Journal of Quality Measurement and Analysis *JQMA* 20(3) 2024, 49-64
e-ISSN: 2600-8602 *http://www.ukm.my/jqma*
https://doi.org/10.17576/jqma.2003.2024.04

URBAN $PM_{2.5}$ POLLUTION DYNAMICS IN PETALING JAYA, MALAYSIA: A TEMPORAL APPROACH

(Dinamik Pencemar $PM_{2.5}$ Bandar di Petaling Jaya, Malaysia: Suatu Pendekatan Temporal)

ZAINOL MUSTAFA, AMINA BELAL*, AHMED MAMI & MAS NORDIANA RUSLI

ABSTRACT

In 2021, Malaysia experienced a 25% increase in fine particulate matter ($PM_{2.5}$) concentrations compared to 2020. During this period, Petaling Jaya was recognised as one of the most polluted cities in the country. The study intended to investigate the dynamics of daily average concentrations of particulate matter smaller than 2.5 micrometres ($PM_{2.5}$) and other air pollutants with notable significant levels in 2021 in Petaling Jaya, Malaysia, for the year 2021 relative to the levels in 2020. To achieve this, an autoregressive distributed lag (ARDL) model was employed. Results from the paired sample *t*-test indicated sulphur dioxide (SO_2) as having significantly higher concentrations in 2021 compared to 2020. The ARDL bound test established a long-term association between SO_2 and $PM_{2.5}$. The Augmented Dickey (ADF) unit root test supported the suitability of the ARDL model by demonstrating variable integration at different levels. The ARDL model analysis revealed that SO_2 had a significant long-term negative impact on $PM_{2.5}$, while exhibiting a significant effect in the short term. An adjustment speed of 34% indicated that the system could rectify approximately one-third of any deviation from the long-term equilibrium between SO_2 and $PM_{2.5,}$ one day following a disturbance. Various reasons could be cited for the discrepancies in model performance across different time frames and pollutants, such as seasonal fluctuations, changes in human activities, adjustments to regulations, and external influences. This study provides crucial insights into the dynamic interactions between air pollutants and contributes to more effective air quality management strategies.

Keywords: $PM_{2.5}$; Petaling Jaya; paired sample *t*-test; autoregressive distributed lags model

ABSTRAK

Pada tahun 2021, Malaysia mengalami peningkatan sebanyak 25% pada kepekatan partikel terampai halus ($PM_{2.5}$) berbanding tahun 2020. Pada waktu tersebut, Petaling Jaya dikenalpasti sebagai salah satu bandar paling tercemar dalam negara. Kajian ini bertujuan untuk menyiasat dinamik kepekatan purata harian partikel terampai bersaiz kurang daripada 2.5 micrometer ($PM_{2.5}$) dan pencemar udara lain yang mempunyai aras yang ketara pada tahun 2021 di Petaling Jaya, Malaysia, untuk tahun 2021 relatif kepada aras pada tahun 2020. Untuk tujuan ini, model autoregresif lat tertabur (ARDL) telah digunakan. Hasil daripada ujian-*t* sampel berpasangan menunjukkan yang sulfur dioksida (SO_2) mempunyai kepekatan tinggi yang signifikan pada tahun 2021 berbanding tahun 2020. Ujian terikat ARDL memantapkan perkaitan jangka masa panjang antara SO_2 dan $PM_{2.5}$. Ujian Augmented Dickey (ADF) menyokong kesesuaian model ARDL dengan menunjukkan integrasi pembolehubah pada pelbagai aras. Analisis model ARDL menunjukkan SO_2 mempunyai impak negatif jangka masa panjang yang signifikan kepada $PM_{2.5}$, sambil mempamerkan kesan signifikan dalam jangka masa pendek. Pelarasan kepantasan pada 34% menunjukkan bahawa sistem ini mampu membetulkan lebih kurang satu pertiga daripada sebarang penyimpangan daripada keseimbangan jangka masa panjang diantara SO_2 and $PM_{2.5,}$ sehari selepas gangguan. Pelbagai punca boleh dipetik berkenaan percanggahan pada prestasi model merentasi pelbagai jangka masa dan pencemar, contohnya turun-naik musiman, perubahan aktiviti manusia, pelarasan peraturan, dan pengaruh luar. Kajian ini memberi pandangan penting kepada interaksi dinamik antara pencemar udara dan menyumbang kepada pengurusan strategi kualiti udara yang lebih efektif.

Keywords: $PM_{2.5}$; Petaling Jaya; ujian-*t* sampel berpasangan; model autoregresif lat tertabur

1. Introduction

Particle air pollution, which is widely known as particulate matter pollution (PM), is a critical global health issue, causing an estimated four million deaths annually (Thangavel *et al*. 2022). Fine particulate matter ($PM_{2.5}$), with its ability to penetrate deep into the respiratory system and bloodstream, poses severe health risks, including respiratory and cardiovascular diseases, and cancer (World Health Organisation 2021). In urban settings, vehicular and industrial emissions are major sources of $PM_{2.5}$, and their impact on air quality and public health is exacerbated under certain conditions such as the COVID-19 pandemic. Studies, including those conducted during the lockdown phases of the pandemic, have observed significant variations in air pollutant levels, thereby demonstrating the influence of human activity on air quality (Abdullah *et al*. 2020; Zulkarnain *et al.* 2023).

Fine particulate matter ($PM_{2.5}$) is one of six air pollutants recorded by the Department of Environment of Malaysia (DOE). It is widely regarded as the most detrimental to both the environment and human health because of its widespread presence and extensive range of health effects (Darunikorn *et al*. 2023). Exposure to $PM_{2.5}$, which presents a substantial health threat, with outdoor $PM_{2.5}$ emerging as the most critical environmental determinant of mortality in the area, linked to 130,000–320,000 additional deaths in ASEAN countries in 2019 (Ravi *et al*. 2022). Studies in the Asia-Pacific region have repeatedly demonstrated connections between extended exposure to $PM_{2.5}$, heightened all-cause mortality, cardiovascular disease, type 2 diabetes mellitus, kidney disease, and chronic obstructive pulmonary disease (Nguyen *et al*. 2022).

