Science and Technology Indonesia
e-ISSN:2580-4391 *p*-ISSN:2580-4405
Vol. 11, No. 1, January 2026

Research Paper

Modeling Cointegrated Nonstationary Air Pollution Data: A Forecasting Study of NO_2 and SO_2 in Indonesia (1950–2022)

Arisman Adnan[1*], Gustriza Erda[1], Wamiliana[2], Edwin Russel[2]

[1]*Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Riau, Riau, 28293, Indonesia*
[2]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Lampung, 35145, Indonesia*
***Corresponding author**: arisman.adnan@lecturer.unri.ac.id*

Abstract
Air pollution from nitrogen dioxide (NO_2) and sulfur dioxide (SO_2) poses serious threats to human respiratory health and contributes to environmental degradation through acid rain formation. In Indonesia, despite rapid industrialization and increasing emissions, studies examining the interrelated dynamics between NO_2 and SO_2 at the national level remain limited, with most research focusing only on provincial areas and short time periods. This study fills this gap by analyzing the dynamic relationship between NO_2 and SO_2 using comprehensive national-level time series data from 1950 to 2022. The analysis examines short-term adjustments, long-term equilibrium patterns, directional causality, and shock responses between the two pollutants. The analysis focuses on identifying the best statistical model to capture the interaction between the two variables. Granger causality tests, impulse response functions (IRFs), and forecast error variance decomposition are applied to examine causal links and response dynamics. The data exhibits nonstationary but cointegrated with rank r=1, indicating a long-run equilibrium correlation between two pollutants. Consequently, the Vector Error-Correction Model, VECM(4), is selected as the most appropriate model. The study also provides 10-year forecasts for both pollutants insights into potential future air pollution trends in Indonesia, with NO_2 rising from 5.29 to 8.09 million tons and SO_2 from 3.38 to 5.10 million tons, underscoring the urgent need for integrated emission control policies that address both pollutants simultaneously rather than in isolation.

Keywords
Forecasting, Granger-Causality, Nitrogen Dioxide, Sulfur Dioxide, VAR(P), VECM

Received: 10 June 2025, Accepted: 7 November 2025
https://doi.org/10.26554/sti.2026.11.1.161-173

1. INTRODUCTION

The advancement of technology and rapid industrial development across various countries have intensified concerns regarding environmental degradation, particularly due to air pollution from industrial emissions. Nitrogen dioxide (NO_2) and sulfur dioxide (SO_2) are among the primary air pollutants of anthropogenic origin, coming from the combustion of fossil fuels in power plants and motor vehicles. It is well-documented that both gases have harmful effects on human health, particularly the respiratory system, and contribute to broader environmental issues (Jion et al., 2023; Nouri et al., 2023)). Many studies have highlighted that NO_2 and SO_2 emissions are on the rise, driven by increasing energy use and growing industrial activity. NO_2 is a reactive gas with a pungent odor, typically produced during high-temperature combustion processes, while SO_2 is a toxic gas released during the burning of sulfur-containing fuels, like petroleum and coal (White et al., 2020). The continuous release of these pollutants emphasizes the necessity of monitoring and modeling their behavior to inform effective environmental management strategies.

Many studies have been conducted showing the harmful effects of exposure to NO_2 and SO_2, especially when the two are mixed (Lee et al., 2008; Nouri et al., 2023; Yang et al., 2024). Exposure to high concentrations of the mixture of NO_2 and SO_2 can cause irritation and burning sensations in the throat, eyes, and nose. Once they enter the atmosphere, they react with oxygen, water vapor, and nitrogen to form acidic compounds such as sulfuric acid and nitric acid, which cause acid rain (He, 2024). Acid rain has a wide-ranging damaging impact on the environment, causing long-term negative impacts on water bodies, plants, and soil. These studies show that NO_2 and SO_2 in air pollution are also associated with respiratory diseases and related to global environmental problems such as acid rain. Acid rain has been demonstrated to exert long-term detrimental effects on ecosystems, including degradation of soil quality, acidification of water bodies, and damage to vegetation.

Furthermore, both NO_2 and SO_2 have been demonstrated to be significantly associated with respiratory illnesses and are widely recognized as key contributors to global environmental issues, such as atmospheric acidification and poor air quality (Yang et al., 2024).

Over the past 15 years, nitrogen dioxide emissions have gained increasing attention as a critical environmental concern, particularly due to their elevated concentrations near roadways resulting from fossil fuel combustion (Carslaw, 2005). Subsequent research across Europe has reinforced the substantial impact of traffic-related NO_2 emissions on ambient air pollution in proximity to highways (Casquero-Vera et al., 2019). The growing value of NO_2 is linked to the widespread implementation technologies for diesel emission control, such as Diesel Oxidation Catalysts (DOCs), which are designed to convert nitric oxide (NO) into NO_2. The process in question has been demonstrated to develop the oxidation of carbon monoxide (CO), hydrocarbons, and particulate matter.

The concentration levels of nitrogen dioxide (NO_2) and sulfur dioxide (SO_2) in Indonesia are quite concerning. According to The Air Quality Index (2025), the NO_2 concentration stands at 21 ppb, about 1.58 times above the World Health Organization's suggested limit of 13.29 ppb. In contrast, the SO_2 level is 13 ppb, which is 0.85 times below the WHO guideline of 15.27 ppb. Despite this, improving the accuracy of predictions for both pollutants remain essential to strengthen air quality monitoring and pollution control efforts in the country.

The Vector Error Correction Model (VECM) is an effective statistical approach for monitoring and forecasting air pollution trends. VECM is particularly suitable for analyzing pollutants with shared emission sources and atmospheric interactions, as it can simultaneously capture short-term fluctuations and long-term equilibrium relationships while accounting for mutual influences between variables. VECM is a good model for forecasting NO_2 and SO_2 emissions in Indonesia because it (1) deals with non-stationary but cointegrated data, (2) captures both short-run fluctuations and long-run equilibrium, (3) allows integration of environmental and health variables, and (4) provides meaningful forecasts to inform pollution control policies and health management strategies. Moreover, no study has yet employed the VECM model and the 1950-2022 dataset to forecast the impact of air pollution caused by nitrogen dioxide (NO_2) and sulfur dioxide (SO_2) emissions in Indonesia.

Research on application of VECM has been conducted by Alsaber et al. (2022), who used a time series approach to evaluate the connection among air pollution and COVID-19 cases in Kuwait. Their study found that air pollutants, temperature, and wind speed had a significant impact on the increase in COVID-19 hospitalizations, explaining around 24% of the rise in both the short and long term. Another study by Kurniawan et al. (2023) also employed VECM to investigate the influence of industrial activities-such as import and export values-along with wind speed and ozone (O_3), on tropospheric NO_2 concentrations in West Java. The finding exposed that in the short term, tropospheric NO_2 and O_3 bring effect to each other through photochemical reactions. In the future, industrial exports and wind speed have a major effect on NO_2 concentrations, which are also linked to the rising incidence of pneumonia caused by air pollution. These models have recognized popularity in capturing short-term dynamics and long-term equilibrium, enabling more accurate predictions of pollution levels and providing insights into future air quality trends.

In Indonesia, the number of studies on NO_2 and SO_2 emissions using VECM remains limited, with the majority concentrating on urban regions such as Jakarta and Makassar. For example, a study by Faridah et al. (2024) analyzed the distribution of air quality in Jakarta using NO_2 and SO_2 parameters. The results showed significant spatial variations, with high amount of SO_2 found in area of East Jakarta and NO_2 in area of West Jakarta. In addition, research conducted by Muhtar et al. (2021) also measured changes in SO_2 and NO_2 levels in Makassar City using Landsat 8 OLI satellite imagery before and during the COVID-19 cases, which presented a significant decrease in both pollutants. However, these studies only focused on provincial-level areas and limited time periods, so there is still a lack of understanding of the long-term dynamics of NO_2 and SO_2 pollution at the national level.

