Forecasting the COVID-19 Pandemic's Effect on Unemployment in Malaysia

Nur Syazatul A. Abdul Aziz, Chuan Hui Foo

Department of Mathematics, Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris, 35900 UPSI Tanjung Malim, Perak, Malaysia

ABSTRACT

This study aims to forecast Malaysia's unemployment rates using the Autoregressive Integrated Moving Average (ARIMA) model, a popular tool for analyzing time-series data. Monthly unemployment data from January 2010 to October 2022, obtained from the Malaysia Labour Market Interactive Data portal, was analyzed using the Box-Jenkins methodology. After testing several models, ARIMA(2, 1, 0) was identified as the best-fit model based on diagnostic tests, including the Ljung-Box test, and model evaluation criteria such as Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE). The forecast from November 2022 to October 2023, following the pandemic period, indicates a slight increase in Malaysia's unemployment rate. These findings are significant as they provide a valuable tool for policymakers and private sectors to anticipate economic challenges and develop strategies to address potential rises in unemployment. The forecasted data can assist the government in planning job creation and workforce development initiatives to mitigate unemployment's economic and social impacts. This study also underscores the ARIMA model's usefulness in short-term forecasting, offering insights into future trends for better financial planning.

Keywords: ARIMA model, Box-Jenkins methodology, Pandemic, Time-series forecasting, Unemployment.

INTRODUCTION

Unemployment is a critical indicator of a nation's economic health, reflecting not only the performance of the labor market but also the overall efficiency of the economy in providing employment opportunities for its citizens. A high unemployment rate can have adverse effects on the social, political, and economic stability of a country, including an increase in poverty, income inequality, and even social unrest. In Malaysia, unemployment has long been a crucial socio-economic issue, affecting the country's overall economic performance and individual well-being. The Department of Statistics Malaysia (DOSM) notes that unemployment in Malaysia has fluctuated over the years, with notable spikes during periods of economic downturns and external shocks, such as the 1997 Asian financial crisis and the 2020 COVID-19 pandemic. The pandemic, in particular, caused a surge in unemployment to 4.8% in 2020, its highest in nearly three decades, disproportionately affecting youth and graduates. Ibrahim and Mahyuddin (2017) emphasized the growing youth unemployment issue, while the World Bank (2019) attributed graduate unemployment to skills mismatch and an oversupply of graduates in certain fields. These domestic structural challenges, alongside external pressures like global economic shifts and trade policies, underscore the need for continuous monitoring and effective labor market interventions.

Forecasting unemployment rates plays a fundamental role in understanding future labor market conditions and informing economic policies aimed at mitigating unemployment. Accurate forecasting enables governments to anticipate job market fluctuations and develop strategies to address potential crises before they fully manifest. Businesses can also use unemployment forecasts for workforce planning, ensuring labor demand aligns with supply to optimize productivity.

Among the various forecasting methods, the Autoregressive Integrated Moving Average (ARIMA) model, introduced by Box and Jenkins (1976), has emerged as one of the most reliable tools for analyzing and predicting time-series data. ARIMA integrates three components: autoregression (AR), which models the dependency between current and past values; integration (I), which involves differencing to achieve

stationarity; and moving average (MA), which accounts for the dependency between current values and past forecast errors. Represented as ARIMA(*p, d, q*), where p is the number of lagged terms, d is the degree of differencing, and q is the moving average order, the model is particularly effective for short-term predictions where historical data exhibit clear temporal patterns. Zhang et al. (2018) emphasized ARIMA's flexibility in handling both stationary and non-stationary data, making it well-suited for economic indicators like unemployment rates. However, ARIMA's reliance on linear relationships limits its performance in highly volatile environments, prompting researchers like Xiong et al. (2017) to suggest the integration of ARIMA with advanced techniques like machine learning or Artificial Neural Networks (ANN).

Several studies have demonstrated ARIMA's effectiveness in forecasting unemployment rates. Naccarato et al. (2018) applied the ARIMA model to Italy's unemployment rate, highlighting its utility in short-term labor market forecasting. Similarly, Nkwatroh (2016) applied ARIMA in Cameroon, underscoring its adaptability for developing economies with limited data. In Malaysia, Ramli et al. (2018) compared ARIMA with Holt's Exponential Smoothing and found ARIMA to produce more accurate results, especially for short-term forecasts. The authors recommended ARIMA for national labor market planning due to its ability to handle the dynamic nature of unemployment data. Moreover, Abdullah (2012) and Banerjee (2014) showcased the ARIMA model's robustness in predicting other economic indicators, such as gold prices and stock markets.

Despite its advantages, ARIMA does face limitations. The model assumes linearity and performs optimally with stable data, which may not always be the case for volatile economic variables like unemployment. Further, the process of differencing non-stationary data to achieve stationarity can sometimes oversimplify complex economic dynamics. Nonetheless, ARIMA remains a preferred choice for short-term unemployment forecasting, particularly in developing economies like Malaysia, where data quality and availability constrain the adoption of more sophisticated models. Researchers like Zhang (2018) have also highlighted ARIMA's ability to capture the lagged effects of economic policies and external shocks, a valuable feature for understanding the Malaysian labor market's response to factors such as minimum wage laws, trade agreements, and oil price fluctuations.

This study aims to fill the gap in recent research by applying the ARIMA model to forecast Malaysia's unemployment rate using data from January 2010 to October 2022. Given the structural shifts caused by the COVID-19 pandemic and the need for updated forecasts, this paper seeks to provide reliable predictions for Malaysia's unemployment rate from November 2022 to October 2023. By leveraging the ARIMA model's ability to handle complex time-series data, the forecasts generated can serve as a valuable tool for policymakers to anticipate labor market trends and implement strategies to reduce unemployment.