The significant population growth and corresponding economic development have been major contributors to the rising levels of air pollution in Southeast Asia. Fossil fuels, particularly oil and coal, are the primary sources of fuel in the power sector, and the demand for electricity is increasing at a rate of approximately 6% per year. The combustion of these fuels is a significant contributor to $PM_{2.5}$. Other sources of $PM_{2.5}$ emissions in urban areas include construction, industrial activities, and transportation. In rural areas, open burning practices used for farmland management and forest clearing also contribute to $PM_{2.5}$ emissions (IQAir 2022).

Several studies have pointed out various sources of $PM_{2.5}$. These sources include natural sources, such as biomass burning, which makes a significant contribution during the southwestern monsoon season owing to Indonesian peatland fires (Dahari *et al*. 2019; Suradi *et al.* 2021). Additionally, urban traffic combustion, industrial activities, and motor vehicles have been identified as primary sources of $PM_{2.5}$ in urban areas such as Kuala Lumpur, contributing to particulate air pollution (Chow *et al*. 2019; Rahman *et al.* 2015). Long-range transportation of $PM_{2.5}$, which impacts pollution levels in areas such as Skudai, Johor Bahru, has also been observed in regions such as Sumatera, Indonesia, and China during different monsoon seasons (Fujii *et al*.2015).

Petaling Jaya, located in Kuala Lumpur, the capital city of Malaysia, is facing substantial air pollution challenges. Despite the existence of industrial operations and open burning, the automotive industry has emerged as the main contributor to hazardous emissions in the region (Ganeshwaari & Koshy 2022). A study by Amil *et al.* (2016) revealed the significant impact of fine particulate matter $PM_{2.5}$ in the air of Petaling Jaya's urban-industrial district in the Klang Valley. The average mass concentration of $PM_{2.5}$ was 28±18 μg^3, approximately threefold the annual guidelines of the World Health Organization (WHO).

Many academic studies have focused on Petaling Jaya's urban environment, its population density, and air pollution. Shafie *et al.* (2022) employed geographic methodologies to analyze the distribution and impacts of air pollutants, aiming to understand effects of urban air pollution on Klang Valley residents. Amir (2007) provided a historical analysis of air pollution trends in Petaling Jaya, offering insights into the changing urban air quality. Moreover, Hadipour *et al.* (2009) highlighted the critical role of integrated urban environmental planning by developing mathematical models to evaluate the environmental impacts of transportation-related air pollution in Petaling Jaya. These investigations highlight the significant consequences of urbanisation and air pollution on the health of the population and Petaling Jaya's environmental integrity.

This study addresses the gap in understanding the dynamics of $PM_{2.5}$ concentrations and their association with other air pollutants demonstrated highly significant concentrations indentified by the application of paired-sample *t*-test in 2021. The Autoregressive Distributed Lag (ARDL) model was employed to investigate the short-term and long-term effects of air pollutants utilised by the Department of Environment, Malaysia, which exhibited a significant increase in 2021 compared with 2020 on $PM_{2.5}$, in Petaling Jaya, Malaysia. This study contributes to the empirical body of knowledge on air pollution dynamics. This study provides actionable insights for policymakers and environmental authorities to mitigate air quality issues.

The paper is organized as follows: following the introduction, the definition of fine particulate matter ($PM_{2.5}$) is provided, along with an overview of some existing studies on its health effects and source emissions. The materials and methods section discusses the study area and dataset collection and treatment. This is followed by a description of the methodology, which outlines the analytical methods and statistical techniques employed in the study. The empirical results section presents the findings from the applied techniques and model analysis, including the effects of various pollutants on $PM_{2.5}$ levels. The discussion section interprets these results in the context of existing literature and policy implications. Finally, the conclusion summarizes the key findings of the research.

2. Materials

2.1. *Study area*

This study's research area, Petaling Jaya which is located in the Klang Valley with the coordinate of 3° 08'N latitude and 101° 44'E longitude, is the primary city in the Selangor Darul Ehsan state and is situated in the Petaling District. It has a total area of 97.2 km² and is overseen by the Petaling Jaya City Council (PJCC), which serves as the local governing body. The selection of Petaling Jaya as the focus of this study was influenced by its inclusion by IQAir (2022), who identified it as having high levels of $PM_{2.5,}$ following the Klang Valley region who identified it as having high $PM_{2.5}$ concentrations after the Klang Valley region. As of July 2022, Petaling Jaya had a population of over 619,925 people and 278,800 properties, and it currently stands as the most significant growth centre in Selangor (Rosli *et al.* 2023).

2.2. *Datasets collection and treatment*

Air quality datasets for Petaling Jaya continuous air quality stations were collected from the Air Quality Division of the Department of Environment of Malaysia (DOE). Daily average concentration datasets consisted of air pollutants: $PM_{2.5}$ (µg/m3), SO_2 (ppm), NO_2 (ppm), CO (ppm), and O_3 (ppm) over a period of two years in 2020 and 2021.

The total number of data points utilised in this analysis for 2020 was 1830 (5 variables × 366 days). Similarly, the total number of data points for 2021 was 1825 (five variables × 365 days). From the overall dataset, the total number of missing data points was approximately 6%. To address the issue of unavailable or missing data, the *k*-nearest neighbour algorithm (KNN) was implemented, and computations were carried out using the XLSTAT add-in software. This method has been widely applied in air pollution studies (Azid *et al.* 2013; 2014; Islam *et al.* 2022).

3. Methodology

3.1. *Paired sample t-test*

A paired sample *t*-test analysis was employed to evaluate the average of two variables within a group. This technique is beneficial for examining and comparing the average of the samples both before and after treatment or for a specific period of time, this test is the most suitable for use in circumstances where the underlying distribution is normal and the sample sizes are substantial for any distribution (Imam *et al.* 2014). To determine a sufficiently large t count before and after treatment, the following equation was used:

$$t = \frac{\sum d}{\sqrt{\frac{n\,(\sum d^2) - (\sum d)^2}{n-1}}}$$

Whereas:
d = mean difference per paired value
n = number of samples

To conduct a paired-sample *t*-test, the following assumptions must be met:
(1) The differences between the values obtained at two different times must be normally distributed.
(2) Values were sampled independently.
(3) Values were measured on an interval or ratio scale.
(4) The data consisted of related pairs of values from two different times, with each measurement at one time paired with the corresponding measurement at the other time.

For the paired sample *t*-test, the degrees of freedom were *n-1*. If the calculated *t*-value is less than the *t*-table, the null hypothesis H_0, which states that no significant difference was observed, is rejected, and the alternative hypothesis H_1, which states that there is a significant difference between the two groups, is accepted. Conversely, if the calculated t-value is greater than the *t*-table, the null hypothesis is accepted and the alternative hypothesis is rejected.