This study investigates the dynamic relationship between nitrogen dioxide (NO_2) and sulfur dioxide (SO_2) in Indonesia using vector time series data spanning start in 1950 to 2022. The analysis employs the Vector Error Correction Model (VECM) in capturing both the short-term dynamics and long-term equilibrium between the two variables. Specifically, the study examines: (1) short-term adjustments and interactions, (2) long-term equilibrium patterns through cointegration analysis, (3) directional causality using Granger-causality tests, and (4) shock responses between the pollutants through impulse response functions. Furthermore, this study utilizes the best-fitting VECM model to provide 10-year forecasts of NO_2 and SO_2 concentrations, offering insights to inform integrated pollution control policies.

2. EXPERIMENTAL SECTION

2.1 Statistical Methods

In the context of time series data analysis involving multiple variables, the appropriate methodological approach is the analysis of multivariate time series, where the data are represented in the form of a vector time series. A vector time series consisting of m variables can be expressed shown in Equation (1):

$$\mathbf{Z}_t = \begin{pmatrix} Z_{1t} \\ Z_{2t} \\ \vdots \\ Z_{mt} \end{pmatrix} \tag{1}$$

which is an $m \times 1$ vector of observations recorded at ordered and equally spaced time intervals, where t is the time index and $\mathbf{Z}_t$ is a random vector (Brockwell and Davis, 1991; Pena et al., 2001).

The random vector $\mathbf{Z}_t$ can be specified in terms of NO_2 and SO_2 emissions as in Equation (2):

$$\mathbf{Z}_t = \begin{pmatrix} NO_{2t} \\ SO_{2t} \end{pmatrix} \tag{2}$$

where NO_{2t} is nitrogen dioxide at time t, and SO_{2t} is sulfur dioxide at time t that are taken from Indonesia during 1950 to 2022 (Source: https://ourworldindata.org/cleanest-air-lessons).

2.2 Dynamic Model

The major goal of multivariate time series modeling and investigation is to understand how the active relationships between the variables and to enhance prediction and forecasting accuracy (Pena et al., 2001; Tsay, 2005, 2014; Wei, 2006, 2019). A fundamental supposition in multivariate or vector time series analysis is that the variables are correlated, a premise that necessitates the incorporation of autocorrelation structures within the model. Therefore, understanding the nature of the relationships among variables is essential to develop an appropriate model that yields reliable and accurate predictions (Lütkepohl, 2005; Tsay, 2014).

Another critical assumption in multivariate time series analysis is data stationarity. To verify the validity of this assumption, a visual inspection of the data plots is necessary to observe the behavior of the series and assess its stationarity. Furthermore, an examination of the autocorrelation function (ACF) plot can yield additional insights. A more rigorous and analytical approach involves the application of statistical tests, such as the Augmented Dickey-Fuller (ADF) test or the test of unit root, to get the stationarity of the data (Warsono et al., 2019a,b). The model is formulated as in Equation (3) using the ADF test with a lag length of p:

$$\Delta\mathbf{Z}_t = \alpha_0 + \gamma\mathbf{Z}_{t-1} + \sum_{i=1}^{p-1} \gamma_i^* \Delta\mathbf{Z}_{t-i} + \varepsilon_t \tag{3}$$

where ΔZ_t=Z_t-Z_{t-1} is the first difference of the vector time series at time t, α_0 is the constant (drift) term, γ is the coefficient on the lagged level variable, $\gamma_i{}^*$, are the coefficients on the lagged first differences for i=1,2,..., p-1, p is the maximum lag order, and σ_t is white noise. To test the null hypothesis that the data are nonstationary (i.e., $H_0 : \gamma = 0$). The statistic test is shown in Equation (4):

$$DF_\tau = \frac{\lambda}{Se(\lambda)} \tag{4}$$

The τ statistic, known as the ADF test, approximately follows a t-ratio (Brockwell and Davis, 1991). The null hypothesis (H_0) is rejected if the p-value is below $\alpha = 0.05$.

2.3 Cointegration

In the context of multivariate time series data modeling, it is essential to understand how the variables are connected, especially if they share a long term relationship, known as cointegration. This concept was first introduced by Engle and Granger (1987), and later expanded by Johansen (1988) who provided a method for estimating and inference. A vector time series Z_t is defined integrated of order one, denoted as $I(1)$, when its first difference (1-B)Z_t is stationary. In the event that variables implicated in the multivariate time series analysis exhibit a cointegration relationship, the model VAR foumd is not the best one (Tsay, 2014; Wei, 2019). The prevailing recommendation is to employ the Vector Error Correction Model (VECM) to evaluate the data. To determine the existence of a cointegration relationship between the variables in the vector time series, a null hypothesis test is conducted. It has been demonstrated that there is an r positive eigenvalue with following statistical test as in Equation (5):

$$Tr(r) = -T \sum_{i=r+1}^{k} \ln\left(1 - \hat{\delta}_i\right) \tag{5}$$

If the maximum eigen value test is used, the formula can be expressed as in Equation (6):

$$\delta_{\max}(r, r+1) = -T \ln\left(1 - \hat{\delta}_i\right) \tag{6}$$

2.4 The Model of Vector Autoregressive

The VAR (Vector Autoregressive) model is widely used in multivariate time series analysis due to its relatively simple underlying assumptions. Moreover, the interpretation of the results is uncomplicated. The VAR model treats all variables symmetrically, meaning each variable is modeled as a linear function of past values of itself and other variables in the system. In the context of VAR modeling, the dependent variable on the left-hand side is represented by a vector comprising multiple variables. Conversely, the right side encompasses lagged values of these variables, thereby reflecting the autoregressive nature of the model. The model VAR of order p is a mathematical representation of a system of equations used to evaluate the dynamic effect among variables over time. The quantity p is represented as shown in Equation (7):

$$Z_t = \sum_{i=1}^{p} \Theta_i Z_{t-i} + \varepsilon_t, \tag{7}$$

where Z_t is a $k \times 1$ vector of observations at time t, Θ_i is a $k \times k$ coefficient matrix associated with the lagged vector Z_{t-i} for $i = 1, 2, \ldots, p$, where p denotes the lag order, and ε_t is the $k \times 1$ vector of shocks at time t.

2.5 The Vector Error Correction Model (VECM)

The presence of a cointegration relationship in the middle of variables requires a modification of the VAR model into a

Vector Error Correction Model (Hamilton, 1994; Lütkepohl, 2005; Pena et al., 2001; Tsay, 2014). The VECM is a limited VAR model that was developed for the purpose of analyzing nonstationary time series data (Hamilton, 1994; Wei, 2019) that exhibits cointegration. The VECM model is basically a tool that can help us to estimate the long and short-term effects among variables. The VECM(p) model can be expressed as shown in Equation (8):

$$\Delta Z_t = \Pi Z_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta Z_{t-i} + \varepsilon_t, \tag{8}$$

where Δ is the differencing operator, $\Delta Z_t = Z_t - Z_{t-1}$, and Z_{t-1} is the endogenous variable at lag 1. Π is the cointegration matrix, ε_t is the $k \times 1$ white-noise vector, and $\Pi = \alpha\beta^{\mathsf{T}}$, where α is the adjustment matrix ($k \times r$) and β is the cointegration matrix ($k \times r$). Γ_i is the coefficient matrix ($k \times k$) for the i-th differenced endogenous variable.