The objectives of this study are twofold: (1) to identify the best-fitted ARIMA model for forecasting Malaysia's unemployment rate, and (2) to forecast the future values of unemployment in Malaysia from November 2022 to October 2023 based on the selected model. The study employs the Box-Jenkins approach to identify, estimate, and diagnose the ARIMA model, ensuring that the selected model provides accurate and reliable forecasts. While recognizing ARIMA's limitations, the study underscores its efficacy in providing timely unemployment forecasts in Malaysia, contributing to labor market planning and economic policy interventions. Future research can further explore hybrid models that integrate ARIMA with advanced techniques to improve forecasting accuracy and address non-linear patterns in unemployment data.

METHODOLOGY

The framework of analysis in this study serves as the conceptual and procedural foundation for applying the ARIMA model to forecast unemployment trends in Malaysia. The primary goal is to identify the best-fitting model for predicting unemployment and use it to forecast future values, providing insights for policymakers to address unemployment-related challenges. This framework is grounded in time series analysis and follows the systematic structure of the Box-Jenkins methodology, which ensures that the analysis is thorough, accurate, and reliable. This methodology has been validated in various contexts for analyzing time-series data (Nyoni & Nyoni, 2020) and is particularly suited for datasets with temporal correlations, such as monthly unemployment figures.

To begin with, the study utilizes secondary data sourced from the Malaysia Labor Market Interactive Data, a database managed by the Department of Statistics Malaysia. This data encompasses monthly unemployment figures from January 2010 to October 2022, providing a total of 154 observations. This extensive dataset not only ensures statistical robustness but also captures unemployment trends over a significant period, allowing for more precise forecasting (Zhang, 2018). The monthly nature of the data meets the time-series requirements for applying the ARIMA model.

The analysis framework revolves around the ARIMA model, a widely used technique for forecasting in time-series contexts. ARIMA, which stands for Auto-Regressive Integrated Moving Average, is ideal for this study because it allows the modelling of relationships between current and past observations while also addressing potential non-stationarity in the data. The ARIMA model incorporates three components:

1)   Auto-regressive (AR) aspect, which captures relationships between current observations and their historical values. AR models the dependency of the current value of the series, Y_t, on its p-lagged values, such that

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t \qquad (1)$$

where ϕ_i are the autoregressive coefficients, Y_{t-i} are the lags ($i = 1, 2, \ldots p$) and e_t is white noise.

2)   Integrated (I) aspect, which addresses non-stationarity through differencing. This involves differencing the series d times to achieve stationarity

$$Y_t' = Y_t - Y_{t-1}$$

for first-order differencing ($d = 1$).

3)   Moving average (MA) component, which considers dependencies on residual errors. MA captures the dependency of the series on past error terms

$$Y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} \qquad (2)$$

where θ_j are the moving average coefficients.

These three components are configured using specific parameters (p, d, q), representing the lag order, degree of differencing, and order of the moving average, respectively.

The Box-Jenkins methodology forms the backbone of the framework and involves four distinct stages: model identification, parameter estimation, diagnostic checking, and forecasting. The first stage, model identification, begins with analysing the raw unemployment data to assess its stationarity. Stationarity, a key requirement for applying ARIMA, ensures that the statistical properties of the time series, such as the mean and variance, remain consistent over time. If the data is non-stationary, differencing techniques are applied to transform it into a stationary series (Wong *et al*., 2020). Graphical tools such as autocorrelation function (ACF) and partial autocorrelation function (PACF) plots are employed to examine the data and identify potential values for the ARIMA parameters.

The second stage, parameter estimation, involves determining the specific values for the ARIMA parameters (p, d, q) that best fit the unemployment data. This is achieved through statistical techniques such as maximum likelihood estimation or non-linear least squares estimation. The suitability of the parameters is further validated using statistical tests, including p-values, where values less than 0.05 indicate statistical significance.

Once the model parameters are estimated, the third stage, diagnostic checking, ensures the adequacy and reliability of the fitted ARIMA model. Residual analysis plays a critical role in this stage, where residuals (the differences between observed and predicted values) are examined for patterns. The Ljung-Box test is applied to assess whether the residuals exhibit randomness, a key indicator of a well-fitted model. If the residuals show

significant autocorrelation, the model is deemed inadequate and requires refinement, repeating the earlier stages of parameter estimation and model identification.

The final stage, forecasting, involves applying the best-fitting ARIMA model to predict future unemployment values. This stage leverages the historical patterns captured by the model to generate forecasts for the next 12 months, covering the period from November 2022 to October 2023. Evaluation metrics, including Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE), are used to assess the accuracy of the forecasts. A lower MSE and MAPE indicate that the model can provide precise predictions, which are essential for reliable policy formulation.

Figure 1 integrates a conceptual flowchart that visualizes the sequential process of the analysis. It outlines the progression from data collection to normalization, stationarity checks, model parameterization, and eventual forecasting. It ensures clarity in the steps taken and provides a structured pathway for achieving the research objectives.

The analysis framework is designed to produce accurate unemployment forecasts and provide a systematic approach that can be replicated or adapted for similar time series analyses in other contexts. Integrating the Box-Jenkins methodology with the ARIMA model offers a robust platform for understanding unemployment dynamics in Malaysia and equipping policymakers with actionable insights to mitigate the social and economic consequences of rising unemployment rates.
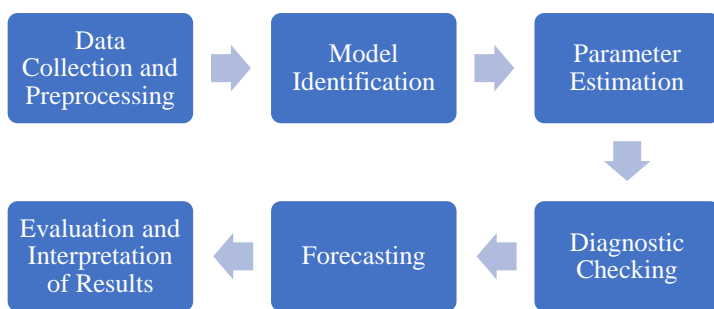
Figure 1. A framework of analysis: The Box-Jenkins approach

DATA ANALYSIS

The data utilized in this study are sourced from the Malaysia Labour Market Interactive Data provided by the Department of Statistics Malaysia (DOSM). This dataset contains monthly records of unemployment figures spanning from January 2010 to October 2022, providing a total of 154 observations. The dataset reflects the number of unemployed individuals in Malaysia each month and serves as the foundation for the time series analysis conducted in this study.