The daily average $PM_{2.5}$ concentrations from 2020 and 2021 were compared using a paired sample *t*-test. The null hypothesis (H_0, the mean difference between the 2021 and 2020 paired daily $PM_{2.5}$ levels is zero) and the alternative hypothesis (H_1, the true mean difference between the 2021 and 2020 daily $PM_{2.5}$ levels deviates from zero, which indicates an increase) were explored using the paired sample *t*-test. Case–control studies or repeated measures studies frequently used the paired sample *t*-test. This test was also used for the variable selection criteria. Variables with significantly increased mean levels in 2021 compared to 2020 were considered for the analysis. This approach ensured that the mean levels of the included variables were characterised by a meaningful and statistically supported increase. This approach enhanced the relevance of the analysis in identifying temporal trends and changes in $PM_{2.5}$.

Subsequently, a descriptive statistical analysis was performed on the obtained higher significant air pollutants for 2021, including the mean, median, first quartile, third quartile, maximum, minimum, and standard deviation. These calculations were performed to investigate the distributions of the obtained variables. The Pearson correlation coefficient was used to evaluate the linear relationship between significant variables.

3.2. *Augmented Dickey-Fuller (ADF) unit root test*

In the field of time series analysis, it is widely known that nonstationary data series may lead to spurious regression. Therefore, testing for stationarity is an essential step in both forecasting and dynamic modeling (Özcan & Öztürk 2019). This is important to ensure that the results of the analysis are valid and accurate. Additionally, autoregressive distributed lag (ARDL) estimation is required to test for unit root to ensure that all variables satisfy the underlying assumptions that data series are integrated at different levels. Therefore, time-series properties of the variables were tested with the application of Augmented Dickey-Fuller (ADF)

3.3. *Dynamic Autoregressive Distributed Lags (ARDL) approach*

This research aimed to investigate the dynamic relationship between daily average $PM_{2.5}$ concentrations, and any of the air pollutants sulphur dioxide (SO_2), nitrogen dioxide (NO_2), carbon monoxide (CO), or ground level ozone (O_3), which displayed higher daily average concentrations in 2021 than in 2020, using the paired sample *t*-test. The model is specified in the functional form as follows:

$$PM_{2.5} = f\ (SO_2, NO_2, CO, O_3) \tag{1}$$

This can be expressed in econometric form as:

$$PM_{2.5,t} = \beta_0 + \beta_1 SO_{2,t} + \beta_2 NO_{2,t} + \beta_3 CO_t + \beta_4 O_{3,t} + \varepsilon_t \tag{2}$$

The model below has been designed to convert all variables in Eq. (2) into their log form:

$$\log PM_{2.5,t} = \beta_0 + \beta_1 \log SO_{2,t} + \beta_2 \log NO_{2,t} + \beta_3 \log CO_t + \beta_4 \log O_{3,t} + \varepsilon_t \tag{3}$$

where $logPM_{2.5\,,t}$ represents the daily average of $PM_{2.5}$ concentrations, *t* represents the time period from January 1, 2021 to December 31, 2021. β_0 represents the constant (intercept) while β_1 to β_4 are the coefficients of regressors, while ε_t represents the error term or the white noise.

The long-term associations known as cointegration among the variables of interest were empirically examined using the ARDL bound test approach (Pesaran *et al.* 2001). This method was selected as the bounds test procedure that possesses several advantages over the traditional cointegration methods (Ali *et al*. 2021). Furthermore, this approach enables the estimation of the cointegration relationship using ordinary least squares (OLS) to identify the appropriate model lag order. Other multivariate cointegration approaches, such as that of Johansen-Juselius (1990), involve variable pre-testing for unit roots. In contrast, bounds testing does not require pre-testing of the unit root model variables. The test is valid irrespective of whether the underlying regressors in the model are purely I(0) or I(1) (Özcan & Öztürk 2019). Third, the test was comparatively more efficient with finite or small sample sizes, such as the sample size used in this study. Nevertheless, the test may fail in the presence of the I(2) series (Chandio *et al*. 2019). Eq. (3) can be written in ARDL form as follows:

$$\begin{aligned}\Delta\, log\, PM_{2.5,t} = \alpha_0 + \alpha_1 \Delta\, log\, PM_{2.5,t-1} + \alpha_2\, \Delta log SO_{2,t-1} + \alpha_3 \Delta\, log\, NO_{2,t} \\ + \alpha_4 \Delta\, log\, CO_t + \alpha_5 \Delta log O_{3,t} + \sum_{i=1}^{q} \gamma_{1i} \Delta log PM_{2.5,t-i} \\ + \sum_{i=0}^{p} \gamma_{2i} \Delta log SO_{2,t-i} + \sum_{i=0}^{p} \gamma_{3i} \Delta log NO_{2,t-i} \\ + \sum_{i=0}^{p} \gamma_{4i} \Delta log CO_{2,t-i} + \sum_{i=0}^{p} \gamma_{5i} \Delta log O_{3,t-i} + \varepsilon_t\end{aligned} \quad (4)$$

where Δ is the first difference function. p and q represents the obtained optimal lag. The long-term dynamics are characterized by α_1, α_2, α_3, and 4, while the short-term dynamics are represented by γ_1, γ_2, γ_3, γ_4 and γ_5. and q and p are the lag periods. Furthermore, the joint significance of the lagged variable coefficient was assessed using the F-test to confirm a long-term relationship between $PM_{2.5}$ and the obtained significant air pollutants. H_0 suggests the absence of a long-term relationship between $PM_{2.5}$ and any significant air pollutants, and assumes that the coefficients have identical values. Thus, H_0: $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5$ was assessed according to Pesaran et al.(2001). An F-test value exceeding the upper critical bound does not support H_0 and the variables exhibit cointegration. Conversely, an F-test value below the lower critical bound supports H_0, and indicates no variable cointegration. An inconclusive result was obtained when the F-test value was between the upper and lower critical bounds. The error correction model can be estimated according to the following formula in Eq. (5):

$$\begin{aligned}\Delta\, log\, PM_{2.5,t} = \alpha_0 + \sum_{i=1}^{q} \gamma_{1i} \Delta log PM_{2.5,t-i} + \sum_{i=0}^{p} \gamma_{2i} \Delta log SO_{2,t-i} \\ + \sum_{i=0}^{p} \gamma_{3i} \Delta log NO_{2,t-i} + \sum_{i=0}^{p} \gamma_{4i} \Delta log CO_{2,t-i} \\ + \sum_{i=0}^{p} \gamma_{5i} \Delta log O_{3,t-i} + \delta ECT_{t-1} + \varepsilon_t\end{aligned} \quad (5)$$

where ECT_{t-1} represent a one lagged error correction term and δ represent the speed coefficient of adjustments towards long-run equilibrium, in other words, if the system is moving out of equilibrium in one direction then it will pull it back to equilibrium.