2.6 Normality Test of Residuals

Prior to the model application for additional analysis, it is imperative to verify several assumptions. A fundamental assumption is the normality of the residuals, which can be assessed visually by examining the QQ plot or the histogram of the residuals. Furthermore, the Jarque-Bera (JB) test is frequently utilized to statistically assess the normality of data (Jarque and Bera, 1987). The JB test statistic is defined as shown in Equation (9):

$$JB = \left[\frac{n}{6} b_1^2 + \frac{n}{24}(b_2 - 3)^2\right], \tag{9}$$

where n is the number of samples, b_1 is the sample skewness, and b_2 is the sample kurtosis. The JB test statistic follows a χ^2 distribution with two degrees of freedom.

2.7 Test of Stability

To verify that constructed model satisfies the requisite criteria for stability, it will be evaluated using the property polynomial of the VAR(p) process. According to Hamilton (1994) and Lütkepohl (2005), the model is said stable if all roots of the characteristic polynomial have a modulus value less than 1, meaning they lie strictly inside the unit circle in the complex plane. The VAR(p) model is given in Equation (10):

$$Z_t = C + \Phi_1 Z_{t-1} + \cdots + \Phi_p Z_{t-p} + \varepsilon_t \tag{10}$$

and the polynomial associated with the coefficient matrices of VAR(p) process is called the polynomial properties. The processes of VAR(p) is defined to be stable when the determinant condition in Equation (11) is satisfied, that is:

$$\det(IKp - \Phi z) = \det(I_K - \Phi_1 z - \cdots - \Phi_{pzp}) \tag{11}$$

meaning that each root has a modulus less than one.

2.8 Granger Causality, Impulse Response Function (IRF), Forecasting and Proportion Prediction Error Covariance

After selecting the optimal model and verifying its assumptions, further analysis will be conducted. First, the Granger-causality theory will be employed to examine the direction of causality between NO_2 and SO_2, and vice versa for SO_2 and NO_2. The approach employed in the analysis aligns with the recommendations of Hamilton (1994) and (Tsay, 2014). Second, the Impulse Response Function (IRF) is a concept that explore the ramifications of a sudden shock, measuring one standard deviation, affecting a single variable. This theoretical framework is further expanded to explore the cascading effects of such an event on other variables (Tsay, 2014; Wei, 2019). Third, the forecasting process will be executed using the best-fitting VECM model to generate. Finally, the forecast error variance decomposition will be employed to describe the impact of other variables on a given variable in predicting future values. Additionally, the impact of other variables on the long-term forecast of that variable will also be calculated (Florens et al., 2007; Tsay, 2014).

This study begins by analyzing the trend of NO_2 and SO_2 concentration data in Indonesia from 1950 to 2022, using the Augmented Dickey-Fuller (ADF) stationarity test to verify the non-stationary nature of the data. After differencing the data, both data series become stationary. Furthermore, to analyze the dynamic relationship between the two pollutants, the Vector Error Correction Model (VECM) model is applied, with the Johansen cointegration test indicating a long-term connection between NO_2 and SO_2. Granger-causality analysis is performed to test the direction of influence between the two variables, which results in the finding of a two-way causal relationship. Forecasting for the next decade is carried out using the identified VECM model.

3. RESULTS AND DISCUSSION

3.1 Data Exploration

Figure 1 displays the trends of NO_2 and SO_2 concentrations. It is clearly seen that both pollutants have demonstrated a precipitous increase commencing in the 1990, accompanied by substantial fluctuations in comparison to the preceding decades. This pattern suggests that the time series data for NO_2 and SO_2 are likely non-stationary.

Furthermore, the Augmented Dicky Fuller test (ADF test) is performed using SAS, it was found that the null hypothesis (H_0) of the data non-stationary. Table 1 shows the Dickey-Fuller unit root test. The test shows that H_0 is not rejected, with p-values for NO_2 and SO_2 of 0.9999 each (before differencing). Therefore, it concluded that the NO_2 and SO_2 data are not stationary. After first differencing, the data for both NO_2 and SO_2 became stationary and can be used for further analysis.

Figures 2(a) and 2(b) present the time series plots and autocorrelation function (ACF) plots for NO_2 and SO_2. As illustrated in Figure 2(a), the NO_2 time series demonstrates an upward trend, and its ACF exhibits a slow decay, suggesting non-stationarity. In a similar vein, Figure 2(b) for the SO_2

Table 1. Dickey-Fuller Unit Root Test

Variable	Type	Before Differencing			After Differencing		
		Rho	p-value	Tau	Rho	p-value	Tau
NO_2	Zero Mean	3.02	0.9984	3.43	-69.74	< 0.0001	-5.64
	Single Mean	2.37	0.9986	2.12	-98.41	0.0007	-6.73
	Trend	-1.16	0.9848	-0.51	-138.57	0.0001	-8.03
SO_2	Zero Mean	3.56	0.9994	6.41	-30.25	< 0.0001	-2.88
	Single Mean	3.30	0.9997	4.91	-67.42	0.0007	-4.45
	Trend	1.90	0.9996	1.31	-224.03	0.0001	-7.98

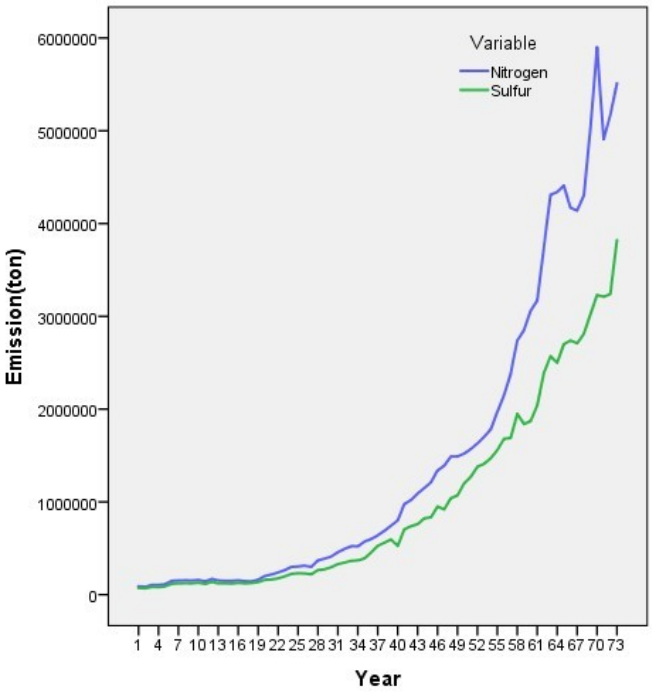


Figure 1. Plot data Nitrogen Dioxide and Sulfur Dioxide From the Year 1950 to 2022

time series demonstrates an upward trend with a gradually decaying ACF, thereby also indicating non-stationarity. After differencing once (d = 1), then the Augmented Dickey-Fuller (ADF) unit root test was performed. The study suggest rejected of the null hypothesis, as evidenced by the p-values for SO_2 falling below the 0.05 significance level. The findings indicate that the differenced data meet the stationarity assumption, as outlined in Table 1.

3.2 Test for Lag Optimum

The objective of this study is to reveal the optimal lag for the VAR(p) model applied to NO_2 and SO_2 time series data. To this end, the AICC (Corrected Akaike Information Criterion), HQC (Hannan-Quinn Criterion), AIC (Akaike Information Criterion), and SBC (Schwarz Bayesian Criterion) criteria will be employed. These information criteria balance model fit against complexity, with lower values indicating better model performance.

Table 2 shows the results of the criterion analysis based on AICC, HQC, AIC, and SBC for the VAR(2), VAR(3), and VAR(4) models. All four information criteria consistently decrease as the lag order increases from 2 to 4, with the minimum values achieved at lag 4. Specifically, the AIC decreases from 46.85521 (VAR(2)) to 46.36881 (VAR(4)), while the SBC decreases from 47.17389 to 46.95162. Therefore, based on the unanimous agreement of all information criteria, the VAR(4) model with four lags is selected as the optimal specification for subsequent analysis.