The choice of this dataset is driven by its official and credible nature, ensuring accuracy and reliability. As the DOSM is the primary agency responsible for collecting and disseminating labor market statistics in Malaysia, its data are widely recognized as authoritative. These records are collected and processed using standardized statistical methodologies, minimizing potential errors and inconsistencies.

The time frame from 2010 to 2022 was selected to capture a comprehensive view of unemployment trends over more than a decade. This period covers various economic cycles, including periods of growth, recession, and the significant disruptions caused by the COVID-19 pandemic. Analyzing this extended period allows for a more nuanced understanding of unemployment patterns in Malaysia. It enables the ARIMA model to account for both long-term trends and short-term fluctuations.

The dataset is structured monthly, making it well-suited for time series analysis. Monthly unemployment data provide the granularity required for identifying seasonal trends, abrupt changes, and other temporal patterns

that might influence forecasting accuracy. By working with a dataset of this nature, the study can generate detailed and reliable future employment projections.

In terms of scope, the data cover the entire geographical region of Malaysia, encompassing unemployment figures across all states and territories. This nationwide coverage ensures that the findings and forecasts derived from the study represent the country as a whole, making them valuable for both policymakers and stakeholders in various sectors. The original dataset is in raw numerical form and, therefore, required pre-processing to ensure its compatibility with the ARIMA modeling process. This includes steps such as normalization and differencing, as detailed in subsequent sections. Furthermore, the dataset was subjected to rigorous statistical checks, including stationarity and normality tests, to verify its suitability for time series forecasting.

By leveraging this rich and comprehensive dataset, the study aims to provide an accurate and insightful analysis of Malaysia's unemployment trends. The findings will serve as a critical input for government agencies, private sector organizations, and researchers aiming to address unemployment issues and develop informed labor market strategies.

ARIMA The methodology employed in this study focuses on the application of the ARIMA (Auto-Regressive Integrated Moving Average) model to forecast unemployment trends in Malaysia. This process involves several systematic steps to ensure the accuracy and validity of the forecasts. The stages include data collection, pre-processing, the application of the Box-Jenkins approach, model evaluation, and forecasting. Each stage is elaborated below.

Data Pre-processing

Before employing ARIMA modelling, the raw data must undergo preprocessing to ensure it is suitable for analysis. A critical assumption of ARIMA is that the data must be stationary. Therefore, the first task is to test for stationarity using the Augmented Dickey-Fuller (ADF) test. Stationarity implies that the mean, variance, and autocovariance of the data remain constant over time.

If the ADF test reveals that the dataset is non-stationary, differencing is applied iteratively until stationarity is achieved. Differencing involves subtracting the value of an observation from its previous value, a process that stabilizes the data by eliminating trends or seasonality.

Normalization is another key pre-processing step. The raw unemployment data is first evaluated for normality using statistical tests. The Kolmogorov-Smirnov or Shapiro-Wilk test can be applied to determine whether the dataset follows a normal distribution. The dataset is normalized to bring all data points into a common scale, which simplifies analysis and ensures compatibility with the ARIMA model. The normalization process is performed using the formula:

$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{3}$$

where X represents the raw data value, while X_{min} and X_{max} denote the minimum and maximum values within the dataset, respectively. The normalized data lies within the range of 0 to 1. The hypothesis for this test is as follows:

- Null Hypothesis (H_0): The dataset follows a normal distribution.

- Alternative Hypothesis (H_1): The dataset does not follow a normal distribution.

The test employs a significance level (α=0.05), with a p-value below this threshold leading to rejection of H_0 (Suwardo *et al*., 2009). For the unemployment dataset, the p-value was less than 0.05, indicating non-normality (Figure 2). The plot reveals systematic departures from the straight line. Since the raw data does not follow a normal distribution, it must be normalized to transform them into normal data through Eq. (3).

Stationarity is then assessed by plotting the time series data and observing the trends and variability over time. Non-stationary data exhibits a changing mean or variance. To address this, differencing is applied to remove trends and achieve stationarity. Stationarity is confirmed through visual inspection of time series plots and by examining Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. A stationary dataset is essential because ARIMA models assume constant statistical properties over time.
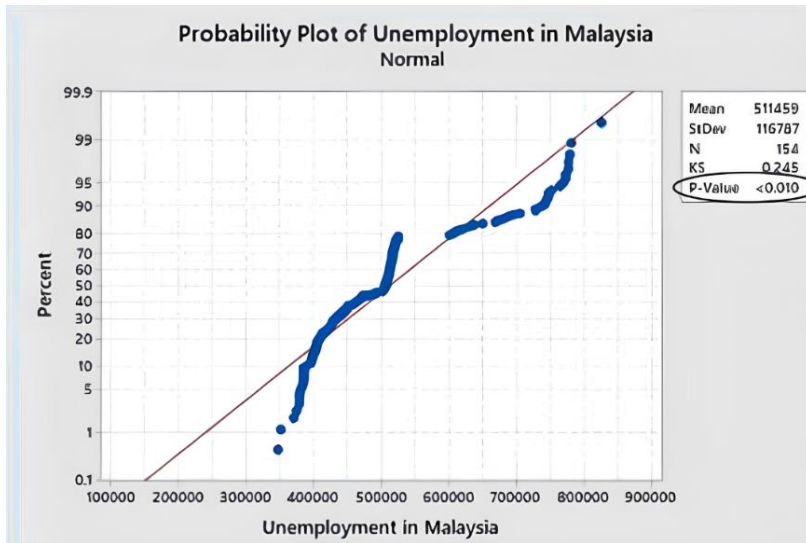
Figure 2. Normal probability plot for unemployment in Malaysia (2010-2022)

