Available online at http://iorajournal.org/indx.php/orics /index

Operations Research: International Conference Series

e-ISSN: 2722-0974
p-ISSN: 2723-1739

Vol. 6, No. 4, pp. 169-182, 2025

Investigating Long-Run and Short-Run Dynamics of Palm Oil Production with Key Factors Using the VECM Method

Lathifah Zahra[1*], Gustriza Erda[2]

[1,2]*Statistics Study Program, Faculty of Mathematics and Natural Science, Riau University, Riau, Indonesia*

**Corresponding author email: lathifah.zahra5955@student.unri.ac.id*

Abstract

This study investigates the long-run and short-run relationships among palm oil production, rainfall, the number of bunches per palm (NOB), and average bunch weight (BTR) using the Vector Error Correction Model (VECM). Monthly data from 2015 to 2024 obtained from PT Perkebunan Nusantara IV (PTPN IV) Regional III, Sei Rokan Estate, were analyzed. Descriptive statistics indicate high variability in rainfall and relatively balanced distributions for production, NOB, and BTR. The Augmented Dickey-Fuller (ADF) test confirmed that all variables became stationary after first differencing, and the Johansen cointegration test identified three cointegrating relationships, suggesting both short-run and long-run linkages among variables. The VECM estimation results reveal positive long-run relationships for palm oil production (ECT = 0.052), rainfall (ECT = 0.090), and NOB (ECT = 0.042), indicating that these variables move toward long-run equilibrium in the same direction. In the short run, previous rainfall significantly affects both current palm oil production and NOB, with coefficients of 0,203 and 0,178, respectively, highlighting the critical role of rainfall fluctuations in influencing short-term productivity and fruit development. Model evaluation using the Root Mean Square Error (RMSE) shows low prediction errors across all variables, with rainfall having the highest RMSE (1.334) and NOB the lowest (0.962), confirming the model's strong predictive performance. Overall, the findings demonstrate that the VECM approach effectively captures both long-run equilibrium and short-run dynamics among key determinants of palm oil productivity in the Sei Rokan plantation.

Keywords: Palm oil production, rainfall, number of bunches, bunch weight, vector error correction model

1. Introduction

PT Perkebunan Nusantara IV (PTPN IV) is a state-owned enterprise engaged in the plantation sector, focusing primarily on palm oil commodity management. PTPN IV Regional III oversees 20 estates located across several regencies and cities in Riau Province. As a strategic plantation commodity, palm oil serves as the core of PTPN IV's operational activities. The company manages extensive oil palm plantations and carries out all stages of production, including cultivation, maintenance, harvesting, and processing into value-added derivative products. Palm oil plays a vital role as the company's main production support and contributes significantly to achieving national targets for crude palm oil (CPO) supply (PTPN IV, 2025).

Palm oil production is influenced by several interrelated factors, including rainfall, the number of bunches per palm (NOB), and the average bunch weight (BTR). Appropriate rainfall is essential, as excessive rainfall may hinder female flower formation, while insufficient rainfall can reduce vegetative growth, ultimately lowering production. According to Krisdiarto et al. (2017), optimal rainfall conditions support the formation of more fruit bunches, thereby increasing NOB and contributing directly to higher Fresh Fruit Bunch (FFB) yields. Furthermore, BTR plays a substantial role since heavier bunches result in greater total FFB production. These factors are interdependent and collectively determine palm oil productivity. Therefore, a time series analysis method is needed to identify the temporal relationships among palm oil production, rainfall, NOB, and BTR.

Time series analysis observes variables over specific time intervals. Typically, this analysis employs the ARIMA model, which is suitable for univariate data. However, as stated by Prahutama et al. (2019), when the data involve more than one variable, a multivariate time series approach such as the Vector Autoregressive (VAR) model is more appropriate. The VAR model is used when each variable in the system acts as an endogenous variable influenced by the lagged values of all other endogenous variables.

The VAR model requires several assumptions to be met, including stationarity in mean and variance, white noise residuals, and the absence of correlation among error terms. According to Juselius (2019), if the data in a VAR model are found to be non-stationary, the appropriate model to use is the Vector Error Correction Model (VECM). VECM is a multivariate time series model that captures both short-term and long-term relationships among non-stationary variables that exhibit cointegration (Winarno et al., 2021). Although this method is widely applied in econometric studies, such as those involving inflation data, its application to non-economic data such as palm oil productivity remains relatively rare.