Table 2. The Values of Information Criteria

Information Criteria	VAR(2)	VAR(3)	VAR(4)
AICC	46.87655	46.65163	46.44707
HQC	46.98194	46.78581	46.60003
AIC	46.85521	46.60719	46.36881
SBC	47.17389	47.05688	46.95162

3.3 Test for Checking the Cointegration

To ascertain the presence of cointegration in the relationship between NO_2 and SO_2 time series data, it is necessary to verify that the NO_2 and SO_2, time series data are nonstationary (see Table 1). Nonstationary data can be I(1), or integrated with order one. Subsequent to the differentiation of the time series data on a single occasion (d = 1), the NO_2 and SO_2, time series data become stationary or integrated with order 0, I(0). To ascertain the presence of cointegration, the Johansen test is employed on the optimum lag, specifically on VAR(4). The calculation of the cointegration test is displayed in Table 3.

Table 3, shows cointegration test in the first stage tested Ho: Rank = 0 (r = 0) with alternative Ha: Rank > 0. The test results in the first stage obtained p-value = 0.0002. Because Ho is rejected, it is concluded that Rank > 0. In the second stage tested Ho: Rank = 1 (r = 1) with alternative Ha: Rank > 1. The test results in the first stage obtained p -value = 0.1188 > 0.05. Because Ho is not rejected, hence, the conclusion is that the rank cointegration is r = 1. Because the NO_2 and SO_2 time series data have a cointegration relationship, the model to be used is the VECM (4) model with rank cointegration r = 1.

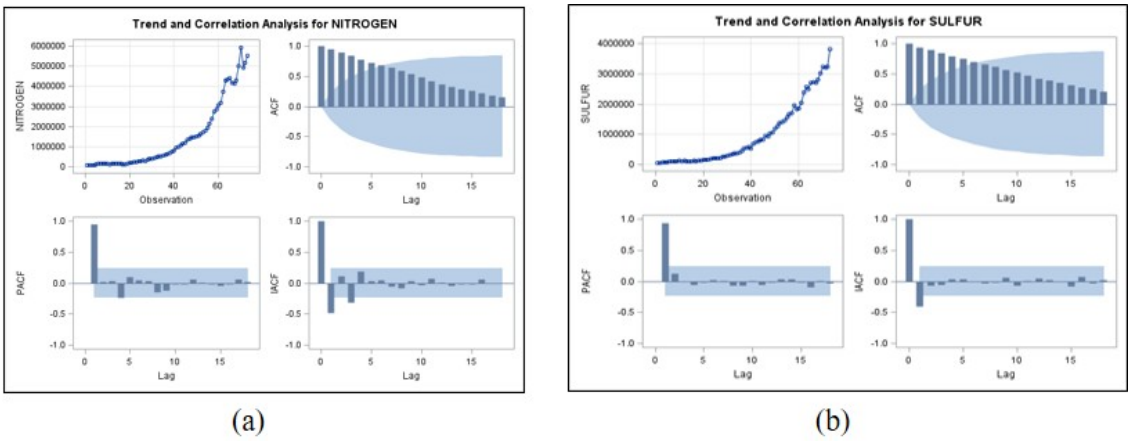


Figure 2. (a) Plot Data and ACF for NO_2 (b) Plot Data and ACF for SO_2

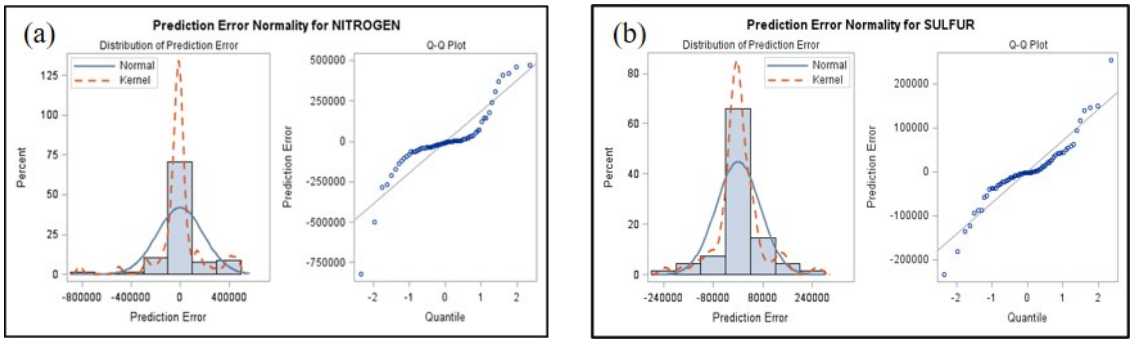


Figure 3. (a) Error Normality Prediction for NO_2, and (b) Error Normality Prediction for SO_2

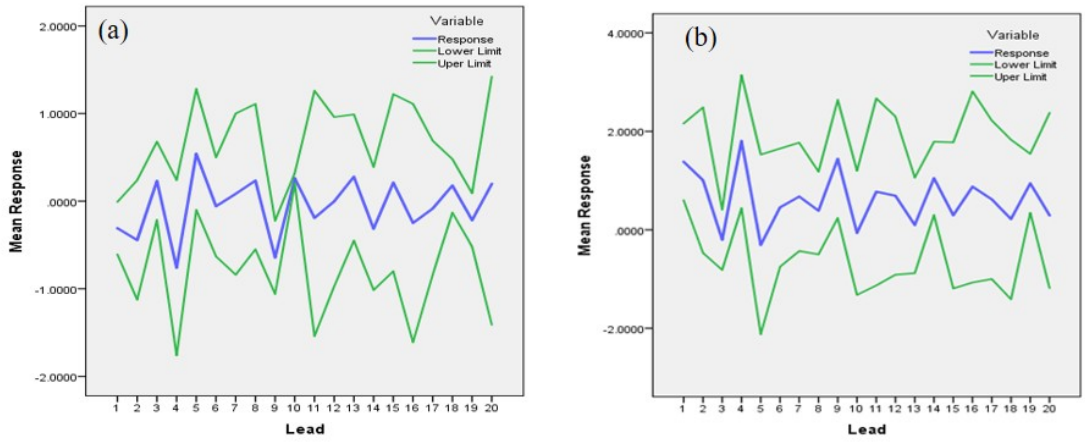


Figure 4. Impulse Response Function for Shock in Variable NO_2 and (a) the Response of Nitrogen Dioxide (NO_2), and (b) the Response of Sulfur Dioxide (SO_2)

Table 3. Cointegration Rank Test Using Trace

H_0: Rank = r	H_1: Rank > r	Eigenvalue	Trace	p-value	Drift in ECM	Drift in Process
0	1	0.3317	30.2355	0.0002	Constant	Linear
1	2	0.0346	2.4303	0.1188	Constant	Linear

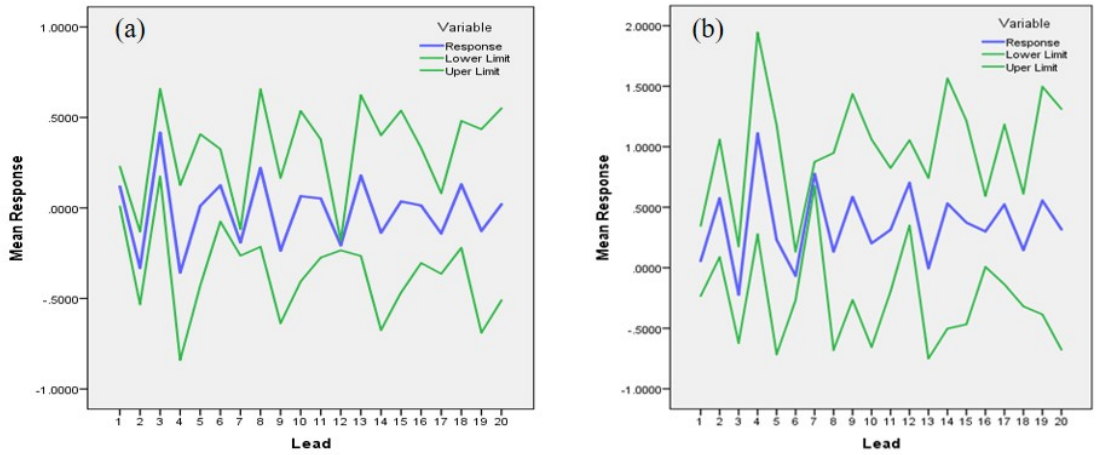


Figure 5. IRF for Shock in Variable SO_2 and (a) the Response of Nitrogen Dioxide (NO_2), and (b) the Response of Sulfur Dioxide (SO_2).