Model Selection: Box-Jenkins Approach

Autoregressive Integrated Moving Average (ARIMA) models are a type of Box-Jenkins methodology. One of the steps in determining whether the data exhibits seasonal, cyclic, or random variation patterns is to use a time series plot (Figure 3). Since the time series plot shows random variation and does not display any cyclical pattern, the data can be analysed using an ARIMA model.
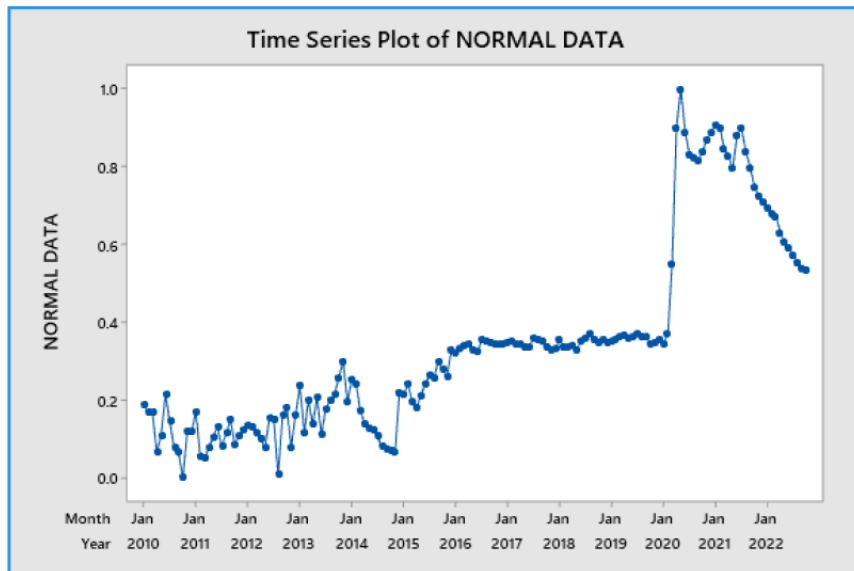
Figure 3. Time series plot showing random variation and no cyclical pattern

The Box-Jenkins methodology forms the core of the ARIMA modelling process. It involves four distinct stages:

(1)   Model Identification

At this stage, the stationary dataset is analysed to determine the parameters p, d, and q of the ARIMA model:

- p: The order of the autoregressive (AR) component.

- d: The degree of differencing applied to the data.

- q: The order of the moving average (MA) component.

To identify p and q, this study examines the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. The ACF plot shows the correlation between the data and its lagged values, helping to determine the order of the moving average process (q). Based on the Autocorrelation Function (ACF) plot in Figure 4, it is evident that the ACF decays slowly. This indicates that the data is non-stationary. If the ACF plot quickly drops to zero, it suggests that the data is stationary (Pherwani & Vyjayanthi, 2017).
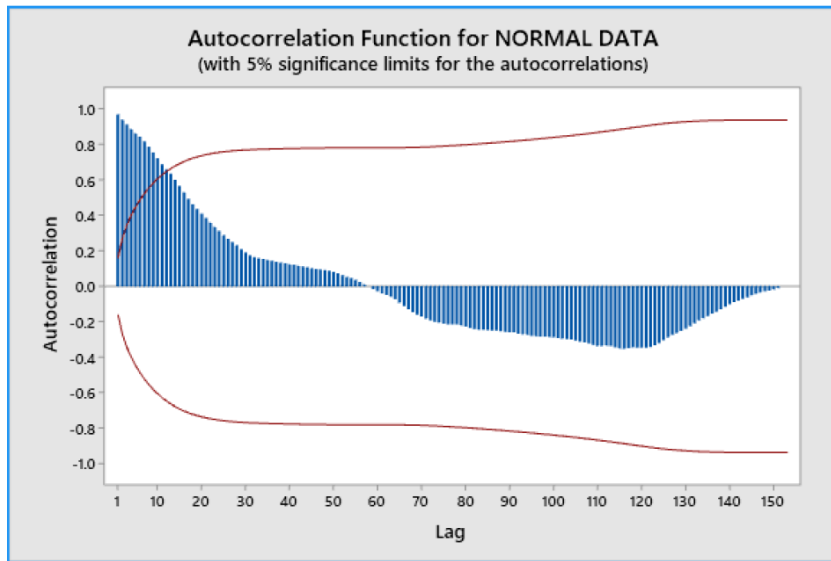
Figure 4. ACF plot for unemployment in Malaysia from January 2010 to October 2022