Based on the above background, this study aims to analyze the relationship among palm oil production, rainfall, NOB, and BTR in PTPN IV Regional III, Sei Rokan Estate, using the VECM method. This method was chosen because it effectively handles non-stationary and cointegrated data while distinguishing between long-term and short-term effects among variables. Consequently, the VECM approach provides richer insights than other forecasting methods that focus solely on historical data patterns without considering the interdependence of influencing factors. The study utilizes monthly historical data from 2015 to 2024, including production volume (kg), rainfall (mm), NOB (bunches/palm), and BTR (kg). The data were processed using R-Studio software with the available VAR and VECM packages. The objective of this study is to examine the short-term and long-term effects among palm oil production, rainfall, NOB, and BTR during the 2015–2024 period.

2. Literature Review

2.1. Palm Oil Production

Palm oil cultivation involves a long and structured process to achieve optimal yields. Sustainable production requires balanced management of production factors and harvesting practices. Oil palm yield performance is influenced by three aspects: genetic yield, site yield potential, and actual yield. Genetic yield reflects the maximum potential of a palm variety, site yield potential depends on environmental conditions, and actual yield represents the real output under natural and management limitations (Lubis, 2008).

In general, internal factors affecting palm oil productivity consist of biological and socio-economic aspects. Biological factors include soil fertility, seed quality, fertilizer and pesticide use, and weed control, whereas socio-economic factors involve production costs, market prices, labor availability, education level, income, institutional support, and access to credit (Soekartawi, 2002).

2.2. Rainfall

According to BMKG (2025), rainfall refers to the amount of rainwater that falls in a particular area over a certain period and is usually expressed in millimeters (mm). Each 1 mm of rainfall indicates that 1 liter of water has fallen on an area of 1 square meter. Information about rainfall is useful in various fields such as agriculture, hydrology, and climatology. In oil palm plantations, rainfall is one of the main factors influencing plant growth and production yield.

Rainfall plays an important role in determining the productivity of oil palm. Excessive rainfall can interfere with the formation of female flowers, which later develop into fruit. Conversely, if rainfall is too low, the plants will experience prolonged water shortages, which can hinder vegetative growth. Therefore, the ideal rainfall level for oil palm growth is around 2000–2500 mm per year, with an effective water requirement ranging between 1300–1500 mm per year (Lubis, 2008).

2.3. Number of Fresh Fruit Bunces/Palm (NOB)

Fresh Fruit Bunches (FFB) are parts of the oil palm plant that contain palm fruits. FFBs are harvested as the main product of oil palm before being further processed into various products such as crude palm oil (CPO), palm kernel oil, and their derivatives. In the palm oil processing industry, FFB serves as the primary raw material; therefore, its quantity and quality must be maintained to ensure smooth production processes (Krisdiarto et al., 2017).

One of the key indicators in assessing oil palm productivity is the number of bunches per palm, or Number of Bunches (NOB). The higher the NOB, the greater the potential CPO production. Regular monitoring of NOB helps evaluate the effectiveness of maintenance practices, climatic conditions, and fertilization patterns, thereby supporting efficiency and sustainability in palm oil production (Hasibuan, 2020).

2.4. Average Bunch Weight (BTR)

The Average Bunch Weight (BTR) represents the average weight of each Fresh Fruit Bunch (FFB) harvested from oil palm. It is calculated by dividing the total weight of harvested FFBs by the number of bunches collected. BTR is useful for evaluating production performance and serves as a basis for future plantation management planning.

According to PKT Group (2024), BTR is closely related to palm oil yield, as heavier bunches indicate higher total FFB production. BTR is influenced by factors such as plant age, nutrient availability, rainfall, and field maintenance quality. Therefore, consistent BTR monitoring is essential not only to estimate production potential but also to assess the effectiveness of fertilization programs and plantation management strategies.

2.5. Vector Error Correlation Model (VECM)

One of the multivariate time series analysis methods used for forecasting is the Vector Autoregressive (VAR) model, which extends the autoregressive approach to capture relationships among multiple variables. In building a VAR model, all variables must be stationary in their mean. If a variable is non-stationary, differencing is required to stabilize it. Non-stationary variables may also exhibit cointegration, which indicates a long-term relationship among them. According to Wei et al. (2006), when cointegration exists, the appropriate method to use is the Vector Error Correction Model (VECM).