Table 4. Parameter Estimates for Model VECM(4) with Cointegration Rank r=1.

Equation	Parameter	Estimate	Standard Error	*t* Value	*p*-value	Variable
D_NITROGEN	CONST1	5765.14753	23127.37226	0.25	0.8040	1
	AR1_1_1	-2.93784	0.49064	-5.99	0.0000	NITROGEN(t-1)
	AR1_1_2	4.67546	0.78083	5.99	0.0000	SULFUR(t-1)
	AR2_1_1	1.62856	0.41170	3.96	0.0002	D_NITROGEN(t-1)
	AR2_1_2	-3.29535	0.66168	-4.98	0.0000	D_SULFUR(t-1)
	AR3_1_1	0.92534	0.30613	3.02	0.0030	D_NITROGEN(t-2)
	AR3_1_2	-1.93982	0.54031	-3.59	0.0000	D_SULFUR(t-2)
	AR4_1_1	1.09782	0.33753	3.25	0.0010	D_NITROGEN(t-3)
	AR4_1_2	-1.72814	0.41872	-4.13	0.0000	D_SULFUR(t-3)
D_SULFUR	CONST2	5400.79824	8623.77721	0.63	0.5336	1
	AR1_2_1	0.39560	0.18295	2.16	0.0320	NITROGEN(t-1)
	AR1_2_2	2.02090	0.34321	5.89	0.0000	SULFUR(t-1)
	AR2_2_1	0.25757	0.15352	1.68	0.0987	D_NITROGEN(t-1)
	AR2_2_2	-1.16481	0.24839	-4.69	0.0000	D_SULFUR(t-1)
	AR3_2_1	-0.59943	0.15396	-3.89	0.0001	D_NITROGEN(t-2)
	AR3_2_2	-0.75677	0.20147	-3.76	0.0002	D_SULFUR(t-2)
	AR4_2_1	0.38102	0.12586	2.40	0.0197	D_NITROGEN(t-3)
	AR4_2_2	-0.73842	0.15613	-4.73	0.0000	D_SULFUR(t-3)

As illustrated in Table 3, the initial stage of the cointegration test was utilized to assess the hypothesis (Ho: r = 0) and its alternative (Ha: r > 0). The outcomes of this test yielded a *p*-value of 0.0002. Therefore, Ho is rejected, which indicates that Rank > 0. In the second experiment, we tested Ho: r = 1 with the alternative Ha: r > 1. The *p*-value obtained from the initial stage of the experiment was 0.1188, which exceeds the 0.05 threshold. Therefore, Ho is not rejected, and it can be said that the rank cointegration is r = 1. Given the established cointegration relationship between the NO_2 and SO_2 time series data, the VECM (4) model with rank cointegration r = 1 is chosen.

3.4 The Estimation of Parameter VECM(4) with Cointegration Rank r=1

From the results of the analysis above, the model that will be used to describe the relationship pattern between NO_2 and SO_2 time series data is the model VECM(4) with cointegration rank $r = 1$. The NO_2 and SO_2 time series vector data is represented as shown in Equation (12):

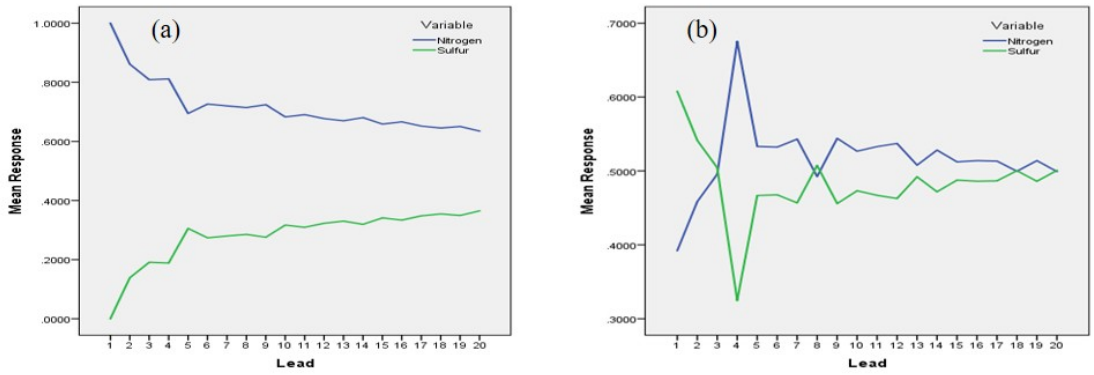
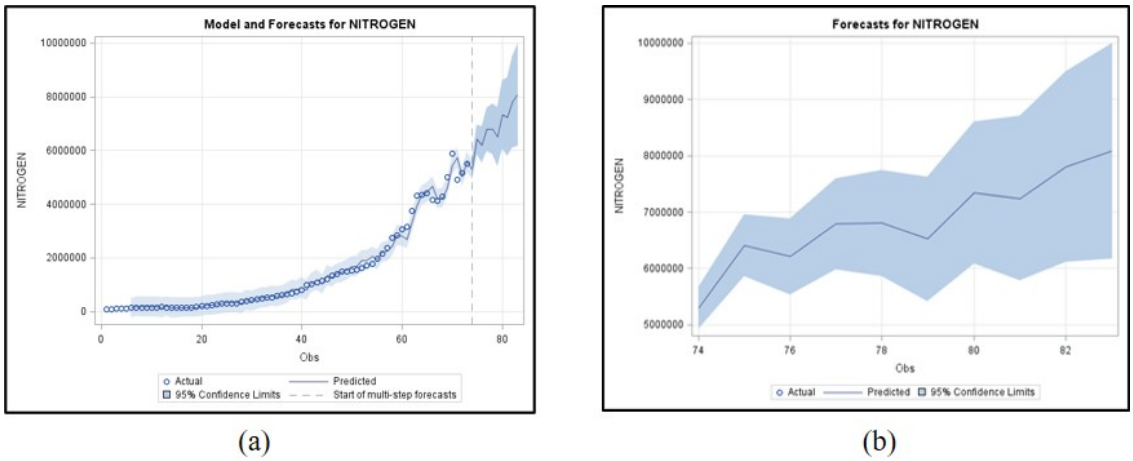


Figure 6. Proportion of Predictions Error Covariance for (a) Nitrogen Dioxide, and (b) Sulfur Dioxide.



(a) (b)

Figure 7. (a) Model and Forecasting for Nitrogen (b) Forecast of Nitrogen

$$Z_t = \begin{bmatrix} \mathrm{NO}_{2t} \\ \mathrm{SO}_{2t} \end{bmatrix} \tag{12}$$

From the calculation of estimating the parameters of the VECM(4) model with cointegration rank $r = 1$, the estimated model is obtained as shown in Equation (13):

$$\Delta Z_t = C + \Pi Z_{t-1} + \Phi_1 \Delta Z_{t-1} + \Phi_2 \Delta Z_{t-2} + \Phi_3 \Delta Z_{t-3} + \varepsilon_t \tag{13}$$

The parameter coefficients used for constructing the VECM model are based on Table 4. Table 4 presents the estimated coefficients, standard errors, t-values, and p-values for both equations in the VECM system.