Meanwhile, the PACF plot isolates the direct correlation of a lagged observation, which helps identify the order of the autoregressive process (p). The Partial Autocorrelation Function (PACF), as shown in Figure 5, indicates a truncated spike, suggesting that the unemployment data should be differentiated to achieve a stationary state. Additionally, the stationarity of the time series data can be determined using a Box-Cox plot (Figure 6).
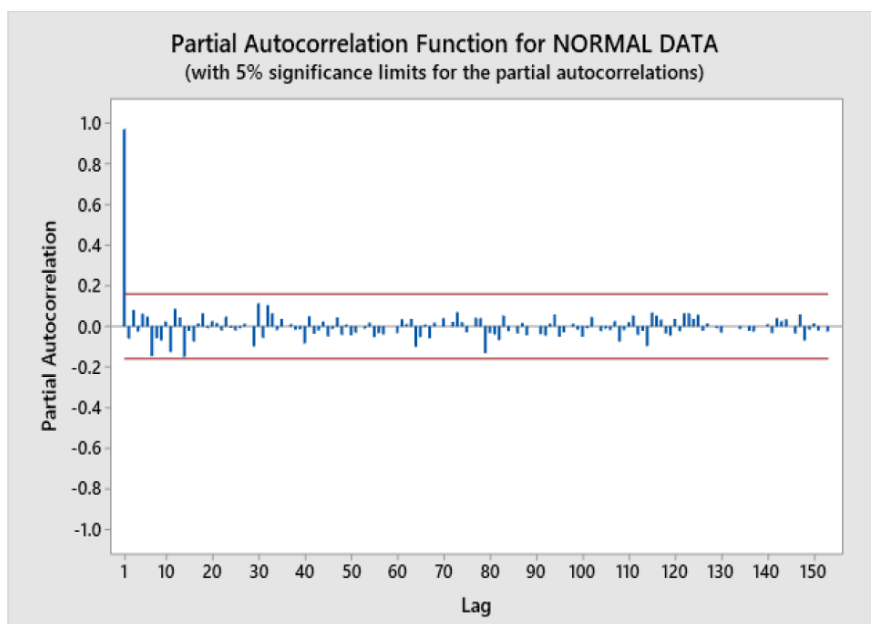
Figure 5. PACF plot indicating the need for differentiation to achieve stationarity
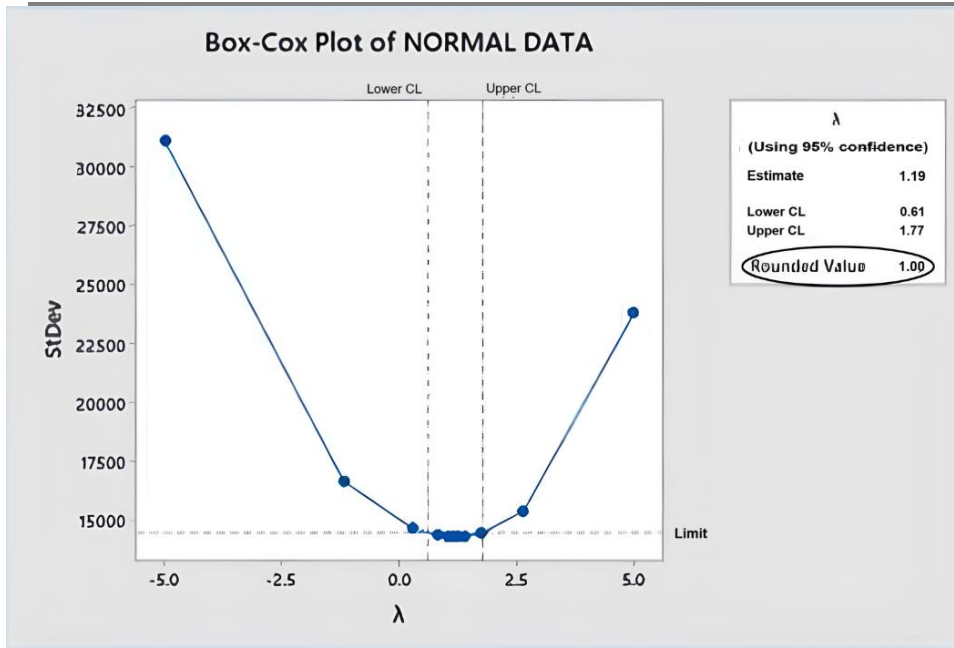
Figure 6. Box-Cox Transformation plot showing a rounded value of 1.00

Figure 7 shows that the rounded value in the Box-Cox Transformation plot is 1.00, which is equivalent to using the original data. Therefore, no transformation is needed if the optimal confidence interval includes 1.00. However, if the confidence interval does not include 1.00, a transformation is required. In conclusion, since the rounded value is 1.00, no transformation is necessary before proceeding with the differentiation process.

Differencing 1 was performed, and the ACF and PACF correlograms were plotted to determine the appropriate ARIMA model. Based on the ACF plot for differencing 1, no significant lags were observed, indicating that the value of q is 0. The PACF plot for differencing 1 shows three significant lags, suggesting that the value of p is 3. Consequently, the candidate models considered were ARIMA(3, 1, 0), ARIMA(2, 1, 0), and ARIMA(1, 1, 0).