The ARDL error correction term is a vital component of the model that determines the adjustment speed of the variables to long-term equilibrium after a shock. This term aids in the identification of the long-term relationship between variables and corrects for short-term deviations from that relationship. The term is calculated as the lagged difference coefficient of the dependent variable in the ARDL model, and indicates a cointegrating relationship among the variables. Such a relationship indicates a long-term equilibrium that is sustained even if the variables deviate from it in the short term. Furthermore, the term ensures appropriate model specification and reliably estimates the long-term relationship between the variables (Shuaibu *et al*. 2022). A statistically significant and negative coefficient (ECT_{t-1}) indicates that a long-term imbalance between $PM_{2.5}$ and obtained significant variables obtained by the paired sample *t*-test will converge to the association of long-term equilibrium. To confirm the stability of the model, Breusch-Godfrey serial correlation LM, Heteroskedasticity by Breusch-Pagan-Godfrey, and Jarque–Bera normality tests along with CUSUM proposed by Brown *et al*. (1975) have been employed to conduct a comprehensive evaluation of the model's stability.

3.4. *CUSUM test*

After the long-run relationship between variables under consideration is confirmed, the research applies the Cumulative Sum (CUSUM) test. Earlier literature by (Pesaran *et al.* 2001) proposed this evaluation to indicate the appropriate fit of the ARDL model. This examination involves plotting the residuals of the ECM. If the statistics in the plot fall within critical limits at the 5% significance level, it suggests that the coefficients of the ARDL model are stable.

4. Results and Discussions

4.1. *Summary statistics*

The summary statistics (see Table 1) present an overview of the statistical distribution of $PM_{2.5}$ concentrations in 2020 and 2021. In 2020, the mean $PM_{2.5}$ concentration was 19.59±6.70 μg/m^3, which was below the Malaysian Ambient Air Quality Standards (MAAQS) that defines the permissible $PM_{2.5}$ threshold of 35 μg/m^3, this indicates the concentration was in compliance with regulatory limits.

Additionally, the first and third quartiles of $PM_{2.5}$ concentrations in 2020 were 15.16 μg/m^3 and 23.26 μg/m^3, respectively indicating 75% of the daily average concentrations did not exceed the MAAQS. The median, minimum (Min), and maximum (Max) daily $PM_{2.5}$ concentrations in 2020 were 18.58 μg/m^3, 7.60 μg/m^3, and 47.74 μg/m^3, respectively. The standard deviation (SD) of 6.67 μg/m3 indicated moderate variability.

Table 1: Summary statistics of $PM_{2.5}$ concentrations (μg/m^3) in 2020–2021

Variable	Mean	First quartile	Median	Third quartile	Max.	Min.	SD
$PM_{2.5}$ 2020	19.59	15.16	18.58	23.26	47.74	7.60	6.70
$PM_{2.5}$ 2021	24.08	18.86	23.12	28.14	55.60	6.64	7.41

The mean $PM_{2.5}$ concentration in 2021 was 24.08±7.41 μg/m^3, which reflected an increase compared to 2020. The quartile, and maximum values also showed similar increasing patterns. The SD in 2021 was 7.41 μg/m^3, which is marginally higher than that in 2020. Figure 1 depicts the time series graph of $PM_{2.5}$ concentrations in 2020 and 2021. The $PM_{2.5}$ concentrations in the air significantly increased in 2021.

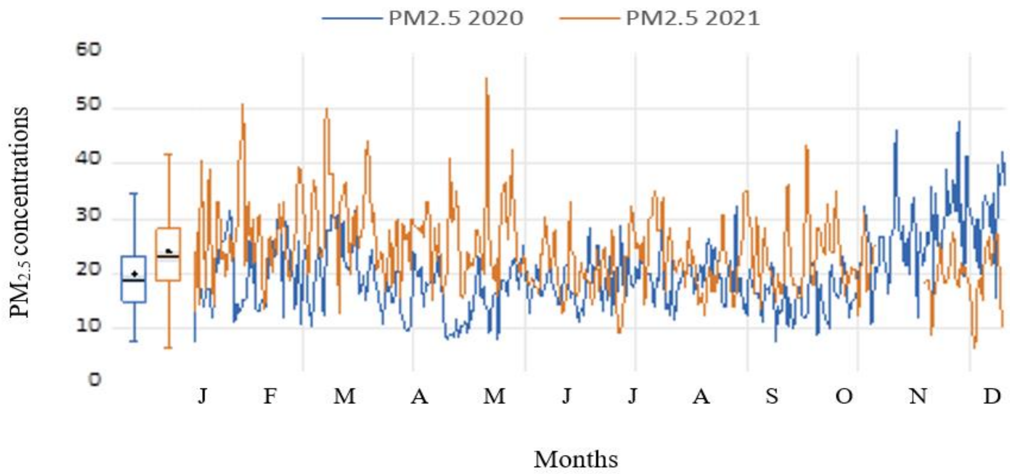

Figure 1: Visual representation of $PM_{2.5}$ concentrations in 2020–2021

4.2. *Paired sample t-test results*

Given that our dataset consists of 365 daily average concentrations for each air pollutant ($PM_{2.5}$, SO_2, NO_2, and O_3), and considering the Central Limit Theorem (CLT) which supports that with a sample size of this size, the distribution of sample means will approach normality. Additionally, with the justification by Kwak and Kim (2017) that the skewed population distribution does not influence the distribution of sample means as sample size increases. According to the central limit theorem, with a sufficiently large sample size, the means of samples will be normally distributed, thus allowing us to apply the paired sample *t*-test.