Table 4 shows that the constants (CONST1 and CONST2) for both equations are not significant ($p > 0.05$), indicating no deterministic trend in pollutant changes. The error correction coefficients (AR1_1_1 = -2.93784 for NO_2; AR1_2_1 = -0.13955 for SO_2) reveal that NO_2 adjusts much faster to restore long-run equilibrium compared to SO_2. Most short-run coefficients (AR2, AR3, AR4) are statistically significant, with lagged SO_2 changes showing strong negative effects on both pollutants, confirming the dynamic interdependence between NO_2 and SO_2 and validating the VECM(4) specification.

The estimation of the model is presented in Equation (14):

$$\begin{bmatrix} \Delta NO2_t \\ \Delta SO2_t \end{bmatrix} = \begin{bmatrix} 5765.1475 \\ 5400.7982 \end{bmatrix} + \begin{bmatrix} -2.9378 & 4.6754 \\ -0.1395 & 0.2221 \end{bmatrix} \begin{bmatrix} NO2_{t-1} \\ SO2_{t-1} \end{bmatrix} + \begin{bmatrix} 1.6285 & -3.2953 \\ 0.2575 & -1.1648 \end{bmatrix} \begin{bmatrix} \Delta NO2_{t-1} \\ \Delta SO2_{t-1} \end{bmatrix} + \begin{bmatrix} 0.9253 & -1.9398 \\ -0.0440 & -0.7567 \end{bmatrix} \begin{bmatrix} \Delta NO2_{t-2} \\ \Delta SO2_{t-2} \end{bmatrix} + \begin{bmatrix} 1.0978 & -1.7281 \\ 0.3018 & -0.7384 \end{bmatrix} \begin{bmatrix} \Delta NO2_{t-3} \\ \Delta SO2_{t-3} \end{bmatrix} \tag{14}$$

and the covariance of innovation is given in Equation (15).

$$\mathrm{Cov}(\varepsilon_t) = \begin{bmatrix} 3570204108.5 & 8337639430 \\ 8337639430 & 4964037656 \end{bmatrix} \tag{15}$$

It can be seen from Table 4 shows that the constants for both variables are not significant. This means that there is no

strong deterministic trend in the changes in the variables. The lag coefficients 1 to lag 4 of the changes in NO_2 and SO_2 are mostly significant (p-value <0.05) indicating that the changes in NO_2 and SO_2 in the previous period have an effect on the current changes in NO_2. In addition, several lag coefficients of SO_2 changes are significantly negative indicating a short-term negative adjustment effect.

Table 5. Long-Run Beta Estimates When r = 1

Variable	1
NITROGEN	1.00000
SULFUR	-1.59146

Table 5 shows that NO_2 and SO_2 have a long-term equilibrium relationship where NO_2 is proportional to 1.59 times SO_2.

Table 6. Adjustment Alpha Estimates When r = 1

Variable	1
NITROGEN	-2.93784
SULFUR	-0.13955

Table 6 shows that the Alpha for NO_2 is -2.93784, indicating that if NO_2 is above its long-term equilibrium value, it will adjust downward in the next period to return to equilibrium. Meanwhile, the alpha value for SO_2 is also negative but smaller (-0.1395). This suggests that SO_2 adjusts more slowly or is less responsive to deviations from equilibrium.

Table 7. The Product of Adjustment and Cointegration Matrices

Parameter Alpha*Beta Estimates	NO_2	SO_2
NO_2	-2.9378	4.6754
SO_2	-0.1395	0.2221

Table 8. White Noise Diagnostics for Univariate Model

Variable	Durbin Watson	Normality	
		Chi-Square	p-value
NITROGEN	1.7029	85.18	<0.0001
SULFUR	1.9607	32.01	<0.0001

Table 7 describes the long-term relationship between NO_2 and SO_2, and it can be used to understand the influence between these two pollutants in the VECM model. The coefficient of NO_2 on SO_2 is 4.6754, meaning that a 1-unit change in the concentration of NO_2 in the previous period will lead to an increase of 4.6754 units in the concentration of SO_2 in the following period. The coefficient of SO_2 on NO_2 is -0.1395, indicating that a 1-unit increase in SO_2 concentration in the earlier period will reduce the concentration of NO_2 by 0.1395 units in the next period. This recommend that although there is positive long-term relationship between NO_2 and SO_2, there is also a mechanism influencing the direction of this relationship, with SO_2 having a dampening effect on NO_2 concentrations.

Normality of residual is used to check the error distribution. The Durbin-Watson test is employed to test for the presence of autocorrelation in the residual model. The value for both variables is close to 2 signifying that there is no significant autocorrelation in the residual so that the model used is good enough in capturing the autocorrelation pattern in the data.

To assess the normality of residuals, the Jarque-Bera (JB) test and visual inspection through Q-Q plots are employed (see Table 8 and Figure 3). Table 8 and Figure 3 show that residuals are not normally distributed based on the Jarque-Bera test ($p <$ 0.0001 for both variables), with Q-Q plots revealing deviations in the tails. However, with 73 observations, the Central Limit Theorem ensures that parameter estimates remain asymptotically valid despite non-normality. Additionally, VECM models are relatively robust to distributional deviations, particularly for cointegration analysis and forecasting. The Durbin-Watson statistics near 2.0 confirm no significant autocorrelation, indicating the model adequately captures temporal dynamics. Therefore, the model is considered acceptable for the study's objectives.

3.5 Stability Model

The stability of the model can be shown by using the roots of the Autoregressive (AR) characteristic polynomial in the VECM model.

Table 9 shows that the roots whose modulus is smaller than or equal to 1, then the model is unstable. Although there is a root with modulus exactly 1 (Root 1), most of the other roots have moduli smaller than 1, revealing that the model is stable overall.

3.6 Test for the Fitness of the Univariate Model

From the model of VECM(4) with cointegration Rank r=1 from Equation (14), the univariate model can be presented as follows:

$$\begin{aligned}\Delta NO2_t = {} & 5765.1475 - 2.9378\,NO2_{t-1} + 4.6754\,SO2_{t-1} + 1.6285\,\Delta NO2_{t-1} \\ & - 3.2953\,\Delta SO2_{t-1} + 0.9253\,\Delta NO2_{t-2} - 1.9398\,\Delta SO2_{t-2} \\ & + 1.0978\,\Delta NO2_{t-3} - 1.7281\,\Delta SO2_{t-3}\end{aligned} \tag{16}$$

$$\begin{aligned}\Delta SO2_t = {} & 5400.7982 - 0.1395\,NO2_{t-1} + 0.2221\,SO2_{t-1} + 0.2575\,\Delta NO2_{t-1} \\ & - 1.1648\,\Delta SO2_{t-1} - 0.0440\,\Delta NO2_{t-2} - 0.7567\,\Delta SO2_{t-2} \\ & + 0.3018\,\Delta NO2_{t-3} - 0.7384\,\Delta SO2_{t-3}\end{aligned} \tag{17}$$