Upon comparing these models, the ARIMA(3, 1, 0) model stands out with the smallest residual sum of squares (RSS), indicating that it fits the data better than the other two (Table 1). Although ARIMA(2, 1, 0) and ARIMA(1, 1, 0) also demonstrate reasonable fits, their RSS values are slightly higher, making them less optimal in comparison. Additionally, the ARIMA(0, 1, 0) model was deemed unsuitable because it lacked both autoregressive and moving average terms, limiting its ability to capture the structure in the data.
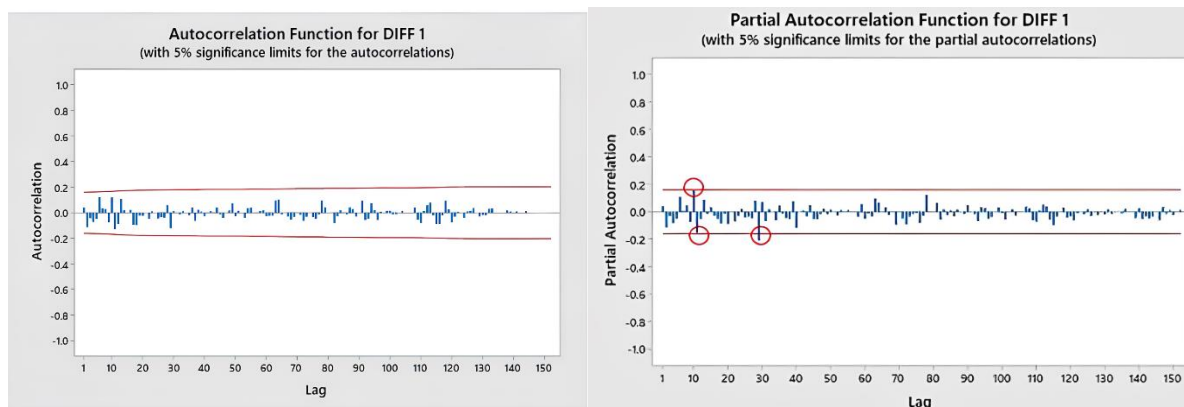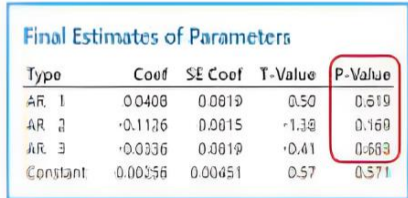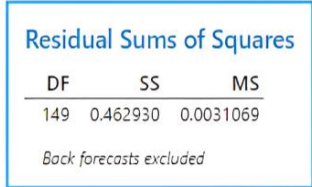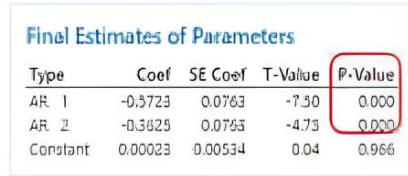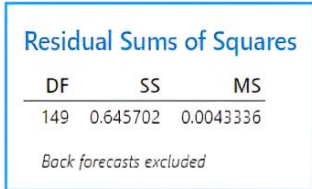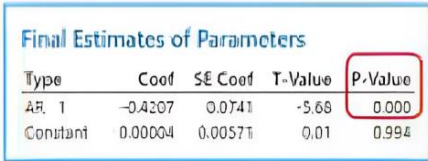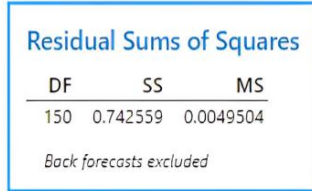

Figure 7. ACF and PACF for the First-Order Differenced Data (DIFF 1)

Overall, the ARIMA(3, 1, 0) model appears to provide the best balance between complexity and fit, making it the most appropriate choice for the dataset under consideration.

Table 1 Analysis of three candidate ARIMA models

ARIMA MODEL	PARAMETERS	RESIDUAL SUM OF SQUARE
ARIMA (3, 1, 0)	**Final Estimates of Parameters** Type — Coef — SE Coef — T-Value — P-Value AR 1 — 0.0408 — 0.0819 — 0.50 — 0.619 AR 2 — -0.1176 — 0.0815 — -1.39 — 0.160 AR 3 — -0.0336 — 0.0819 — -0.41 — 0.683 Constant — 0.00256 — 0.00451 — 0.57 — 0.571	**Residual Sums of Squares** DF — SS — MS 149 — 0.462930 — 0.0031069 Back forecasts excluded
ARIMA (2, 1, 0)	**Final Estimates of Parameters** Type — Coef — SE Coef — T-Value — P-Value AR 1 — -0.5723 — 0.0763 — -7.50 — 0.000 AR 2 — -0.3825 — 0.0795 — -4.73 — 0.000 Constant — 0.00023 — 0.00534 — 0.04 — 0.966	**Residual Sums of Squares** DF — SS — MS 149 — 0.645702 — 0.0043336 Back forecasts excluded
ARIMA (1, 1, 0)	**Final Estimates of Parameters** Type — Coef — SE Coef — T-Value — P-Value AR 1 — -0.4207 — 0.0741 — -5.68 — 0.000 Constant — 0.00004 — 0.00571 — 0.01 — 0.994	**Residual Sums of Squares** DF — SS — MS 150 — 0.742559 — 0.0049504 Back forecasts excluded

(2)   Parameter Estimation

Parameter estimation is the second stage in the Box-Jenkins methodology for ARIMA modeling. This stage involves determining the values of the parameters p, d, and q that best fit the chosen ARIMA model to the data. Accurate parameter estimation ensures the model's predictive power and reliability.

The process begins with identifying the autoregressive order, p, which signifies the number of lag observations influencing the current value of the time series. This is assessed using the Partial Autocorrelation Function (PACF) plot, which helps determine which past data points have significant correlations with the present.

This study uses maximum likelihood estimation (MLE) to determine the values of the model parameters. The p-value of the estimated parameter must be less than 0.05.

For ARIMA(3, 1, 0) model, three autoregressive (AR) terms are considered. The coefficients of the AR terms (AR1, AR2, AR3) and the constant are shown with their corresponding standard errors, t-values, and p-values. None of the AR terms have statistically significant coefficients, as all their p-values exceed the conventional significance threshold (0.05). This suggests that the AR terms do not contribute significantly to explaining the variation in the data. The residual sum of squares (SS) for this model is 0.462930, with a mean square (MS) of 0.0031069.

The ARIMA(2, 1, 0) model includes two autoregressive terms. Unlike the ARIMA(3,1,0), both AR1 and AR2 terms are statistically significant, as indicated by their p-values of 0.000. These significant coefficients imply that including these AR terms effectively captures the underlying data patterns. The residual sum of squares

(SS) for this model is 0.645702, and the mean square (MS) is 0.0043336, both higher than those in the ARIMA(3,1,0), indicating a slightly poorer fit in terms of residual error compared to the previous model.

For ARIMA(1, 1, 0), it includes only one autoregressive term. The AR1 term is statistically significant, with a p-value of 0.000, while the constant term is not significant (p-value of 0.994). The residual sum of squares (SS) is 0.742559, and the mean square (MS) is 0.0049504, which are the highest among the three models, suggesting a worse fit in terms of residuals compared to both ARIMA(2,1,0) and ARIMA(3,1,0).