Table 2 presents the results of the paired-sample *t*-test. The mean difference (diff) in $PM_{2.5}$ concentrations between the two years was -4.496 μg/m^3, which indicated a statistically significant increase from 19.589 μg/m^3 in 2020 to 24.085 μg/m^3 in 2021 (t = -7.95, p < 0.05). Similarly, the paired *t*-test for SO_2 concentration revealed a significant increase, with a mean difference of 0.000 (t = 9.20, p < 0.05).

Table 2: Paired sample *t*-test of 2020–2021

Variable	Mean 2020	Mean 2021	diff	Standard error	*t*-value	*p*-value
$PM_{2.5}$ 2021 – $PM_{2.5}$ 2020	19.59	24.09	-4.50	0.57	-7.95	0.00
SO_2 2021 – SO_2 2020	0.001	0.001	0.000	0.000	9.20	0.00
NO_2 2021 – NO_2 2020	0.018	0.019	-0.001	0.001	-1.70	0.09
CO 2021 – CO 2020	0.825	1.018	-0.193	0.020	-9.55	1.00
O_3 2021 – O_3 2020	0.005	0.010	-0.004	0.001	-12.15	1.00

NO_2 exhibited insignificant results from 2020 to 2021 (mean difference = -0.001, t = -1.70, p > 0.05). In contrast, CO concentrations decreased significantly (mean difference = -0.193, t = -9.55, p > 0.05). Additionally, the O_3 concentrations markedly decreased (mean difference = -0.004, t = -12.15, p > 0.05). These findings emphasise the significant temporal shifts in air quality parameters, where the 2021 $PM_{2.5}$ and SO_2 concentrations were notably higher in 2021.

The daily mean $PM_{2.5}$ concentration in 2021 was 24.08 ±7.41 μg/m^3 (see Table 3). The minimum and maximum concentrations were 6.64 μg/m^3 and 55.60 μg/m^3, respectively. This suggests that $PM_{2.5}$, which exceeded the MAAQS, reached a maximum value of 55.60 µg/m^3 that exceeded the MAAQS.

Table 3: Summary statistics of $PM_{2.5}$ and SO_2 daily mean levels in 2021

Variable	Mean	First quartile	Median	Third quartile	Max	Min	SD
$PM_{2.5}$ (µg/m^3)	24.08	18.86	23.124	28.14	55.60	6.64	7.41
SO_2 (ppm)	1.176×10^{-3}	1×10^{-3}	1.1×10^{-3}	0.001	3×10^{-3}	1×10^{-3}	4.5×10^{-4}

Additionally, in 2021, the daily average concentration of SO_2 exhibited a maximum value of 3×10^{-3}ppm. The first quartile value of 1×10^{-3} ppm indicates that 25% of the days had relatively low SO_2 levels, while the third quartile value of 1.176×10^{-3} ppm indicated that 75% of the days had SO_2 levels below this threshold, highlighting a tendency towards lower concentrations. The standard deviation of 8.9×10^{-4} ppm reflects some variability in daily SO_2 levels, indicating fluctuations but not extreme differences. Overall, the data suggest that, while there were variations, the majority of days had relatively low SO_2 concentrations.

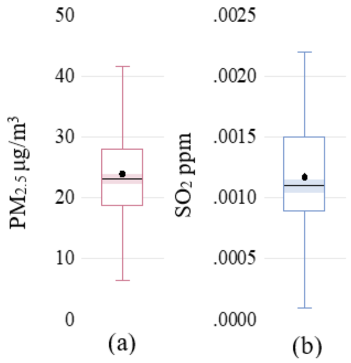

Figure 2: Boxplot analysis on daily average concentrations of (a) $PM_{2.5}$ and (b) SO_2 in Petaling Jaya station in Malaysia

The median of 23.124 µg/m^3 of $PM_{2.5}$ concentration is slightly lower than the mean of 24.08 µg/m^3, which indicates a right-skewed distribution of the data, as shown in Figure 2. This distribution suggests that, although a majority of the observations cluster around or below the mean, there exists a subset of relatively higher values that disproportionately influence the mean, causing it to be slightly higher than the median. Practically, this discrepancy implies that the average $PM_{2.5}$ concentration may be fairly inflated due to occasional spikes or outliers in pollution levels, which may have significant implications for public health and environmental policy.

Similarly, in 2021, the daily average SO_2 concentration exhibited a slight excess over the median, with values of 1.176×10^{-3} and 1.1×10^{-3} ppm, respectively. This suggested a skewed distribution at higher concentrations. This discrepancy implies that, while the majority of observations may cluster around or below the median, the presence of occasional spikes or outliers in SO_2 levels leads to an upward shift in the mean.

4.3. *Correlation matrix*

The correlation matrix in Table 4 demonstrates the relationship between $PM_{2.5}$ and SO_2 concentrations in 2021. The correlation coefficient of -0.112 suggested a weak and statistically significant negative correlation between $PM_{2.5}$ and SO_2 at the 5% significance level.

Table 4: Correlation matrix of $PM_{2.5}$ and SO_2 daily mean levels in 2021

	$PM_{2.5}$	SO_2
$PM_{2.5}$	1	
SO_2	-0.112*	1

* significance at the 5% level.

This result implied that the SO_2 daily concentrations tended to decrease or increase as the daily concentrations of $PM_{2.5}$, increased or decreased, respectively. The negative correlation suggested that $PM_{2.5}$ and SO_2 levels in the examined dataset were inversely associated.

4.4. *Unit root test*

Unit root test was conducted to assess the stationarity properties of the variables $PM_{2.5}$ and SO_2 at the level and the first difference. The tests examined H_0, assuming a unit root is present in a time series (the time series dataset does not exhibit stationarity), while H_1 assumes that no unit root is observed in the tested time-series dataset.

Table 5: Agumented Dicky-Fuller (ADF) test results

	$PM_{2.5}$	SO_2
Constant	-10.5594***	-3.7373***
Constant and trend	-7.8266***	-4.7311***
No constant and trend	-0.6477**	-1.2112

Notes: *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

The ADF results in Table 5 indicate that $PM_{2.5}$ and SO_2 are stationary at the level when including a constant and trend. Accordingly, the highly negative test statistics did not support H_0 of the unit root.. The stationarity results were less conclusive without a constant trend, particularly for SO_2. When no constant and trend were included, SO_2 exhibited non-stationarity ($p < 5\%$ significance level). Nevertheless, SO_2 demonstrated stationarity at the first difference with a 1% significance level. Thus, $PM_{2.5}$ was I(0), while SO_2 was I(1).

