Journal of Advanced Research in Computing and Applications

Journal homepage: https://karyailham.com.my/index.php/arca
2462-1927

Forecasting the Unemployment Rate in Malaysia using ARIMA Models

Suhaila Bahrom[1,*], Eszleen Sies[2], Nurul Nisa' Khairol Azmi[3], Nurul Aini Abdul Wahab[3]

[1] Department of Mathematics, Centre for Foundation Studies, International Islamic University Malaysia, Gambang Campus, 26300 Pahang, Malaysia
[2] College of Computing & Informatics, Universiti Tenaga Nasional, Putrajaya Campus, Jalan Ikram-UNITEN, 43000 Kajang, Selangor, Malaysia
[3] Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Negeri Sembilan, Seremban Campus, 70300 Seremban, Negeri Sembilan, Malaysia

ARTICLE INFO

Article history:
Received 1 December 2025
Received in revised form 1 January 2026
Accepted 2 February 2026
Available online 8 March 2026

Keywords:
Unemployment rate; ARIMA; time series analysis; structural breaks

ABSTRACT

Considerable fluctuations in Malaysia's labour market have been driven by both global economic uncertainties and domestic transformations. Events such as the COVID-19 pandemic exposed the labour markets to certain vulnerabilities. Gaps remain in the literature regarding the continuous validation of forecasting models of the unemployment trend in the presence of extreme events such as COVID-19, particularly over extended and post-pandemic periods. It is essential to develop an accurate forecast for unemployment trends to inform effective labour market policies and economic planning. This study analyses the monthly unemployment rate in Malaysia from January 2016 to March 2025 with the aim of modeling and forecasting Malaysia's monthly unemployment rate using a time series approach. A structural break analysis, assuming one break, identified a significant change in trend occurring in February 2020, which aligns with the onset of the COVID-19 pandemic. This suggests that the pandemic had a notable impact on the labour market. The identified break point is then used as a basis for modeling the data using ARIMA, considering the change in trend to improve forecasting accuracy. Two distinct forecasting models are developed: one assumes no pandemic, using data before February 2020, and the other reflects the post-pandemic period. The forecasts for both periods highlight how the pandemic altered the unemployment trend and model structures. These findings underscore the importance of accounting for structural breaks in time series modelling to improve forecasting accuracy and provide critical evidence for labour market policy and planning in line with the United Nations Sustainable Goal (SDG) 8.

1. Introduction

Unemployment is a key indicator of a nation's economic well-being, with implications for household livelihoods, social stability, and policy direction. In Malaysia, the labour market has experienced notable fluctuations in recent years, influenced by both global economic conditions and domestic structural changes. The COVID-19 pandemic worsened these challenges, causing a sharp

* *Corresponding author.*
E-mail address: suhaila_b@iium.edu.my

https://doi.org/10.37934/arca.42.1.93100

rise in unemployment and exposing the vulnerability of employment systems during periods of disruption [1].

Improving the ability to forecast unemployment trends is important for informed labour market planning. More reliable projections can help policymakers anticipate developments, design targeted measures, and support recovery initiatives. The Autoregressive Integrated Moving Average (ARIMA) model has been widely used in short-term economic forecasting for its clarity and adaptability [2]. Malaysian studies have reported its usefulness in modelling unemployment variations [3,4], although most have centred on specific crisis periods without assessing model consistency over longer horizons or adapting to post-pandemic labour conditions. Globally, advanced approaches such as machine learning have gained attention [5-8], but applications in Malaysia's unemployment forecasting remain limited.

A significant gap in the literature is the treatment of structural breaks, which are sudden changes in data patterns that arise from policy shifts, economic shocks, or significant events such as pandemics [9-13]. If left unaccounted for, these changes may reduce the accuracy of forecasts and obscure important shifts in labour market behaviour. In Malaysia, the COVID-19 outbreak represents a significant structural break, introducing dynamics that conventional continuous-trend models may not fully capture.