The results suggest that while the ARIMA(3,1,0) has the lowest residual sum of squares, its parameters are not statistically significant, which undermines its reliability as a predictive model. On the other hand, ARIMA(2,1,0) provides a balance with significant parameters and a moderately low residual error, making it the most interpretable and likely the best model among the three for capturing the data dynamics. The ARIMA(1,1,0), despite its simplicity, has the highest residual error, indicating it may not adequately explain the data patterns.

Once a model is chosen, its parameters need to be estimated. Table 2 presents the parameters derived for the selected models.

Table 2 Estimates of parameters for the ARIMA(2, 1, 0) model

Type	Coefficient ($\hat{\phi}$)	Standard Error (SE)	T-Value	P-Value
AR(1)	-0.5723	0.0763	-7.5	0
AR(2)	-0.3625	0.0763	-4.75	0
Constant	0.00023	0.00534	0.04	0.966

Based on Eq. (1) and Eq. (2), the ARIMA(2, 1, 0) model is written as:

$$y_t = \left(\theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p}\right) + \left(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}\right)$$

$$= 0.00023 - 0.5723 y_{t-1} - 0.3625 y_{t-2} + e_t$$

(3)   Diagnostic Checking

This section focuses on validating the adequacy of the selected ARIMA model, specifically ARIMA(2,1,0), using residual analysis and statistical tests. Figure 8 shows no significant lags exceeding the 5% significance level. This indicates that the residuals are uncorrelated and suggests that the model has captured the underlying patterns in the data effectively. The lack of structure in the residuals confirms the model's adequacy for forecasting purposes.
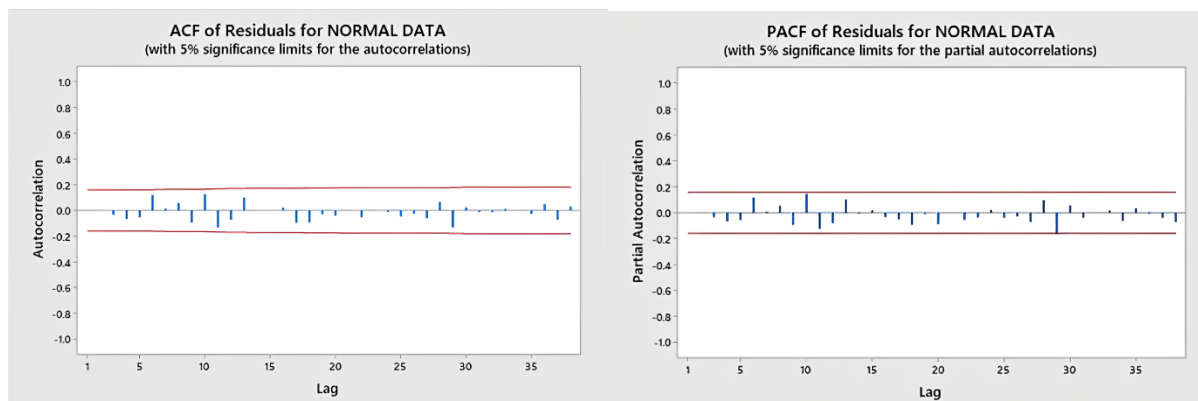

Figure 8. ACF and PACF of residual plot for ARIMA(2, 1, 0) model

While uncorrelated residuals from ACF/PACF plots suggest a good model, the Ljung-Box test provides a rigorous statistical check to validate this assumption, ensuring no systematic patterns have been overlooked. The Ljung-Box test was conducted on the residuals, with the p-value for the Modified Box-Pierce statistic being greater than 0.05 (Figure 9). This result indicates that the residuals exhibit white noise behavior, implying that the model does not suffer from lack of fit and adequately captures the data dynamics.

Based on the diagnostic results, the ARIMA(2,1,0) model is deemed suitable and robust for forecasting. The residual analysis confirms that the model satisfies key assumptions necessary for reliable time series predictions, such as independence and randomness of residuals.

Figure 9. Ljung-Box test on ARIMA(2, 1, 0) model

(4)   Data Forecasting

Using the best-fitted ARIMA(2, 1, 0) model, Malaysia's unemployment rate was forecasted for 12 months. The model was trained on 90% of the 154 monthly unemployment data points, ensuring robustness. Figure 10 includes forecasted values beyond the historical data and confidence intervals, which help assess the uncertainty of future predictions. The forecast revealed a slight upward trend in unemployment from November 2022 to October 2023.
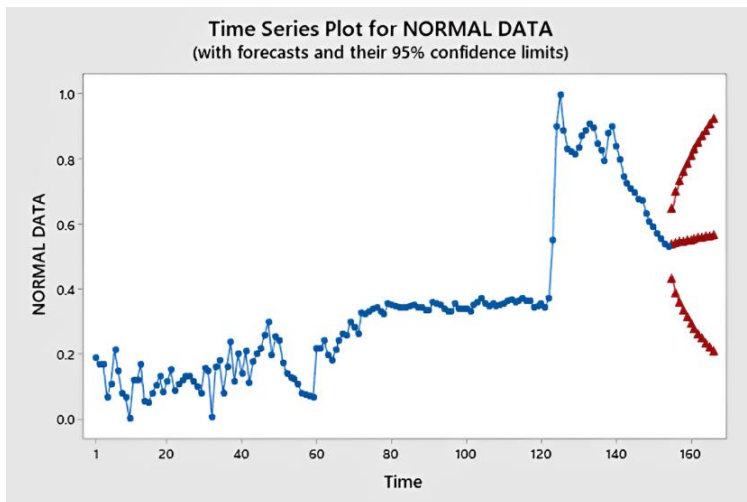

Figure 10. Time series forecasting of unemployment data (Nov 2022 to Oct 2023)

The historical data from January 2010 to November 2022 show that it fluctuates over time. However, there are some significant spikes or jumps, particularly around the 120th month of the data. This sharp change could indicate an outlier, a structural break, or some external factor influencing the data at that point. Such disruptions in the time series can pose challenges for accurate forecasting. While the model captures the general trend of the data, it might be sensitive to such abrupt changes, which is something to consider when evaluating its forecasts.

