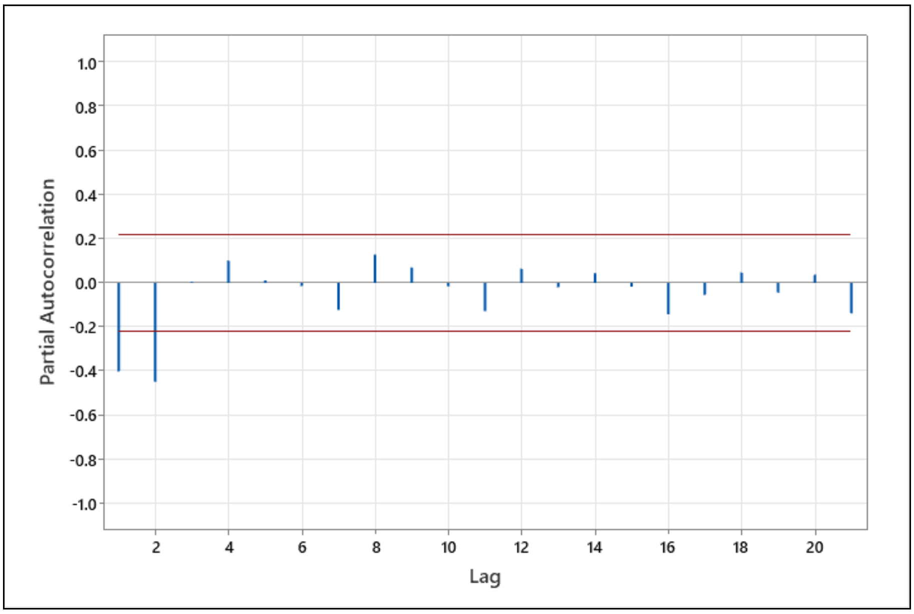

Fig. 6 PACF plot of unemployment rates after differencing

From Fig. 5, the ACF plot shows that it has cut off at lag 3. Meanwhile from Fig. 6, the PACF plot shows that it has cut off at lag 2.

Table 1 *P*-value and Mean Squared Error of possible models

Model	*P*-value of coefficient parameter	*P*-value of Ljung Box				MSE
		Lag 12	Lag 24	Lag 36	Lag 48	
ARIMA (1,1,0)	AR (1) : 0.000	0.001	0.014	0.109	0.128	91.1505
ARIMA (2,1,0)	AR (1) : 0.000 AR (2) : 0.000	0.840	0.837	0.964	0.993	70.5360
ARIMA (0,1,1)	MA (1) : 0.000	0.194	0.378	0.780	0.871	78.4743
ARIMA (0,1,2)	MA (1) : 0.000 MA (2) : 0.130	0.191	0.414	0.743	0.876	78.2650
ARIMA (0,1,3)	MA (1) : 0.000 MA (2) : 0.487 MA (3) : 0.027	0.629	0.761	0.945	0.980	74.3787

Table 1 compares ARIMA models based on their *p*-values, Ljung-Box test results at different lags, and Mean Squared Error (MSE). The ARIMA (2,1,0) model is found to perform the best, with the lowest MSE of 70.5360 and well-behaved residuals, as indicated by high *p*-values from the Ljung-Box test. Significant autocorrelation is observed in the residuals of the ARIMA (1,1,0) model but a higher MSE of 91.1505 is reported. The ARIMA (0,1,2,) and ARIMA (0,1,3) could not be choose as it did not fulfil the significant *p*-value of coefficient parameter. Overall, the ARIMA (2,1,0) model is considered the most efficient in fitting the data.