VECM is used to analyze both short-term and long-term relationships among variables. It is an extension of the VAR model with an added error correction component to account for short-term deviations from long-term equilibrium. The equation form of the VAR model with a lag length of p is as follows:

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + \varepsilon_t \quad (1)$$

Description:

Y_t : A vector $(Y_{1t}, Y_{2t}, \dots, Y_{Kt})$ with dimensions $K \times 1$.
A_i : A coefficient matrix with dimensions $K \times K$ for the i-th lag, with $i = 1, 2, \dots, p$.
ε_t : An error vector with dimensions $K \times 1$ $(\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{Kt})$.
K : The number of endogenous variables.
p : The lag order.

The VECM model with $(p-1)$ lags is obtained by differencing the VAR (p) model, which can be expressed as follows:

$$\Delta Y_t = \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-2} + \cdots + \Gamma_{p-1} \Delta Y_{t-p} + \varepsilon_t \quad (2)$$

Description:

Π : Cointegration matrix with dimensions $K \times K$ $(I_K - A_1 - \cdots - A_p)$.
Γ_1 : Coefficient matrix with dimensions $K \times K$ (-$(A_{i+1} + \cdots + A_p)$), with i=1, 2, …, p-1).

In equation (2), Π can be decomposed as $\Pi = \alpha\beta^T$, with β^Trepresents the *Error Correction Term* (ECT) that captures the long-run equilibrium relationships among the variables. Therefore, equation (2) can also be rewritten as:

$$\Delta Y_t = \alpha\beta^T Y_{t-1} + \Gamma_1 \Delta Y_{t-2} + \cdots + \Gamma_{p-1} \Delta Y_{t-p} + \varepsilon_t \quad (3)$$

2.6. Data Standardization

Data standardization is applied when variables have different measurement units. The Min-Max Normalization method is used to scale data so that all variables are on a comparable range, typically with a mean of 0 and variance of 1. This method not only equalizes data scales but also improves analysis accuracy (Srinivas et al., 2024). The Min-Max Normalization formula is as follows:

$$x_{new} = \frac{x - x_{min}}{x_{max-x_{min}}} \quad (4)$$

Description:

x_{new} : Normalized value.
x : Original value to be normalized.
x_{min} : Minimum value in the dataset.
x_{max} : Maximum value in the dataset.

2.7. Stationarity Test

Stationarity in the mean of time series data can be tested using a unit root test such as the Augmented Dickey-Fuller (ADF) test. The ADF test aims to determine whether unit roots exist in the model. The hypotheses used in the ADF test are as follows:

a) Null hypothesis (H_0): $H_0 = \delta = 0$ (there is a unit root, meaning the data are non-stationary in the mean).
b) Alternative hypothesis (H_1): $H_1 = \delta \neq 0$ (there is no unit root, meaning the data are stationary in the mean).

According to Wei et al. (2006), the ADF test statistic can be formulated as follows:

$$\tau_{statistic} = \frac{\hat{\delta} - 1}{se\ (\hat{\delta})} \tag{5}$$

Description:

τ : $t_{statistic}$ value.
$\hat{\delta}$: Estimated value of the parameter δ.
$se\ (\hat{\delta})$: Standard error of the parameter $\hat{\delta}$.

The testing criterion is determined by comparing the t-statistic value with the critical value of the Dickey-Fuller or t-table. The t-table value is based on the significance level (α) and the number of observations (n). The decision criteria for the ADF test are as follows:

1) If $\tau_{statistic} < \tau_{critical}(\tau_{\alpha,n})$, reject H_0; the data are stationary in the mean.
2) If $\tau_{statistic} > \tau_{critical}(\tau_{\alpha,n})$, fail to reject H_0; the data are non-stationary in the mean.

2.8. Differencing

The differencing transformation is applied when time series data are non-stationary in the mean. Its purpose is to stabilize the mean so that the subsequent analysis becomes more accurate. If the data have no seasonal pattern, non-seasonal differencing is used, while data with seasonal patterns require seasonal differencing. In this process, a backward shift operator (B) is used to shift the data value to the previous period, which is mathematically expressed as follows (Makridakis et al., 1999):

$$BY_t = Y_{t-1} \tag{6}$$

Description:

B : Backward shift.
Y_t : Value Y at period t.
Y_{t-1} : Value Y at period $t-1$.