This study analyses Malaysia's monthly youth unemployment rate (ages 15–30) from January 2016 to March 2025 using the ARIMA model, with a statistically identified structural break in February 2020 included to account for the impact of the COVID-19 pandemic. The objective is to evaluate whether including this break enhances the model's accuracy in capturing labour market dynamics. The research aligns with the United Nations Sustainable Development Goal (SDG) 8: Decent Work and Economic Growth, as more accurate unemployment forecasts can guide policies and programmes that promote sustainable employment and economic resilience.

2. Data and Methodology

2.1 Data Description

This study utilised the monthly unemployment rate of youth in Malaysia, focusing on individuals aged 15 to 30. The dataset covers the period from January 2016 to March 2025 and consists of 111 monthly observations. The data were obtained from the Malaysia Open Data Portal, which provides publicly accessible official statistics. The chosen timeframe captures both pre-pandemic and post-pandemic periods, enabling an examination of potential labour market changes linked to external shocks such as the COVID-19 pandemic.

2.2 Structural Break Analysis (SBA)

A time series plot was also generated to visualise patterns and detect potential anomalies as an initial step in identifying possible structural breaks in the data [11]. Such breaks can change the data's underlying pattern, affecting trend, variance, or timing, and if ignored, may lead to biased estimates and reduced forecasting accuracy [14]. Visual inspection and contextual evidence suggested February 2020 as a potential breakpoint, coinciding with the onset of the COVID-19 pandemic in Malaysia. This date was selected to reflect the significant labour market disruptions that emerged during the pandemic period. Further test on the presence of a structural break was conducted using the Zivot-Andrews Test [15] the Chow test (*F*-test) [16].

The dataset was subsequently divided into two periods: January 2016 to February 2020 (pre-pandemic) and March 2020 to March 2025 (post-pandemic). The Chow test was applied, and if the computed *F*-statistic was significant at the 0.05 level, it confirmed the presence of a structural break in the unemployment rate series around February 2020. The *F*-statistic is given by (1):

$$F = \frac{\left(RSSR_R - SSR_1 - SSR_2\right)/k}{\left(SSR_1 + SSR_2\right)/\left(n-2k\right)} \tag{1}$$

where $RSSR_R$ is the residual sum of squares from the restricted model (pooled data), SSR_1 and SSR_2 are the residual sums of squares from the separate regressions for each period, k is the number of parameters estimated, and n is the total number of observations.

2.3 ARIMA Model Development

This study adopts the Box-Jenkins methodology to develop an effective ARIMA model for forecasting Malaysia's unemployment rate. The methodology consists of four main stages: model identification, parameter estimation, diagnostic checking through residual analysis, and forecasting.

In the model identification stage, the time series is examined for stationarity in both variance and mean. If the series is non-stationary, necessary transformations such as the Box-Cox transformation [17] and differencing are applied. Stationarity in mean is assessed using the Augmented Dickey-Fuller (ADF) test Dickey *et al.*, [18] supported by visual inspection of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. During the parameter estimation stage, the suitable ARIMA model parameters are determined, including the autoregressive order (*p*), differencing order (*d*), and moving average order (*q*). The general form of the ARIMA(*p,d,q*) model is expressed as (2).

$$\phi\left(B\right)\left(1-B\right)^d Y_t = \theta\left(B\right)\varepsilon_t \tag{2}$$

where Y_t denotes the actual time series value at time t, B is the backshift operator such that $BY_t = Y_{t-1}, d$ denotes the degree of differencing applied to achieve stationarity, $\phi\left(B\right)$ is the autoregressive (AR) polynomial of order p, and $\theta\left(B\right)$ is the moving average (MA) polynomial of order q. The term ε_t represents a white noise error at time t. Given the structural break detected at time T_b, the series is divided into two periods:

Period 1 (pre-pandemic)	:	$t = 1, 2, ..., T_b - 1$
Period 2 (post-pandemic)	:	$t = T_b, T_b + 1, ..., T$