3.4 Holt's Linear Trend

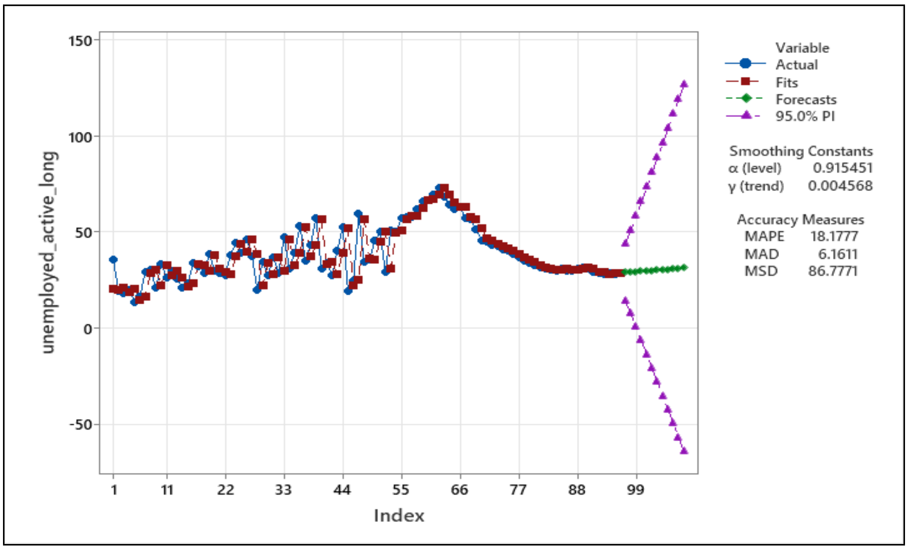

Fig. 7 Smoothing Plot of Unemployment rates

Fig. 7 presents a time series analysis of the unemployment rate using the Double Exponential Smoothing method, which accounts for both the level and trend of the data. The actual observed values are represented by the blue dots, while the fitted values are shown by the red line, illustrating how fluctuations are smoothed to reveal the underlying trend. The forecasted values are indicated by the green points, predicting future unemployment levels based on historical data. The 95% prediction intervals are represented by the purple lines, which widen as the forecast extends, indicating increased uncertainty in longer-term predictions. The smoothing constants (α=0.915, γ=0.004) suggest that significant weight is given to recent observations, with minimal adjustments made for the trend. Overall, the graph is used to effectively illustrate how historical patterns are captured and future values are projected, with uncertainty increasing over time.

3.5 Forecast Performance Evaluation

Table 2 Accuracy Measures of All methods with testing phase

Method	Training			Testing		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Naïve Method	7.1988	0.2164	10.3497	0.6750	0.0227	0.8026
Box Jenkins ARIMA (2,1,0)	6.7405	0.1950	8.2968	0.7144	0.0244	0.9610
Holt Linear	7.2293	0.2108	8.8047	1.3809	0.0473	1.5878

Table 2 compares the forecasting performance of three methods based on training and testing datasets using MAE, MAPE and RMSE. During the training phase, the Box-Jenkins method showed the best performance, with the lowest MAE (6.7405), MAPE (0.1950), and RMSE (8.2968), indicating strong fit to the data. In the testing phase, the Naïve Method achieved the lowest errors, with MAE (0.6750), MAPE (0.0227), and RMSE (0.8026). The Box-Jenkins method followed with slightly higher errors, still showing good generalization. Overall, the best balance of performance was shown by the Box-Jenkins method. Holt Linear Trend exhibited higher errors, making it less robust. Since the Box-Jenkins (ARIMA) model outperformed the Naïve and Holt's Linear models, the forecast for unemployment rates will be based on the ARIMA (2,1,0) model.

Table 3 Accuracy Measures of All methods

Method	Training			
	MAE	MAPE	RMSE	MSE
Naïve Method	6.3747	0.1919	9.6782	93.6674
Box Jenkins ARIMA (2,1,0)	5.9793	0.1686	7.7626	60.2578
Holt Linear	6.2458	0.1825	8.2592	68.2143

Table 3 compares the training performance of three forecasting methods using MAE, MAPE, RMSE, and MSE. Box Jenkins outperforms the others with the lowest errors MAE (5.9793), MAPE (0.1686), RMSE (7.7626), MSE (60.2578), indicating the highest accuracy. Holt Linear shows moderate performance, while the Naïve Method has the highest error values, making it the least effective. Therefore, Box Jenkins is the most reliable method to forecast unemployment rates for the long-term period.

Table 4 The forecast value for the unemployment rate using ARIMA (2,1,0)

Period	Forecast	Lower	Upper
97	30.3681	13.9036	46.8326
98	30.2251	12.5531	47.8971
99	30.0899	11.7971	48.3827
100	30.2403	9.1325	51.3481
101	30.2128	7.7117	52.7138
102	30.1581	6.6791	53.6371
103	30.2046	5.1758	55.2333
104	30.2022	3.9293	56.4751
105	30.1816	2.8671	57.4961
106	30.1953	1.7020	58.6886
107	30.1967	0.6106	59.7828
108	30.1893	-0.3952	60.7738

Table 4 presents the forecast values for the unemployment rate in Malaysia. Although the forecast values may appear similar, they actually differ in their four decimal places.