The application of this operator results in differencing of order *d*, denoted as *I*(*d*), which can be expressed by the following equation:

$$\Delta^d Y_t = (1-B)^d Y_t \tag{7}$$

Description:

$\Delta^d Y_t$: d-th order differencing at period t.
$(1-B)^d$: d-th order differencing operator.
Y_t : Data value at period t.

2.9. Optimum Lag Selection

The selection of lag length in the VAR model aims to avoid autocorrelation problems, which are often influenced by the number of lags used. Common criteria used are the Akaike Information Criterion (AIC) and Schwarz Information Criterion (SIC), where the optimal lag is determined by choosing the model with the smallest value (Qu & Perron, 2007). The formulas for determining the optimal lag using AIC and SIC are as follows:

$$AIC = -\frac{2l}{n} + \frac{2k}{n} \tag{8}$$

$$SIC = -\frac{2l}{n} + \frac{k \ln n}{n} \tag{9}$$

Where *l* can be obtained using the following formula:

$$l = -\frac{n}{2}\left(1 + \ln(2\pi) + \ln\left(\frac{SSE}{n}\right)\right) \tag{10}$$

Description:

l : Log-likelihood function.
k : Number of independent variables.
n : Number of observations.
SSE : Sum of squared errors.

2.10. Model Stability Test

The stability test of the VAR model is conducted by examining the unit roots of the characteristic polynomial. A VAR model is considered stable if all roots have a modulus value less than 1, ensuring the model does not produce explosive patterns over time. If any root has a modulus ≥ 1, the model is deemed unstable and unsuitable for further analysis or forecasting. The modulus represents the absolute value of the characteristic roots, which may be real or complex (Ambala & Anarfo, 2022).

The concept of the modulus is used to assess stability, so the next step is to formulate the VAR(p) model. According to Lütkepohl (2005), the VAR(p) model equation can be expressed as follows:

$$\vec{Y}_t = \vec{\phi}_0 + \phi_1\vec{Y}_{t-1} + \phi_2\vec{Y}_{t-2} + \cdots + \phi_p\vec{Y}_{t-p} + \phi_1\vec{\varepsilon}_t \tag{11}$$

Description:

$\vec{Y}_t$: Data vector of size $k \times 1$ at period t.
$\vec{\phi}_0$: Intercept vector.
ϕ_i : VAR parameter matrix at lag i of size $k \times k$ (with $i = 1,2,\dots,p$).
$\vec{\varepsilon}_t$: Error vector at period t of size $k \times 1$.
k : Number of variables.
p : Lag length.

A VAR(p) model is considered stable if all eigenvalues (ϕ_i) have modulus values less than one, which is equivalent to:

$$det\left(I_{kp} - \phi z\right) = \det(I_{kp} - \phi_1 z - \cdots - \phi_p z^p) \neq 0 \tag{12}$$

The values of z that satisfy this equation are the roots of the model.

2.11. Johansen Cointegration Test

Cointegration testing is conducted to determine the existence of equilibrium relationships among variables, whether stationary or non-stationary. One common approach for testing cointegration is the Johansen method. The test is performed at a 5% significance level by comparing the trace statistic (λ_{trace}) with its critical value. The Johansen method uses two matrices, α and β, to evaluate restrictions on the cointegration vectors, which can be formulated as follows (Johansen, 1988):

$$\Pi = \alpha\beta^\tau \tag{13}$$

Description:

α : Error correction coefficient matrix of size ($K \times r$).
β : Cointegration parameter matrix of size ($K \times r$).

The hypotheses used in the cointegration test are as follows:

a) Null hypothesis (H_0): $rank$ (Π) ≤ r_i (no cointegration vectors exist).
b) Alternative hypothesis (H_1): $rank$ (Π) ≥ r_i (there are cointegration vectors).

Mathematically, the trace statistic is formulated as:

$$\lambda_{trace}(r|K) = -n\sum_{i=r_0+1}^{K} \ln(1 - \hat{\lambda}_i) \tag{14}$$

Description:

$\lambda_{trace}(r|K)$: Trace statistic for testing cointegration rank.
n : Number of observations.
K : Number of endogenous variables.
r : *Rank* of Π, with r_0 = 0, 1, 2, …, K-1.
$\hat{\lambda}_i$: Estimated eigenvalue from the Π matrix estimation.