Accordingly, separate ARIMA models are specified for each period, allowing different ARIMA orders to be estimated based on the series characteristics within each period as (3) and (4):

Pre-pandemic ARIMA $\left(p_1, d_1, q_1\right)$	:	$\phi_1\left(B\right)\left(1-B\right)^{d_1} Y_t = \theta_1\left(B\right)\varepsilon_t, \quad t < T_b$	(3)
Post-pandemic ARIMA $\left(p_2, d_2, q_2\right)$	:	$\phi_2\left(B\right)\left(1-B\right)^{d_2} Y_t = \theta_2\left(B\right)\varepsilon_t, \quad t \geq T_b$	(4)

3. Results and Discussion

3.1 Unemployment in Malaysia

The monthly youth unemployment rate in Malaysia (ages 15-30) from January 2016 to March 2025 averaged about 7%, with the lowest rate recorded at 5.7% and the highest at 10%, as shown in Fig. 1. An upward shift in unemployment becomes apparent in early 2020. Structural break analysis, assuming a single breakpoint, detected a significant shift in March 2020 that coincided with the onset of the COVID-10 pandemic. This period was associated with the widespread disruptions across various industries, aligning with global findings that the pandemic adversely affected labour markets [19,20].

Results from the Zivot-Andrews unit root test [14] suggested a possible break around observation 50, as depicted in Fig. 1. The Chow test has confirmed that a structural break has statistically significantly existed, with a break point specified at observation number 51 (March 2020). The breakpoint marks a transition from a relatively stable pre-pandemic labour market to a more volatile post-pandemic period, influenced by policy measures such as the Movement Control Orders (MCOs) and broader economic challenges in Malaysia.

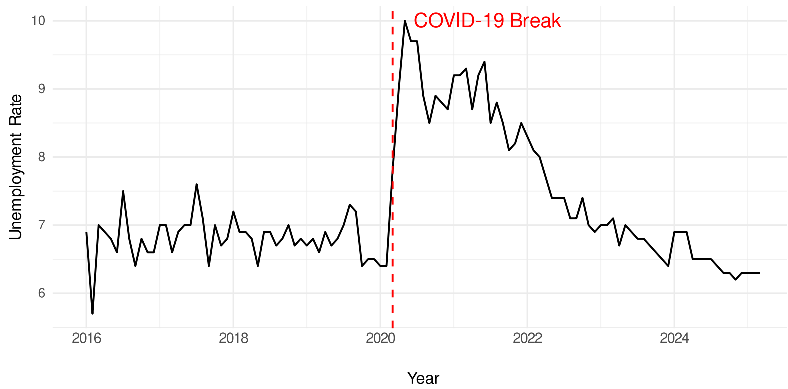

Fig. 1. Monthly unemployment rate in Malaysia with a structural break in March 2020

3.2 Modeling the Unemployment Rate with a Structural Break

As outlined in Sections 2.2 and 2.3, the unemployment rate series was divided into pre-pandemic and post-pandemic periods based on the structural break detected in March 2020. Separate ARIMA models were developed for each period to capture their distinct characteristics. The pre-pandemic model reflected a relatively stable unemployment trend with moderate fluctuations and consistent parameters, whereas the post-pandemic model required adjustments to account for increased volatility and a noticeable upward shift in unemployment levels. This approach follows the recommendations of Box *et al.*, [2] for adapting ARIMA models in the presence of structural breaks.