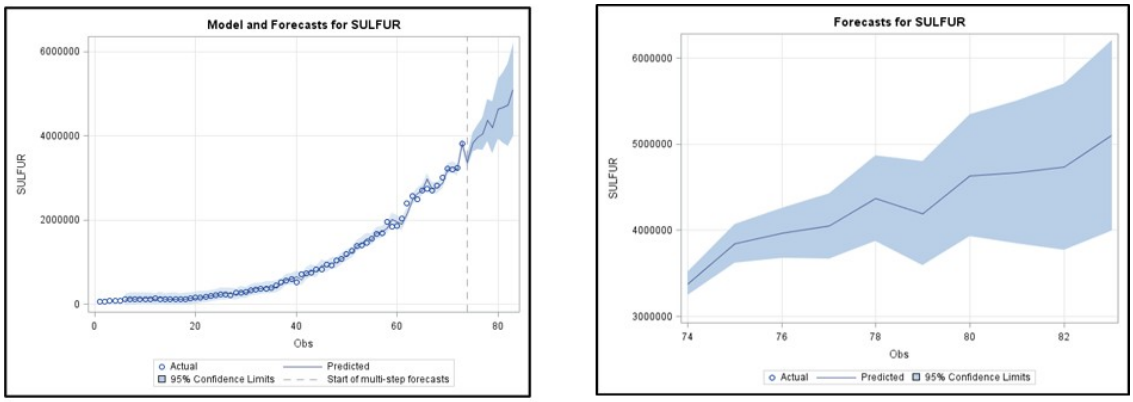
Table 10 displays the significance test of models (16) and (17). Both models are statistically significant ($p <$ 0.0001), with R-squared values of 0.6136 for NO_2 and 0.6817 for SO_2. These moderate R-squared values indicate that approximately 39% of NO_2 variability and 32% of SO_2 variability remain unexplained, likely due to unobserved factors such as meteorological conditions, policy changes, or measurement errors. Therefore, the 10-year forecasts from models (16) and (17) should be

Table 9. AR Characteristic Polynomial Roots

Index	Real	Imaginary	Modulus	Radian	Degree
1	1.00000	0.00000	1.0000	0.0000	0.0000
2	0.50194	0.722848	0.8847	0.9675	55.4322
3	0.50194	-0.722848	0.8847	-0.9675	-55.4322
4	-0.08785	0.86274	0.8672	1.6723	95.8139
5	-0.08785	-0.86274	0.8672	-1.6723	-95.8139
6	-0.55553	0.00000	0.5555	3.1416	180.0000
7	-0.76234	0.55041	0.9403	2.5162	144.1704
8	-0.76234	-0.55041	0.9403	-2.5162	-144.1704

Table 10. ANOVA Diagnostics of Univariate Model

Variable	R-Square	Standard Deviation	F Value	p-value
NITROGEN	0.6136	188949.83748	11.71	< 0.0001
SULFUR	0.6817	70455.92705	15.80	< 0.0001



(a) (b)

Figure 8. (a) Model and Forecasting for Sulfur (b) Forecast of Sulfur

Table 11. Test of Granger-Causality Wald

Test	Group Variable	Null hypothesis	Chi-Square	p-value
1	Group 1. Variable: NO_2 Group 2. Variable: SO_2	Ho: NO_2 is only influenced by itself and not affected by SO_2.	20.59	0.0004
2	Group 1. Variable: SO_2 Group 2. Variable: NO_2	Ho: SO_2 is only influenced by itself and not affected by NO_2.		

interpreted as conditional trend projections based on historical patterns rather than precise predictions, with substantial uncertainty reflected in the forecast confidence intervals.

3.7 Testing for Granger-Causality

To test the causal relationship among the NO_2 and SO_2 variables in the VECM model, a Granger-causality test was carried out using the Wald test statistic, as presented in Table 11.

Table 11 presents the results of Granger-causality tests examining whether past values of one pollutant help predict the current values of the other. In Test 1, the null hypothesis that NO_2 is influenced only by its own past values (and not by SO_2) is strongly rejected ($p = 0.0004$), indicating that SO_2 Granger-causes NO_2. Similarly, Test 2 rejects the hypothesis that SO_2 is independent of NO_2 ($p < 0.0001$), with an even stronger chi-square statistic (52.06), confirming that NO_2 Granger-causes SO_2. The substantially larger chi-square value in Test 2 suggests that NO_2 has a stronger predictive influence on SO_2 than the reverse. Together, these results establish bidirectional Granger-causality: changes in SO_2 concentration can be used to forecast changes in NO_2, and vice versa, though the influence of NO_2 on SO_2 appears more pronounced.

Table 12. Forecasting Nitrogen and Sulfur Dioxide for 10 Years

Variable	Obs	Forecast	Standard Error	95% Confidence Limits	
NITROGEN	74	5290740.9453	189849.83748	4920406.0689	5661075.8216
	75	6406705.2118	279438.77186	5859015.2707	6954395.1529
	76	7210211.6945	344016.97525	6535608.8130	7888032.5761
	77	6785978.9563	409352.56111	5938662.6796	7588295.2331
	78	6801374.9281	480026.98764	5806136.5272	7742563.3290
	79	6521277.3605	562266.17537	5418344.8479	7624179.8730
	80	7341860.8932	645254.57921	6071875.1571	8605636.6293
	81	7243136.0426	746343.89535	5780328.8877	8705943.1976
	82	7806811.1663	848501.94393	6145304.0741	9460954.0741
	83	8085368.2920	975562.84331	6173203.2645	9997436.3296
SULFUR	74	359470.4943	112275.47691	137597.1249	581343.8638
	75	392407.8121	137958.85359	406640.5642	745389.0596
	76	396401.6419	148745.26547	367502.2682	4255576.9957
	77	436912.3738	253140.96577	3872365.2025	4865559.5544
	78	419518.2343	301489.36785	358330.2197	4801706.2489
	79	463386.5039	361547.43104	3342860.3842	5344260.3842
	80	467340.1486	423829.95560	3842710.7700	5504093.5971
	81	510362.9408	563336.69098	3998430.3959	6207182.5663
	82	503860.9213	612742.84331	4173203.2645	6997436.3296
	83	510362.9408	563336.69098	3998430.3959	6207182.5663

3.8 Impulse Response Function

Impulse Response Function (IRF) for VECM model between NO_2 and SO_2 is used to see how one variable responds to a "shock" on another variable in the next few periods. This analysis is also used to understand the short and medium term impacts of sudden changes in one pollutant on itself and other pollutants.

Figure 4 and Figure 5 show that shocks to NO_2 and SO_2 both have a direct impact on the variable itself, but their effects will disappear over time.

3.9 Forecasting and Proportion Prediction Error Covariance

The forecasting of nitrogen dioxide (NO_2) and sulfur dioxide (SO_2) emissions for the subsequent decade was conducted using Models (11) and Model (15). The two models are highly significant if p-values < 0.0001 (see Table 10) and R-squared of 0.6136 and 0.6817, respectively. These findings suggest that Model (14) accounts for 61.36% of the variability in NO_2 emissions, while Model (15) explains 68.17% of the variability in SO_2 emissions. The high significance levels indicate that both models are reliable for short- to medium-term forecasting, as evidenced by the results presented in Tables 4, 10, and 11. Moreover, the Granger-causality test indicate a bidirectional causal connection between NO_2 and SO_2 where SO_2 is the Granger-causality for NO_2, and vice versa.

The percentages reported below are derived from the Forecast Error Variance Decomposition (FEVD), which is calculated from the estimated VECM(4) parameters in Table 4. The FEVD quantifies what proportion of the h-step-ahead forecast error variance in one variable is attributable to innovations (shocks) in another variable, computed using the moving average representation of the VECM model.

Figure 6 presents the results of the proportion predictions error covariance for both NO_2 and SO_2. As illustrated in Figure 6(a), SO_2 exhibits no substantial short-term influence on the one-year-ahead forecast of NO_2. However, its influence increases over time, contributing approximately 35% at h=20 years to the forecast of NO_2 in the long term. In contrast, Figure 6(b) demonstrates that NO_2 exerts a substantial short-term impact on SO_2 forecasts, contributing approximately 39% at h=1 year to the one-year-ahead forecast. This influence increases further over time, reaching approximately 51% to 55% in the long term at h=10-20 years (SAS/ETS, 2014). These findings suggest a stronger predictive influence of NO_2 on SO_2 compared to the reverse.