The testing criteria are as follows:

1) If $\lambda_{trace} > \lambda_{critical}\left(\lambda_{(k-r)}\right)$, reject H_0; there is a cointegration relationship among the variables.
2) If $\lambda_{trace} < \lambda_{critical}\left(\lambda_{(k-r)}\right)$, fail to reject H_0; no cointegration relationship exists among the variables.

2.12. VECM Estimation

The Vector Error Correction Model (VECM) is used to analyze time series data that are non-stationary but cointegrated, allowing the study of both short- and long-term relationships among variables. It requires all variables to be stationary at first differencing and to exhibit cointegration, ensuring the model reflects both temporary fluctuations and long-run equilibrium. The Error Correction Term (ECT) represents the adjustment toward long-run balance when short-term deviations occur. If the ECT coefficient is significant, the model is valid; otherwise, it needs respecification (Granger & Newbold, 1978).

In the estimation of the Vector Error Correction Model (VECM), the hypotheses used are as follows:
a) Null hypothesis (H_0): $\beta_j = 0$ (no long-term relationship among the variables).
b) Alternative hypothesis (H_1): $\beta_j \neq 0$(a long-term relationship exists among the variables).

The significance of both the Error Correction Term (ECT) and the lagged independent variables is tested using the t-test, formulated as:

$$t_{statistic} = \frac{\hat{\beta}}{SE(\hat{\beta})} \tag{15}$$

Description:
t : $t_{statistic}$ value.
$\hat{\beta}$: Estimated coefficient from the VECM model.
SE : Standard error of the estimated coefficient.

The testing criteria are determined by comparing the calculated t-statistic with the t-table value. The t-table is obtained based on the significance level (α) and degrees of freedom (df), calculated as $n - K$, in which n is the number of observations and K is the number of endogenous variables. The decision rules are as follows:

The testing criteria are as follows:
1) If $t_{statistic} > t_{table}(t_{\alpha,df})$, reject H_0; indicating a short- or long-term relationship exists.
2) If $t_{statistic} < t_{table}(t_{\alpha,df})$, fail to reject H_0; indicating no short- or long-term relationship exists.

2.13. Model Evaluation

After obtaining the model equation, it is necessary to evaluate its accuracy. One common method used is the Root Mean Square Error (RMSE), which measures the average error of the VECM model. RMSE is defined as the square root of the average squared difference between the predicted and actual values. A smaller RMSE value, or one closer to zero, indicates that the model performs better in forecasting the data (Van Dao et al., 2019). The mathematical formula for RMSE is presented in Equation 16.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y} - y)^2} \tag{16}$$

Description:
$RMSE$: Root mean square error.
$\hat{y}$: Predicted value from the model.
y : Actual (observed) value.

3. Materials and Methods

3.1. Data and Sources

This study employs historical data obtained from PTPN IV Regional III, Sei Rokan Estate, covering the period from January 2015 to December 2024. The dataset comprises variables including palm oil production volume (kg), rainfall (mm), NOB (bunches/tree), and BTR (kg). The data were sourced from the Production Planning Assistant of PTPN IV Regional III, who systematically records and manages information related to palm oil yield performance. The research applies a literature review approach by examining relevant books and academic journals to support the analysis framework. The Vector Error Correction Model (VECM) is employed to analyze the interrelationships among production output, rainfall, NOB, and BTR.

3.2. Stages of Analysis

In studies employing the Vector Error Correction Model (VECM), data analysis is conducted using R-Studio software, following several systematic stages as outlined below:
1) Inputting the dataset to be analyzed.
2) Standardizing the data using scaling techniques to ensure all variables are within a comparable range.
3) Conducting stationarity tests for each variable using the Augmented Dickey-Fuller (ADF) test to verify that the variables are stationary in mean and free from unit roots.
4) Applying differencing based on the stationarity test results to achieve stable or constant data over time.
5) Determining the optimal lag length by selecting the smallest Information Criterion (IC) value to obtain the best model specification.
6) Testing model stability through unit root analysis to ensure the model remains stable.
7) Performing the Johansen cointegration test to determine whether short-run or long-run relationships exist among the variables.
8) Estimating the VECM to capture both short-run and long-run dynamics among variables based on the Error Correction Term (ECT) and t-statistic values.
9) Developing mathematical equations to represent the interrelationships among the variables used in the analysis.
10) Evaluating the model's performance using the Root Mean Square Error (RMSE) to assess forecasting accuracy.