The ADF test was first applied to assess stationarity in the regional series, yielding a non-significant result (test statistic = -1.78, p = 0.6648), indicating non-stationarity. The Zivot-Andrews unit root test, which accounts for the possibility of a structural break, detected a statistically significant breakpoint at observation 51 (March 2020) with a test statistic of -8.54, exceeding the critical value of -5.08. Following this, the ADF test was applied separately to the pre-pandemic and post-pandemic series before ARIMA specification. Both raw series were found to be non-stationary (p>0.05), but stationarity was achieved for the first-differenced series (p<0.01). Subsequently, the ACF and PACF plots were examined to guide model specification. For the pre-pandemic period, the ACF plot exhibited a significant spike at lags 1, while the PACF plot showed significant spikes at lags

1, 2, and 5, suggesting a possible ARIMA(3,1,2) model as illustrated in Fig. 2. Table 1 summaries the comparison of six possible ARIMA models. Among these, ARIMA(0,1,1) achieved the lowest AIC value (33.03) with competitive RMSE, MAE, and MAPE values, indicating it as the best-fitting model for this period.

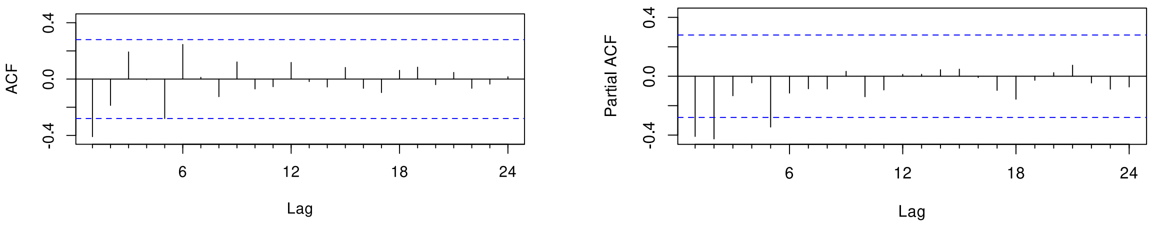

Fig. 2. ACF and PACF plot of the differenced unemployment rate for the pre-pandemic period

Table 1
ARIMA Model comparison (pre-pandemic)

Model	AIC	RMSE	MAE	MAPE (%)
ARIMA(3,1,2)	38.93	0.3037	0.2283	3.38
ARIMA(3,1,1)	37.14	0.3045	0.2279	3.37
ARIMA(2,1,1)	36.25	0.3070	0.2279	3.37
ARIMA(1,1,1)	34.56	0.3087	0.2298	3.40
ARIMA(1,1,2)	36.03	0.3068	0.2284	3.38
ARIMA(0,1,1)	33.03	0.3095	0.2324	3.44

Diagnostic checks, as presented in Fig. 3, show the ACF plot of residuals from ARIMA (0,1,1) model for the pre-pandemic period. The absence of significant autocorrelation, supported by the Ljung-Box test ($p = 0.650$), indicates that the residuals behave like white noise, thereby validating the adequacy of the selected model.

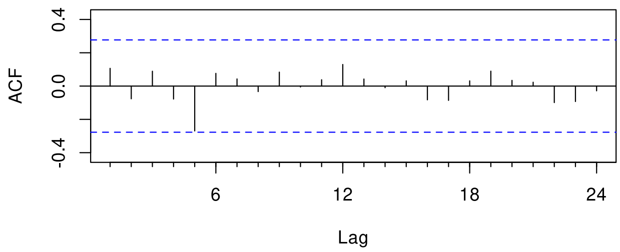

Fig. 3. ACF plot of residuals from the ARIMA(0,1,1) model (pre-pandemic)

During the post-pandemic period, the ACF and PACF plots in Fig. 4 show that the ACF has no significant spike, while the PACF displays a significant spike at lag 3, suggesting the selection of the ARIMA(1,1,0) model. However, ARIMA(0,1,0) achieved the lowest AIC value (43.96), outperforming ARIMA(1,1,0) and ARIMA (1,1,1), with AICs of 45.94 and 47.11, respectively. All three models produced comparable RMSE values, with ARIMA(0,1,0) achieving the lowest MAE (0.2214) and MAPE (2.73%). These results indicate that ARIMA (0,1,0) has better predictive accuracy, as summarised in Table 2. The ACF plot of residuals for ARIMA(0,1,0) in Fig. 5 shows no significant autocorrelation, and the Ljung-Box test produced a *p*-value of 0.514, validating the white noise assumption. Thus, ARIMA(0,1,0) was selected as the most suitable model for forecasting the post-pandemic series.