The goodness-of-fit of the models is visually confirmed in Figures 7 and 10. As illustrated in Figure 7, Model (14) shows that the model used fits the data very well where the expected values and observed values are very close together, while Figure 8 shows a comparable alignment for SO_2 utilizing Model (15). These results further exhibits the predictive capability of both models.

The forecasting results for the subsequent decade indicate an overall upward trend with noticeable fluctuations in both nitrogen dioxide (NO_2) and sulfur dioxide (SO_2) (see Figures 7 and 8). As illustrated in Table 12, the projected NO_2 emissions for the initial year amount to 5,290,740.95 tons, with an escalation to 8,085,368.29 tons by the tenth year. Table 15 presents the ten-year forecasts of NO_2 and SO_2 emissions expressed in tons, including the predicted annual values, stan-

dard errors, and 95% confidence intervals, which indicate the statistical uncertainty of each estimate.

4. CONCLUSIONS

This study examined the relationship between nitrogen dioxide (NO_2) and sulfur dioxide (SO_2) emissions in Indonesia from 1950 to 2022 using vector time series analysis. Both pollutants were integrated of order one (I(1)), and cointegration tests confirmed a long-term equilibrium, supporting the use of a Vector Error Correction Model (VECM). Granger causality revealed a bidirectional relationship, while impulse responses indicated persistent effects lasting about 20 years. Forecasts suggest rising emissions driven by fossil fuel use and industrial growth. NO_2 explained more than half of SO_2 variation, highlighting its dominant influence. The findings emphasize the need for integrated emission control policies targeting NO_2 to indirectly reduce SO_2 levels. Future research should consider additional pollutants and nonlinear or machine learning approaches to capture complex environmental dynamics.

5. ACKNOWLEDGMENT

The author gratefully acknowledges the support provided by Universitas Riau in this study. Appreciation is also extended to the researchers of the Environmental Statistics Laboratory at the Department of Statistics, Universitas Riau and Universitas Lampung, for their valuable collaboration in this joint research project. Their insightful discussions significantly contributed to the quality of this study.

REFERENCES

Alsaber, A. R., P. Setiya, A. T. Al-Sultan, and J. Pan (2022). Exploring the Impact of Air Pollution on COVID-19 Admitted Cases. *Japanese Journal of Statistics and Data Science*, **5**(1); 379–406

Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods*. Springer, New York

Carslaw, D. C. (2005). Evidence of an Increasing NO2/NOX Emissions Ratio from Road Traffic Emissions. *Atmospheric Environment*, **39**(26); 4793–4802

Casquero-Vera, J. A., H. Lyamani, G. Titos, E. Borrás, F. J. Olmo, and L. Alados-Arboledas (2019). Impact of Primary NO2 Emissions at Different Urban Sites Exceeding the European NO2 Standard Limit. *Science of the Total Environment*, **646**; 1117–1125

Engle, R. F. and C. W. J. Granger (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, **55**(2); 251–276

Faridah, R. A. N., H. Z. Hadibasyir, U. E. I. Kiat, and W. T. Pramono (2024). Spatial Analysis of Sulfur Dioxide (SO2) and Nitrogen Dioxide (NO2) Distribution Using Getis-Ord Gi* in DKI Jakarta Region, Indonesia. In *IOP Conference Series: Earth and Environmental Science*. pages 1–11

Florens, J. P., V. Marimoutou, and A. Péguin-Feissolle (2007). *Econometric Modeling and Inference*. Cambridge University Press, New York

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, New Jersey

He, C. (2024). Sulfur Dioxide in the Atmosphere: Dangers and Current Mitigation Strategies. *Science and Technology of Engineering Chemistry and Environmental Protection*, **1**(10); 1–5

Jarque, C. M. and A. K. Bera (1987). A Test for Normality of Observations and Regression Residuals. *International Statistical Review*, **55**(2); 163–172

Jion, M. M. M. F., J. N. Jannat, M. Y. Mia, M. A. Ali, M. S. Islam, S. M. Ibrahim, and A. R. M. T. Islam (2023). A Critical Review and Prospect of NO2 and SO2 Pollution Over Asia: Hotspots, Trends, and Sources. *Science of the Total Environment*, **876**; 162851

Johansen, S. (1988). Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control*, **12**(2–3); 231–254

Kurniawan, R., A. R. B. Alamsyah, A. Fudholi, A. Purwanto, B. Sumargo, P. U. Gio, and A. E. H. Susanto (2023). Impacts of Industrial Production and Air Quality by Remote Sensing on Nitrogen Dioxide Concentration and Related Effects: An Econometric Approach. *Environmental Pollution*, **334**; 122212

Lee, C., A. Richter, H. Lee, Y. J. Kim, J. P. Burrows, Y. G. Lee, and B. C. Choi (2008). Impact of Transport of Sulfur Dioxide from the Asian Continent on the Air Quality Over Korea During May 2005. *Atmospheric Environment*, **42**(7); 1461–1475

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, Berlin, Heidelberg

Muhtar, G. A., H. Mariati, Ridwan, S. Hasanuddin, R. Sandi, and N. Okviyani (2021). COVID-19 Impact to Air Quality Transformation Based on Landsat: A Case from Makassar City, Indonesia. In *International Conference on Research Collaboration of Environmental Science*, volume 802. page 012038

Nouri, F., M. Taheri, M. Ziaddini, J. Najafian, K. Rabiei, A. Pourmoghadas, and N. Sarrafzadegan (2023). Effects of Sulfur Dioxide and Particulate Matter Pollution on Hospital Admissions for Hypertensive Cardiovascular Disease: A Time Series Analysis. *Frontiers in Physiology*, **14**; 1124967

Pena, D., G. C. Tiao, and R. S. Tsay (2001). *A Course in Time Series Analysis*. John Wiley and Sons, New York

SAS/ETS (2014). *User Guide The VARMAX Procedure. In SAS/ETS® 13.2 User's Guide: High-Performance Procedures*. SAS Institute Inc

Tsay, R. S. (2005). *Analysis of Financial Time Series*. John Wiley and Sons, Hoboken, New Jersey, 2 edition

Tsay, R. S. (2014). *Multivariate Time Series Analysis: With R and Financial Applications*. John Wiley and Sons, Hoboken, New Jersey

Warsono, W., E. Russels, W. Wamiliana, W. Widiarti, and M. Usman (2019a). Modeling and Forecasting by the Vector Autoregressive Moving Average Model for Export of Coal

and Oil Data (Case Study from Indonesia Over the Years 2002–2017). *International Journal of Energy Economics and Policy*, **9**(4); 240–247

Warsono, W., E. Russels, W. Wamiliana, W. Widiarti, and M. Usman (2019b). Vector Autoregressive with Exogenous Variable Model and Its Application in Modeling and Forecasting Energy Data: Case Study of PTBA and HRUM Energy. *International Journal of Energy Economics and Policy*, **9**(2); 390–398

Wei, W. W. S. (2006). *Time Series Analysis: Univariate and Multivariate Methods*. Pearson Education, Boston, 2 edition

Wei, W. W. S. (2019). *Multivariate Time Series Analysis and Applications*. John Wiley and Sons, Hoboken, New Jersey

White, P. F., R. P. Conway, D. G. Byrne, D. M. R. O'Riordan, and B. M. Silke (2020). Air Pollution and Comorbidity Burden Influencing Acute Hospital Mortality Outcomes in a Large Academic Teaching Hospital in Dublin, Ireland: A Semi-Ecologic Analysis. *Public Health*, **186**; 164–169

Yang, C.-H., P.-H. Chen, C.-S. Yang, and L.-Y. Chuang (2024). Analysis and Forecasting of Air Pollution on Nitrogen Dioxide and Sulfur Dioxide Using Deep Learning. *IEEE Access*, **12**; 165236–165252