4. Results and Discussion

4.1. Descriptive Data Analysis

Descriptive analysis was conducted to summarize and explain the main characteristics of the dataset, including the maximum, minimum, mean, and median values of palm oil production volume (kg), rainfall (mm), number of bunches per tree (NOB), and average bunch weight (BTR), as presented in Table 1.

Table 1: Descriptive Analysis Output

Variable	Minimum	Maximum	Mean	Median
Palm Oil Production	10,755,100	32,482,670	20,657,939	20,839,830
Rainfall	9.0	491.0	224.6	206.6
NOB	0.553	1.872	1.045	1.009
BTR	8.827	19.480	15.430	16.436

The results show that palm oil production ranged from 10,755,100 kg to 32,482,670 kg, with an average of approximately 20,657,939 kg and a median of 20,839,830 kg, indicating a relatively balanced distribution. Rainfall varied between 9 mm and 492 mm, with an average of 224.6 mm and a median of 206.6 mm, reflecting high variability in rainfall intensity. The NOB ranged from 0.533 to 1.872, with an average of 1.045 and a median of 1.009, suggesting stable data distribution. Meanwhile, the BTR ranged from 8.827 kg to 19.480 kg, with an average of 15.430 kg and a median of 16.436 kg, indicating that most bunches tended to be heavier than the mean.

4.2. Data Standardization

Table 2: Units of Research Variables

Variable	Unit
Palm Oil Production	Kilograms
Rainfall	Milimeters
NOB	Bunches/Tree
BTR	Kilograms

Data standardization using scaling is performed prior to the stationarity test to equalize the measurement scales among the variables used in the analysis. This step is necessary because each variable has different units and value ranges. In this study, production output is measured in kilograms (kg), rainfall in millimeters (mm), the number of bunches (NOB) in bunches per tree, and the average bunch weight (BTR) in kilograms (kg). Without standardization, variables with larger values could dominate the analysis results, leading to biased interpretations. By applying standardization, all variables are converted to a comparable scale, ensuring that the stationarity test is conducted more objectively and produces more accurate results.

4.3. Stasionarity Test

The stationarity test was conducted to determine whether the data are stationary or non-stationary in terms of their mean. This test serves as a prerequisite that must be satisfied before developing the Vector Error Correction Model (VECM). The stationarity test applied in this study is the Augmented Dickey-Fuller (ADF) test, and the results are presented in Table 3.

Table 3: Stationarity Test Results

Variable	t-Statistic	Critical Value (α =5%)	Stationarity Result
Palm Oil Production	-3.933	-2.891	Stationary
Rainfall	-8.163	-2.891	Stationary
NOB	-3.553	-2.891	Stationary
BTR	-2.052	-2.891	Non-stationary

- Hypotheses:
  H_0: $H_0 = \delta = 0$ (there is a unit root, meaning the variable in the model is non-stationary).
  H_1: $H_1 = \delta \neq 0$ (there is no unit root, meaning the variable in the model is stationary).
- Significance Level: $\alpha = 5\% = 0.05$
- Test Statistic:

$$\tau_{statistic} = \frac{\hat{\delta} - 1}{se\,(\hat{\delta})}$$

- Critical Region:
  1. If the t-statistic ($\tau_{statistic}$) < Dickey-Fuller critical value ($\tau_{table} = \tau_{0.05,120}$), then H_0 is rejected, meaning the data are stationary in their mean.
  2. If the t-statistic ($\tau_{statistic}$) > Dickey-Fuller critical value ($\tau_{table} = \tau_{0.05,120}$), then H_0 is not rejected, meaning the data are non-stationary in their mean.
- Decision:
  a. The t-statistic (-3.933) < critical value (-2.891), thus H_0 is rejected or H_1 is accepted.
  b. The t-statistic (-8.163) < critical value (-2.891), thus H_0 is rejected or H_1 is accepted.
  c. The t-statistic (-3.553) < critical value (-2.891), thus H_0 is rejected or H_1 is accepted.
  d. The t-statistic (-2.052) > critical value (-2.891), thus H_0 is not rejected or H_1 is accepted.
- Conclusion:
  At the 5% significance level, there is sufficient evidence to reject H_0 for the palm oil production, rainfall, and NOB variables because their t-statistic values are smaller than the critical value (-2.891). This indicates that these variables do not contain unit roots and are stationary in their mean. In contrast, for the BTR variable, there is insufficient evidence to reject H_0 because its t-statistic value is greater than the critical value, indicating that the BTR variable contains a unit root and is non-stationary in its mean.