4. Conclusion

In conclusion, this study developed three forecasting models to predict Malaysian unemployment rates which are the ARIMA model, Holt's Linear method, and the Naïve method. This study proved ARIMA, is suitable for analysing time series with trends and seasonality while Holt's Linear method is effective for data showing consistent upward or downward movement and the Naïve method assumes future rates will reflect the most recent observed levels. Furthermore, the model performance was evaluated using MAE, MAPE, and RMSE. The ARIMA model outperformed the other two methods with the lowest error values, proving it to be the most reliable for forecasting Malaysian unemployment rates. The study successfully predicted future unemployment rates using the ARIMA model, demonstrating its reliability compared to the other methods. The forecasts provide valuable insights for researchers and policymakers to predict trends and create solutions, highlighting the importance of using complex models like ARIMA for accurate economic planning.

Several suggestions can be made to improve the study and enhance unemployment rate forecasting. First, additional models like Seasonal ARIMA (SARIMA) or advanced machine learning techniques such as Neural Network Autoregression (NNAR) and Support Vector Regression (SVR) could be incorporated to capture complex patterns and address non-linearities or seasonal fluctuations not considered in the current analysis.

Second, incorporating unemployment related factors such as GDP growth, inflation rates, labour force participation, and education levels could lead to more accurate forecasts by providing a deeper understanding of the factors influencing unemployment trends. Including regional data could also reflect local labour market dynamics and spatial disparities. Lastly, expanding the dataset to cover a longer historical period would improve model reliability, particularly in identifying structural changes. Cross-validation techniques and residual analyses would help prevent overfitting and ensure the robustness of models like Box-Jenkins or Holt Linear Trend.

Acknowledgement

The authors would like to thank the Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, for its support.

Conflict of Interest

Authors declare that there is no conflict of interests regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Nurfara Hanim Juhaimi, Kamil Khalid;* ***analysis and interpretation of results:*** *Nurfara Hanim Juhaimi, Kamil Khalid;* ***draft manuscript preparation:*** *Nurfara Hanim Juhaimi, Kamil Khalid. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Watson, I. (2000). Beyond the unemployment rate: Building a set indices to measure the health of the labour market. Australian Bulletin of Labour, 26(3), 175-190.

[2] Khalid, W., Akalpler, E., Khan, S., & Shah, N. H. (2021). The Relationship Between Unemployment and Economic Growth in South Africa: VAR Analysis. Forman Journal of Economic Studies, 17(1).

[3] Sahib, N. S. M. A. J., & Ibrahim, N. (2022). Economic growth analysis and forecasting towards unemployment rate using Arima model. Enhanced Knowledge in Sciences and Technology, 2(1), 433-442.

[4] Manacorda, M., & Petrongolo, B. (1999). Skill mismatch and unemployment in oecd countries marco manacorda. Economica, 66(262), 181-207.

[5] Lavrinovicha, I., Lavrinenko, O., & Teivans-Treinovskis, J. (2015). Influence of education on unemployment rate and incomes of residents. Procedia-Social and Behavioral Sciences, 174, 3824-3831.

[6] Kabeer, N. (2021). Gender equality, inclusive growth, and labor markets. In Women's Economic Empowerment (pp. 13-48). Routledge.

[7] Terziev, V. (2019). Effects of the impact of the active policies on the labor market. Proceedings of SOCIOINT.

[8] Chia, X. Y., & Khalid, K. (2022). Forecasting Analysis of Unemployment Rate in Malaysia based on the Naive, Holt's Linear Trend and Box-Jenkins Models. Enhanced Knowledge in Sciences and Technology, 2(2), 011-020.

[9] Dhakal, C. P. (2017). A naïve approach for comparing a forecast model. International Journal of Thesis Projects and Dissertations, 5(1), 1-3.

[10] Nau, R. (2014). The mathematical structure of arima models. Duke University Online Article, 1(1), 1-8.

[11] Khairina, D. M., Daniel, Y., & Widagdo, P. P. (2021, July). Comparison of double exponential smoothing and triple exponential smoothing methods in predicting income of local water company. In Journal of Physics: Conference Series (Vol. 1943, No. 1, p. 012102). IOP Publishing.