4.5. *Lag selection criteria*

An appropriate variable lag order is important before using the ARDL bound test to examine the cointegration among the investigated variables. This study selects the appropriate lag order using the optimal lag order of the vector autoregression (VAR) model.

Table 6 presents the lag selection criteria based on the VAR model (VAR). According to all optimal lag criteria, it was determined that the criteria indicated that the model yielded better results at lag 2. Furthermore, a suitable lag length in the VAR methodology was confirmed using the inverse roots of the AR characteristics polynomial graph. In the graph, the dots in the unit circle confirm favourable outcomes when utilising a lag of 2 (see Figure 3).

Table 6 : VAR lag order selection criteria

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-278.0361	NA	0.019006	1.712759	1.735939	1.722008
1	-3.932746	543.1774	0.003643	0.060751	0.130291	0.088498
2	10.05482	27.54737*	0.003427*	-0.000335*	0.115566*	0.045911*
3	11.54423	2.915056	0.003480	0.015020	0.177281	0.079765
4	13.30695	3.428409	0.003528	0.028704	0.237325	0.111947
5	16.52718	6.223818	0.003545	0.033473	0.288455	0.135215
6	18.39826	3.593388	0.003592	0.046494	0.347836	0.166734
7	19.17018	1.473011	0.003663	0.066237	0.413940	0.204976

* significance at the 5% level; NA = Not Applicable.

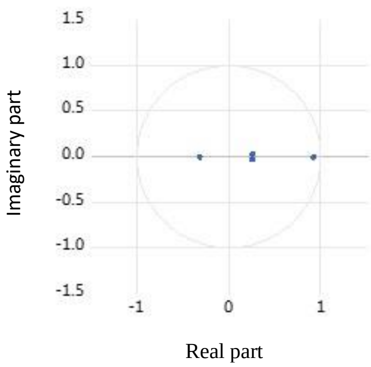

Figure 3: Inverse roots of AR characteristics polynomial graph

4.6. *ARDL bound test for cointegration*

The ARDL bound test is a new approach introduced by M.H. Pesaran and Y. Shin (Pesaran *et al.* 2001) that is applied to detect the existence of cointegration among variables over the Engle and Granger (1987), and Johansen and Juselius (1990) due to its several advantages over these traditional cointegration tests (Ali *et al.* 2021), additionally due to having different integrated variables (Gessesse & He 2020; Madaki & Akanegbu 2023). Confirming cointegration using the ARDL bound test is crucial before examining the long- and short-term relationships between the variables. The *F*-statistics values surpassed the lower and upper bounds at the 1% significance level (see Table 7). Consequently, H_1 of cointegration was supported, thus validating the long-term association between $PM_{2.5}$, and SO_2.

Table 7: ARDL bound test for cointegration

F-bounds test	Lag	*F*-statistic		*p*-value
	(2, 2)	24.26283		0.000
Critical value	10%	5%	2.5%	1%
Lower bound I(0)	2.44	3.15	3.88	4.81
Upper bound I(1)	3.28	4.11	4.92	6.02

4.7. *Dynamic ARDL technique analysis*

Upon confirming a long-term relationship between $PM_{2.5}$ (dependent variable) and SO_2 (independent variable), the long- and short-term equations were estimated using Eqs. (2) and (3) and the ARDL model (Pesaran *et al.*, 2001) (see Table 9). In the long term, SO_2 has a significant negative effect on $PM_{2.5}$. In contrast, the short-term results indicated that SO_2 had a positive and highly significant effect on $PM_{2.5}$. Thus, a 1% increase in SO_2 increased the $PM_{2.5}$ by 0.28%.

The R^2 and adjusted R^2 values were 30% and 29%, respectively, indicating an acceptable model fit. This fit could have been due to the inclusion of only SO_2 as an independent variable as indicated by Pal and Bharati (2019) variables included in a model can impact the value of R-squared, with a greater number of predictor variables possibly resulting in a higher R-squared value. The calculated F-statistic was 34.90 and highly significant at the 1% level. The results suggest that daily SO_2 levels are a significant predictor of daily $PM_{2.5}$. The error correction term $(ECT)_{t-1}$ was statistically significant and negative at the 1% level with a reasonably moderate coefficient. This result indicates that disequilibrium could be adjusted to long-term equilibrium, with 34% reasonably moderate speed due to prior shocks in the daily SO_2 levels.

Table 9: Results of ARDL long- and short-term coefficients

Variable	Coefficient	Standard error	t-Statistic	Probability
Long-term estimation of parameters from ARDL models				
Log(SO_2)	-0.456449	0.006127	-74.49400	0.0000***
Short-term estimation of parameters from ARDL models				
Log $PM_{2.5,t-1}$	-0.337108	0.048428	-6.961074	0.0000***
Log $SO_{2,t-1}$	-0.153873	0.022279	-6.906774	0.0000***
ΔLog $PM_{2.5,t-1}$	-0.005581	0.056142	-0.099400	0.9209
ΔLog SO_2	0.287034	0.058966	4.867762	0.0000***
ΔLog $SO_{2,t-1}$	0.131307	0.058971	2.226644	0.0266**
CointEq(-1)*	-0.337108	0.048321	-6.976457	0.0000***
Diagnostic tests				
R^2 = 0.301881				
Adjusted R^2 = 0.293521				
F-statistic = 34.89703***				
Breusch-Godfrey serial correlation LM test				0.5833
Heteroskedasticity test: Breusch-Pagan-Godfrey				0.8880
Jarque–Bera normality test				0.2651

Notes: *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Model stability was examined using the Breusch-Godfrey serial correlation LM, Breusch-Pagan-Godfrey, and Jarque–Bera normality diagnostic tests. The tests yielded p-values of 0.5833, 0.888, and 0.2651, respectively, indicating that the obtained model was appropriate and that the ARDL model successfully passed all the diagnostic tests. Additionally, the long- and short-term parameter stabilities were investigated using the CUSUMQ test (see Figure 4).

The stability test (Figure 4) demonstrates that the plot is between the 5% significance level critical boundaries, which confirmed the accuracy of the parameters affecting $PM_{2.5}$ concentrations throughout 2021.