4.4. Differencing

Non-stationary data can lead to inaccurate modeling results; therefore, differencing is applied to data that exhibit non-stationarity. The differencing process is carried out based on the results of the stationarity test (Augmented Dickey-Fuller test), which indicate the presence of unit roots or non-stationary behavior. This step aims to stabilize the data by subtracting the current value from its previous period. The procedure is repeated until all data become stable or constant. The ADF test results obtained after applying the first differencing are presented in Table 4.

Table 4: First Differencing Output

Variable	t-Statistic	Critical Value (α =5%)	Stationarity Result
Palm Oil Production	-9.563	-2.891	Stationary
Rainfall	-16.486	-2.891	Stationary
NOB	-9.338	-2.891	Stationary
BTR	-10.877	-2.891	Stationary

Based on the stationarity test results presented in Table 4, the variables production volume, rainfall, NOB, and BTR were found to be stationary at the first difference. This conclusion is supported by all t-statistic ($\tau_{statistic}$) values being lower than the Dickey-Fuller critical value ($\tau_{table} = \tau_{0.05,120}$), indicating the rejection of H_0. Therefore, the differencing process was successfully achieved at the first-difference level.

4.5. Determination of Optimal Lag Length

The optimal lag selection test aims to ensure that the VECM model is free from autocorrelation issues, allowing for accurate analysis of the relationships among variables. The determination of the optimal lag is based on two information criteria: the Akaike Information Criterion (AIC) and the Schwarz Information Criterion (SIC). The optimal lag is selected based on the smallest AIC and SIC values.

Table 5: Optimal Lag Selection Output

Lag	AIC	SIC
1	31.615	32.182
2	31.631	32.576
3	31.515	32.837

Based on Table 5, which presents the lag selection results, the smallest AIC value occurs at lag 3, while the smallest SIC value is found at lag 1. This study adopts lag 1 according to the SIC criterion, taking into account the limited sample size of 120 observations. Using a longer lag, such as lag 3, may reduce the model's degrees of freedom, leading to less efficient parameter estimation, higher coefficient variance, and more insignificant parameters. Moreover, fewer degrees of freedom increase the risk of overfitting, reducing the model's validity (Hanck et al., 2025). The difference between AIC values at lag 1 and lag 3 is relatively small, supporting the choice of lag 1 as the most appropriate. This selection is expected to produce a stable model that effectively captures both short-run and long-run relationships among variables.

4.6. Model Stability Test

After determining the optimal lag length, the next stage involves evaluating the stability of the model. Stability testing is crucial because an unstable model can produce inaccurate analytical results. A VAR model is considered stable if all unit roots have modulus values less than one. When the roots are positive real numbers, the modulus equals the root value itself. However, if the roots are negative real numbers, the modulus corresponds to their absolute values. The results of the model stability test are presented in Table 6.

Table 6: Model Stability Output

Root Value	Modulus Value
0.972	0.972
0.693	0.693
0.693	0.693
0.303	0.303

The number of roots obtained from a model depends on the number of variables (k) included in the system and the selected lag length (p). Generally, the total number of roots equals $k \times p$, since the system can be represented in a matrix form of dimension $(k \times p) \times (k \times p)$. In this study, four roots were obtained because the model used four variables with one selected lag. As shown in Table 6, all modulus values are less than one, indicating that the model is stable and passes the stability test. Therefore, the model can be considered reliable for further analysis, as it exhibits no signs of instability that could distort the estimation results.