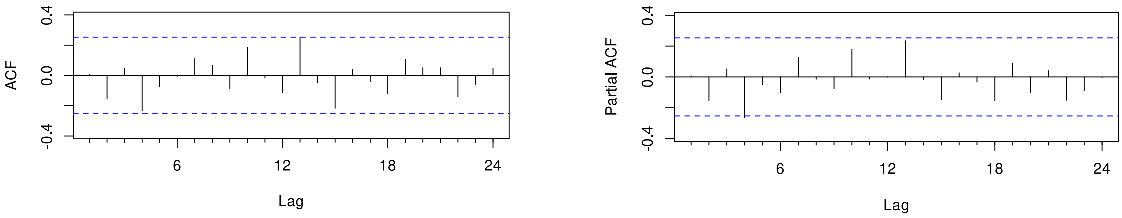

Fig. 4. ACF and PACF plots of the differenced unemployment rate for the post-pandemic period

Table 2
ARIMA Model Comparison (post-pandemic)

Model	AIC	RMSE	MAE	MAPE (%)
ARIMA(1,1,0)	45.94	0.3404	0.2228	2.75
ARIMA(1,1,1)	47.11	0.3379	0.2273	2.82
ARIMA(0,1,0)	43.96	0.3404	0.2214	2.73

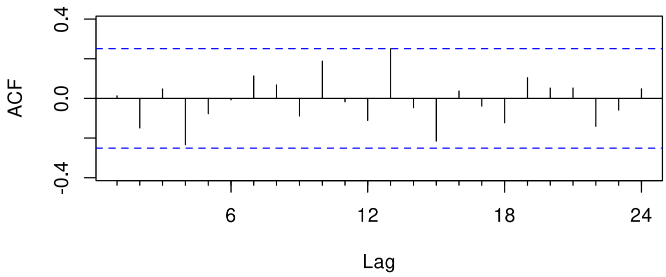

Fig. 5. ACF plot of residuals from the ARIMA(0,1,0) model (post-pandemic)

3.3 12-Month Forecast of the Unemployment Rate: Pre-Pandemic and Post-Pandemic

Following confirmation that the residuals were uncorrelated and exhibited white noise behaviour, forecasting was conducted using the best-fitting model for each period. For the pre-pandemic period, the selected model was ARIMA(0,1,1), as shown in (5), while for the post-pandemic period, the best-fitting model was ARIMA(0,1,0), as shown in (6).

$$y_t = y_{t-1} + \varepsilon_t - 1.0\varepsilon_{t-1} \tag{5}$$

$$y_t = y_{t-1} + \varepsilon_t \tag{6}$$

These models were used to generate 12-month forecasts of Malaysia's unemployment rate, reflecting two distinct economic phases. The pre-pandemic model captured relatively stable dynamics with a flat forecast trend, as shown in Fig. 6, while the post-pandemic model indicated a continued decline in unemployment that began after 2020, accounting for volatility and short-term shocks in the labour market, as shown in Fig. 7.

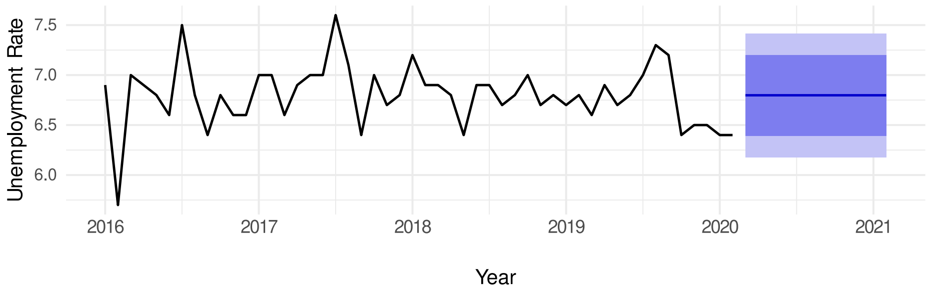

Fig. 6. Forecast of the unemployment rate for a 12-month period (pre-pandemic)