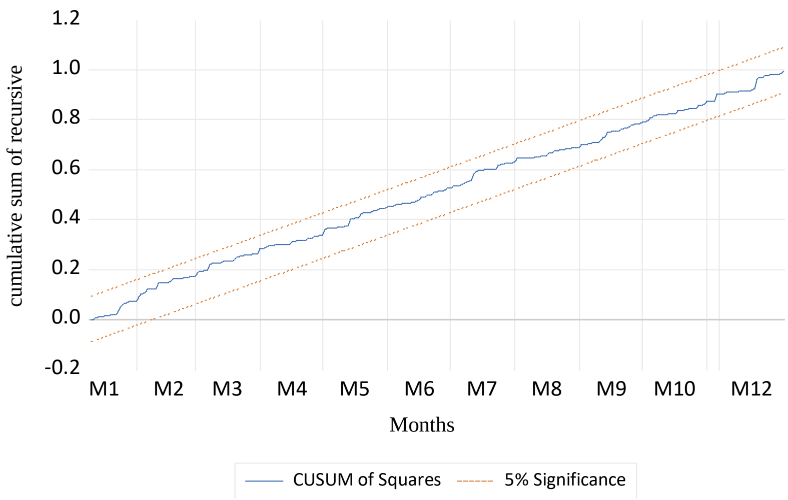

Figure 4: CUSUMQ squares of recursive residuals

5. Conclusion

This study presented convincing evidence of a significant correlation between $PM_{2.5}$ and SO_2 concentrations in Petaling Jaya in 2021. The $PM_{2.5}$ levels were significantly increased compared to those in 2020. According to the paired sample *t*-test, 2021 SO_2 levels were higher than those in 2020. ARDL identified a long-term connection between SO_2 and $PM_{2.5}$, which highlighted the significant effect of SO_2 on $PM_{2.5}$. Empirical evidence indicates that SO_2 substantially and persistently affects $PM_{2.5}$. Additionally, SO_2 significantly and positively affected daily $PM_{2.5}$ levels in the short term. Furthermore, 34% of the overall adjustment towards a new $PM_{2.5}$ equilibrium occurred during the specified period following SO_2 concentration changes. The complex interaction between SO_2 and $PM_{2.5}$ highlighted the interplay and deep connections that require additional exploration and suggested potential contributing elements. Potential factors for variations in model performance across diverse pollutants and periods may encompass changes in human activities, seasonal variations, and regulatory changes. Future investigations should focus on elucidating the intricate mechanisms that underlie the interplay between these pollutants and identifying other factors that may affect their relationships. Gaining such knowledge is critical for devising comprehensive strategies for managing air quality to ensure the protection of both public health and the environment.

6. Practical Implications

Results from this study will be highly valuable to authorities and environmental administrators in developing location-specific interventions aimed at reducing SO_2 emissions and mitigating $PM_{2.5}$ levels. These insights can inform air quality management policies, to implement stricter emission standards for industries and promoting clean energy sources. Such measures would effectively reduce the impacts of SO_2 and $PM_{2.5}$, thereby improving overall air quality.

Acknowledgement

The authors would like to thank the Department of Environment of Malaysia (DOE) for providing the air quality datasets.

References

Abdullah S., Mansor A.A., Napi N.N.L.M., Mansor W.N.W., Ahmed A.N., Ismail M. & Ramly Z.T.A. 2020. Air quality status during 2020 Malaysia Movement Control Order (MCO) due to 2019 novel coronavirus (2019-nCoV) pandemic. *The Science of The Total Environment* **729**: 139022.

Ali M.U., Gong Z., Ali M.U., Wu X. & Yao C. 2021. Fossil energy consumption, economic development, inward FDI impact on CO2 emissions in Pakistan: Testing EKC hypothesis through ARDL model. *International Journal of Finance and Economics* **26**(3): 3210–3221.

Amil N., Latif M.T., Khan M.F. & Mohamad M. 2016. Seasonal variability of $PM_{2.5}$ composition and sources in the Klang Valley urban-industrial environment. *Atmospheric Chemistry and Physics* **16**: 5357–5881.

Amir A. 2007. Air Pollution Trends In Petaling Jaya, Selangor, Malaysia. Master Thesis. Universiti Putra Malaysia.

Azid A., Juahir H., Latif M.T., Zain S.M. & Osman M.R. 2013. Feed-forward artificial neural network model for air pollutant index prediction in the southern region of Peninsular Malaysia. *Journal of Environmental Protection* **4**(12A): 1–10.

Azid A., Juahir H., Toriman M.E., Kamarudin M.K.A., Saudi A.S.M., Hasnam C.N.C., Aziz N.A.A., Azaman F., Latif M.T., Zainuddin S.F.M., Osman M.R. & Yamin M. 2014. Prediction of the level of air pollution using principal component analysis and artificial neural network techniques: A case study in Malaysia. *Water, Air, and Soil Pollution* **225**: 2063.

Brown R.L., Durbin J. & Evans J.M. 1975. Techniques for testing the constancy of the regression relationships over time. *Journal of the Royal Statistical Society: Series B (Methodological)* **37**(2): 149–163.

Chandio A.A., Jiang Y. & Rehman A. 2019. Using the ARDL-ECM approach to investigate the nexus between support price and wheat production: An empirical evidence from Pakistan. *Journal of Asian Business and Economic Studies* **26**(1): 139–152.

Chow J.C., Cao J., Chen L.-W.A., Wang X., Wang Q., Tian J., Ho S.S.H., Watts A.C., Carlson T.B., Kohl S.D. & Watson J.G. 2019. Changes in $PM_{2.5}$ peat combustion source profiles with atmospheric aging in an oxidation flow reactor. *Atmospheric Measurement Techniques* **12**: 5475–5501.

Dahari N., Muda K., Hussein N., Latif M.T., Khan M.F. & Mohamad Khir M.S. 2019. Long-range transport and local emission of atmospheric $PM_{2.5}$ in the southern region of Peninsular Malaysia. *IOP Conference Series: Materials Science and Engineering* **636**: 012005.

Darunikorn K., Jirapornkul C., Limmongkon Y., Junggoth R., Maneenin N. & Sakunkoo P. 2023. $PM_{2.5}$ levels in the Muang District of Khon Kaen Province by ambient $PM_{2.5}$ detectors with real-time sensors. *EnvironmentAsia* **16**(2): 109–117.