The forecast goes beyond the available data, covering the period from November 2022 to November 2023, with predicted values marked by red arrows or lines on the plot. These forecasts are accompanied by uncertainty, shown through widening confidence intervals. The 95% confidence interval represents the range

in which future values are likely to fall, with 95% probability. As the forecast extends further into the future, the intervals expand, indicating greater uncertainty.

Widening confidence intervals are common in time series forecasting because the model has less data to rely on as it predicts further ahead. The model can be more accurate near the end of the observed data, but uncertainty grows as the forecast moves away from the last data point. The confidence intervals also show the limitations of the model and the unpredictability of long-term forecasts.

Given the data's behavior and outliers, it's important to ensure the model fits the dataset. Diagnostic checks, like residual analysis, should be done to confirm that the model assumptions are valid. If the model is correctly specified, the forecast will likely be useful, but if the model doesn't fit the data well, the forecasts may become less reliable as uncertainty increases.

In conclusion, while the plot shows the forecast and confidence intervals, the growing uncertainty suggests that the forecast's accuracy decreases as the time horizon extends.

CONCLUSION

This study successfully achieved its objectives by employing the ARIMA model for forecasting unemployment in Malaysia. The first objective, identifying the most suitable ARIMA model for unemployment forecasting, was accomplished through the application of the Box-Jenkins approach. This methodology involved four critical stages: model identification, parameter estimation, diagnostic checking, and forecasting. Following these steps, the ARIMA (2, 1, 0) model was determined to be the most appropriate for forecasting unemployment trends.

The second objective focused on forecasting the future unemployment rates in Malaysia for the period from November 2022 to October 2023 using the selected ARIMA model. The data, forecasted for 12 months, were transformed back to their original scale for interpretation. The analysis revealed a slight upward trend in unemployment rates over the forecasted period.

This research offers valuable insights and guidance to various stakeholders in Malaysia. For individuals, it highlights potential challenges and opportunities related to future unemployment trends. For policymakers and government agencies, such as the Department of Statistics Malaysia, the findings provide critical data for planning and implementing strategies to mitigate rising unemployment. Furthermore, the private sector can utilize this forecast to adjust hiring strategies and create new job opportunities, contributing to economic stability.

Additionally, addressing unemployment proactively can help reduce associated social issues, such as increased crime rates, which are often linked to economic hardship and job scarcity. By creating more employment opportunities, these negative consequences can be mitigated, fostering a safer and more prosperous society.

Forecasting unemployment is crucial for understanding labour market dynamics and guiding economic policies. This study's identification of the ARIMA (2, 1, 0) model as the best fit demonstrates its utility in providing reliable unemployment forecasts for Malaysia. These forecasts can serve as a valuable tool for policymakers, businesses, and researchers to anticipate future challenges and opportunities in the labour market

ACKNOWLEDGMENTS

REFERENCES

1. Abdullah, L. (2012). ARIMA model for gold bullion coin selling prices forecasting. International Journal of Advances in Applied Sciences, 1(4), 153-158.
2. Banerjee, D. (2014). Forecasting of Indian stock market using time-series ARIMA model. 2014 2nd International Conference on Business and Information Management (ICBIM), 131-135.
3. Box, G. E. P., & Jenkins, G. M. (1976). Time series analysis: Forecasting and control (2nd ed.). Holden-Day.
4. Ibrahim, D. H. M., & Mahyuddin, M. Z. (2017). Youth unemployment in Malaysia: Developments and policy considerations. Outlook and Policy, Annual Report.
5. Naccarato, A., Falorsi, S., Loriga, S., & Pierini, A. (2018). Combining official and Google Trends data to forecast the Italian youth unemployment rate. Technological Forecasting and Social Change, 130, 114–122.
6. Nkwatroh, Louis. (2016). Can Cameroon become an emerging economy by the year 2035? Projections from univariate time series analysis. Journal of Economics and International Finance, 8, 155-167.
7. Nyoni, S. P., & Nyoni, T. (2020). Using artificial neural networks for predicting new dysentery cases in children under 5 years of age in Chitungwiza urban district, Zimbabwe. EPRA International Journal of Research and Development (IJRD), 5(2), 215–221.
8. Pherwani, N., & Vyjayanthi, K. (2017). Using ARIMA Model to Forecast Sales of an Automobile Company. International Journal of Science & Engineering (IJSTE), 4(5).
9. Ramli, S. F., Fidaus, 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.
10. Suwardo, W., Napiah, M., & Kamaruddin, I. (2010). ARIMA models for bus travel time prediction. Journal of the institute of engineers Malaysia, 71(1), 49-58.
11. Wong, W. M., Subramaniam, S. K., Feroz, F. S., Subramaniam, I. D., & Rose, L. A. F. (2020). Flood Prediction using ARIMA Model in Sungai Melaka, Malaysia. International Journal of Advanced Trends in Computer Science and Engineering, 9(4), 5287-5295.
12. Xiong, T., Li, C., & Bao, Y. (2017). Seasonal forecasting of agricultural commodity price using a hybrid STL and ELM method: Evidence from the vegetable market in China. Neurocomputing, 275, 1311–1321.
13. Zhang, M. (2018). Time Series: Autoregressive models AR, MA, ARMA, ARIMA. University of Pittsburgh.
14. Zhang, Q., Jin, Q., Chang, J., et al. (2018). Kernel-weighted graph convolutional network: A deep learning approach for traffic forecasting. In 2018 24th International Conference on Pattern Recognition (ICPR) (pp. 1018–1023). IEEE.