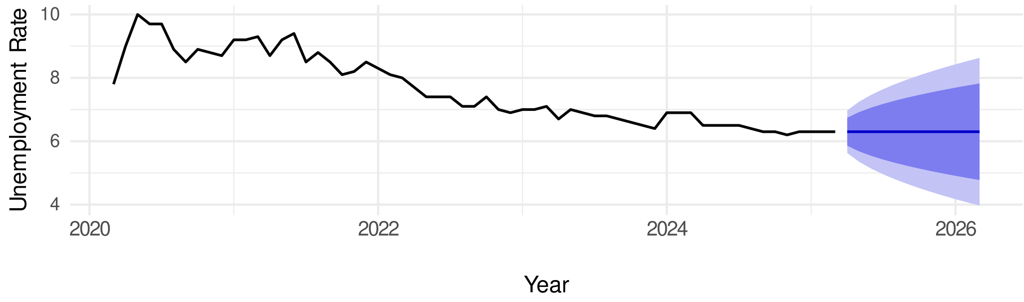

Fig. 7. Forecast of the unemployment rate for a 12-month period (post-pandemic)

4. Conclusion

This applies ARIMA to model and forecast Malaysia's unemployment rate, with emphasis on the structural break in early 2020 due to COVID-19. Results show distinct shifts between the pre- and post-pandemic periods, with ARIMA(0,1,1) and ARIMA(0,1,0) identified as the best-fitting models, respectively. Before the pandemic, unemployment was stable within a moderate range, while COVID-19 caused a sharp rise followed by a gradual decline, indicating economic recovery. The 12-month forecast beyond March 2025 suggests a steady unemployment rate. These findings align with SDG 8, offering insights for policies that promote sustainable employment and economic resilience, and highlight the need for future research exploring alternative forecasting approaches, particularly machine learning techniques such as random forests, gradient boosting, and recurrent neural networks (RNNs), to capture non-linear patterns and incorporate a wider set of macroeconomic variables.

Acknowledgement
This research received no specific funding. The authors gratefully acknowledge support from the Malaysia Official Open Data Portal for providing secondary data for this study.

References

[1] Balcı, Ali İhsan. "THE ANALYSIS OF YOUTH UNEMPLOYMENT IN TÜRKİYE UNDER THE IMPACT OF COVID-19 WITHIN THE FRAMEWORK OF UNEMPLOYMENT HYSTERESIS AND THE NATURAL RATE OF UNEMPLOYMENT." *Ömer Halisdemir Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi* 18, no. 1 (2025): 411-430. doi: 10.25287/ohuiibf.1560055.

[2] G. E. P. Box and G. M. Jenkins, "Time Series Analysis Forecasting and Control.," Operational Research Quarterly (1970-1977), vol. 22, no. 2, p. 199, Jun. 1971, doi: 10.2307/3008255.

[3] Ismail, Nur Afiqah, Nurin Alya Ramzi, and Pauline Jin Wee Mah. "Forecasting the unemployment rate in Malaysia during Covid-19 pandemic using ARIMA and ARFIMA models." *Malaysian Journal of Computing (MJoC)* 7, no. 1 (2022): 982-994.

[4] Ramli, Suraya Fadilah, Muhammad Firdaus, Hazmi Uzair, Muhammad Khairi, and Amirul Zharif. "Prediction of the unemployment rate in Malaysia." *International Journal of Modern Trends in Social Sciences* 1, no. 4 (2018): 38-44.