Engle R.F. & Granger W.J. 1987. Co-integration and error correction: representation, estimation, and testing. *Econometrica* **55**(2): 251–276.

Fujii Y., Tohno S., Amil N., Latif M.T., Oda M., Matsumoto J. & Mizohata A. 2015. Annual variations of carbonaceous $PM_{2.5}$ in Malaysia: Influence by Indonesian peatland fires. *Atmospheric Chemistry and Physics* **15**: 13319–13329.

Ganeshwaari R.G.N. & Koshy M.N. 2022. Residents' preferences on attributes of urban air quality improvement in Petaling Jaya, Selangor, Malaysia. *Journal of Sustainability Science and Management* **17**(2): 112–135.

Gessesse A.T. & He G. 2020. Analysis of carbon dioxide emissions, energy consumption, and economic growth in China. *Agricultural Economics* **66**(4): 183–192.

Hadipour M., Pourebahim S. & Mahmmud A.R. 2009. Mathematical modeling considering air pollution of transportation: An urban environmental planning case study in Petaling Jaya, Malaysia. *Theoretical and Empirical Researches in Urban Management* **4**(13): 75–92.

Imam A., Mohammed U. & Moses Abanyam C. 2014. On consistency and limitation of paired t-test, sign and Wilcoxon sign rank test. *IOSR Journal of Mathematics* **10**(1): 1–6.

IQAir. 2022. World Air Quality Report - Region & City PM2.5 Ranking.

Islam N., Toha T.R., Islam M.M. & Ahmed T. 2022. Spatio-temporal variation of meteorological influence on $PM_{2.5}$ and PM_{10} over major urban cities of Bangladesh. *Aerosol and Air Quality Research* **23**(1): 220082.

Johansen S. & Juselius K. 1990. Maximum likelihood estimation and inference on cointegration - with applications to the demand for money. *Oxford Bulletin of Economics and Statistics* **52**(2): 169–210.

Kwak S.G. & Kim J.H. 2017. Central limit theorem: The cornerstone of modern statistics. *Korean Journal of Anesthesiology* **70**(2): 144–156.

Madaki G.T. & Akanegbu B. 2023. An autoregressive distributed lagged (ARDL) bound testing approach to electricity supply and electricity tariff in Nigeria. *Journal of Global Economics and Business* **4**(13): 227–243.

Nguyen G.T.H., Nguyen T.T.T., Shimadera H., Uranishi K., Matsuo T. & Kondo A. 2022. Estimating mortality related to O_3 and $PM_{2.5}$ under changing climate and emission in continental Southeast Asia. *Aerosol and Air Quality Research* **22**: 220105.

Özcan B. & Öztürk I. 2019. *Environmental Kuznets Curve (EKC): A Manual*. Cambridge, Massachusetts: Academic Press.

Pal M. & Bharati P. 2019. Introduction to correlation and linear regression analysis. In *Introduction to Correlation and Linear Regression Analysis*: 1–18. Singapore: Springer.

Pesaran M.H., Shin Y. & Smith R.J. 2001. Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics* **16**(3): 289–326.

Rahman S.A., Hamzah M.S., Elias M.S., Salim N.A.A., Hashim A., Shukor S., Siong W.B. & Wood A.K. 2015. A long-term study on characterization and source apportionment of particulate pollution in Klang Valley, Kuala Lumpur. *Aerosol and Air Quality Research* **15**(6): 2291–2304.

Ravi V., Thakrar S., Hrath G., Wahyono A.D., Suryadi B., Avery G. & Hill J. 2022. Quantifying impacts of renewable electricity deployment on air quality and human health in Southeast Asia based on AIMs III scenarios. USAID-NREL Partnership Contract No. IAG-19-2115.

Rosli S., Ling O.H.L., Marzukhi M.A., Mohd Yusoff Z., Kwong Q.J. & Yinxue W. 2023. Urban environment and physical activity of Petaling Jaya residents. *International Journal of Sustainable Construction Engineering and Technology* **14**(5): 131–144.

Shafie S.H.M., Mahmud M., Mohamad S., Rameli N.L.F., Abdullah R. & Mohamed A.F. 2022. Influence of urban air pollution on the population in the Klang Valley, Malaysia: A spatial approach. *Ecological Processes* **11**(1): 3.

Shuaibu M., Adamu M.B., Abdullahi S.I., Shehu K.K. & Buba S.I. 2022. International trade and export dynamics in Nigeria: An ARDL vector error correction analysis. *Journal of Humanities, Arts and Social Science* **6**(1): 50–61.

Suradi H., Khan M.F., Sairi N.A., Rahim H.A., Yusoff S., Fujii Y., Qin K., Bari M.A., Othman M. & Latif M.T. 2021. Ambient levels, emission sources and health effects of $PM_{2.5}$-bound carbonaceous particles and polycyclic aromatic hydrocarbons in the city of Kuala Lumpur, Malaysia. *Atmosphere* **12**(5): 549.

Thangavel P., Park D. & Lee Y.-C. 2022. Recent insights into particulate matter ($PM_{2.5}$)-mediated toxicity in humans: An overview. *International Journal of Environmental Research and Public Health* **19**(12): 7511.

World Health Organization. 2021. *WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. Geneva: World Health Organization.

Zulkarnain N., Abdul Hadi M.S., Mohammad N.F. & Shogar I. 2023. Fitting time-varying coefficients SEIRD model to COVID-19 cases in Malaysia. *International Journal of Innovative Computing* **13**(1): 59–68.

Pusat Pemodelan dan Analisis Data (DELTA)
Department of Mathematical Sciences
Faculty of Science & Technology
Universiti Kebangsaan Malaysia
43600 UKM Bangi
Selangor DE, MALAYSIA
E-mail: zbhm@ukm.edu.my

Department of Mathematical Sciences
Faculty of Science & Technology
Universiti Kebangsaan Malaysia
43600 UKM Bangi
Selangor DE, MALAYSIA
*E-mail: p109135@siswa.ukm.edu.my**

Department of Statistics
Faculty of Science
University Of Benghazi
3332+MV5,
Banghazi, LIBYA
E-mail: ahmed.mami@uob.edu.ly

Department of Accounting
Faculty of Business and Economics
Universiti Malaya
50603 Kuala Lumpur
Wilayah Persekutuan Kuala Lumpur
E-mail: annarusli@um.edu.my

Received: 13 June 2024
Accepted: 25 July 2024

*Corresponding author