[5] Ashour, Marwan Abdul Hameed, and Rabab Alayham Abbas Helmi. "Predicting the youth unemployment rate in Iraq until 2035 using artificial intelligence." In *2024 14th International conference on system engineering and technology (ICSET)*, pp. 73-77. IEEE, 2024. doi: 10.1109/ICSET63729.2024.10774860.

[6] Gong, Junzhe, and Chun-Te Lee. "Research on unemployment rate based on machine learning method: a case study of United States from 1976 to 1986." In *International Conference on Cyber Security, Artificial Intelligence, and Digital Economy (CSAIDE 2022)*, vol. 12330, pp. 282-296. SPIE, 2022. doi: 10.1117/12.2647799.

[7] Nirmala, A. P., Arpana Prasad, and Dinesh Babu. "Predictive analysis of Unemployment rate Using Machine Learning Techniques." In *2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS)*, pp. 1-5. IEEE, 2023.doi: 10.1109/ICICACS57338.2023.10100121.

[8] Gogas, Periklis, Theophilos Papadimitriou, and Emmanouil Sofianos. "Forecasting unemployment in the euro area with machine learning." *Journal of Forecasting* 41, no. 3 (2022): 551-566. doi: 10.1002/for.2824.

[9] Luitel, Hari S., and Gerry J. Mahar. "A short note on the application of chow test of structural break in US GDP." *International Business Research* 8, no. 10 (2015): 112. doi: 10.5539/ibr.v8n10p112.

[10] Allaro, Hailegiorgis Bigramo, Belay Kassa, and Bekele Hundie. "A time series analysis of structural break time in the macroeconomic variables in Ethiopia." *African Journal of Agricultural Research* 6, no. 2 (2011): 392-400.

[11] Muthuramu, P., and T. Uma Maheswari. "Tests for structural breaks in time series analysis: A review of recent development." *Shanlax International Journal of Economics* 7, no. 4 (2019): 66-79.doi: 10.34293/economics.v7i4.628.

[12] Ericsson, Neil R., Mohammed HI Dore, and Hassan Butt. "Detecting and quantifying structural breaks in climate." *Econometrics* 10, no. 4 (2022): 33. doi: 10.3390/econometrics10040033.

[13] Hall, Stephen G., George S. Tavlas, Lorenzo Trapani, and Yongli Wang. "On the detection of structural breaks: The case of the Covid shock." *Journal of Forecasting* 44, no. 3 (2025): 1042-1070.doi: 10.1002/for.3238.

[14] Andrews, Donald WK. "Tests for parameter instability and structural change with unknown change point." *Econometrica: Journal of the Econometric Society* (1993): 821-856.doi: 10.2307/2951764.

[15] Zivot, Eric, and Donald W. K. Andrews. "Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis." *Journal of business & economic statistics* 20, no. 1 (2002): 25-44. doi: 10.1080/07350015.1992.10509904.

[16] Chow, Gregory C. "Tests of equality between sets of coefficients in two linear regressions." *Econometrica: Journal of the Econometric Society* (1960): 591-605.

[17] Osborne, Jason. "Improving your data transformations: Applying the Box-Cox transformation." *Practical Assessment, Research, and Evaluation* 15, no. 1 (2010).

[18] Dickey, David A., and Wayne A. Fuller. "Likelihood ratio statistics for autoregressive time series with a unit root." *Econometrica: journal of the Econometric Society* (1981): 1057-1072.

[19] Y. K. Ng, "Youth Unemployment in Malaysia: Trends, Issues and Policy Recommendations," Singapore, 2021. Accessed: Aug. 10, 2025. [Online]. Available: https://www.iseas.edu.sg/wp-content/uploads/2021/08/ISEAS_Perspective_2021_116.pdf

[20] Lim, Beatrice Fui Yee. "Impact of the COVID-19 pandemic on Malaysian labour market and policy responses." *Japan Labor Issues* 6, no. 40 (2022): 37-44.