4.7. Johansen Cointegration Test

The Johansen Cointegration Test was applied to identify the existence of both short-term and long-term stable relationships between two or more variables. This test is used to examine whether a group of non-stationary time series variables are cointegrated. The results are evaluated by comparing the trace statistic with the critical value at a 5% significance level. If the trace statistic is greater than the critical value, the decision is to fail to reject H_0 or accept H_0, indicating that the variables are cointegrated. Conversely, if the trace statistic is smaller than the critical value, the decision is to reject H_0 or accept H_1, meaning that the variables are not cointegrated.

Table 7: Johansen Cointegration Test Results

Hypothesis No. of CE(s)	Trace Statistic	Critical Value (α = 5%)	Cointegration Test Result
r = 0	124.974	53.120	There are cointegrating vectors.
r =1	58.998	34.910	There are cointegrating vectors.
r = 2	22.802	19.960	There are cointegrating vectors.
r = 3	6.441	9.240	No cointegrating vectors exist.

- Hypotheses:
  H_0: $rank\ (\Pi) \leq r_i$ (no cointegrating vectors exist among the variables).
  H_1: $rank\ (\Pi) \geq r_i$ (there are cointegrating vectors among the variables).
- Significance Level: $\alpha = 5\% = 0.05$
- Test Statistic:

$$\lambda_{trace}(r|K) = -n \sum_{i=r_0+1}^{K} \ln(1 - \hat{\lambda}_i)$$

- Critical Region:
  1. If the trace statistic (λ_{trace}) > critical value ($\lambda_{(K-r)}$), then H_0 is rejected, meaning there is no cointegration among the variables.
  2. If the trace statistic (λ_{trace}) < critical value ($\lambda_{(K-r)}$), then H_0 is not rejected, indicating the presence of cointegration among the variables.
- Decision:
  a. r_0: Trace statistic (124.974) > critical value (53.120), thus H_0 is rejected or H_1 is accepted.
  b. r_1: Trace statistic (58.998) > critical value (34.910), thus H_0 is rejected or H_1 is accepted.
  c. r_2: Trace statistic (22.802) > critical value (19.960), thus H_0 is rejected or H_1 is accepted.
  d. r_3: Trace statistic (6.441) < critical value (9.240), thus H_0 is not rejected or H_1 is accepted.
- Conclusion:
  At the 5% significance level, there is sufficient evidence to reject H_0 for r_0, r_1, and r_2 since their trace statistics are greater than the corresponding critical values. This indicates that cointegration exists among the variables. Meanwhile, for r_3, there is insufficient evidence to reject H_0 because the trace statistic is smaller than the critical value, suggesting that the variables remain cointegrated. It can be concluded that there are three cointegrating relationships, namely at r_0, r_1, and r_2, which indicate the presence of both short-term and long-term interrelationships among the variables.

4.8. VECM Estimation

The Vector Error Correction Model (VECM) is an analytical method used to estimate time series data that are initially non-stationary but exhibit cointegration. This model captures both short-term and long-term relationships among variables. VECM can be applied when the data become stationary after first differencing and cointegration exists among variables.

A variable is considered to have a significant influence in the short or long run if the t-statistic exceeds the critical value ($t_{0.05,116}$ = 1.981) or if the p-value is less than the 5% significance level (α = 0.05). Conversely, if the t-statistic is smaller or the p-value is greater than α, the relationship is not significant. Long-term relationships are examined through the Error Correction Term (ECT), while short-term relationships are assessed based on the t-statistic and p-value of the independent variables. The VECM estimation results are presented in Table 8.

Table 8: VECM Estimation Output

Variable	Coefficient	t-statistic	p-value	Result
Palm Oil Production : Intercept	0.0124	0.218	0.828	Not significant.
Palm Oil Production : ECT	0.052	4.814	0.000	*Significant in the long run.
Palm Oil Production : Palm Oil Production (-1)	-0.431	-1.444	0.152	Not significant in the short run.
Palm Oil Production : Rainfall (-1)	0.203	3.216	0.000	*Significant in the short run.
Palm Oil Production : NOB (-1)	0.446	1.465	0.146	Not significant in the short run.
Palm Oil Production : BTR (-1)	0.221	0.496	0.621	Not significant in the short run.
Rainfall : Intercept	0.025	0.270	0.788	Not significant.
Rainfall : ECT	0.090	5.068	0.000	* Significant in the long run.
Rainfall : Palm Oil Production (-1)	0.854	1.753	0.082	Not significant in the short run.
Rainfall : Rainfall (-1)	-0.057	-0.551	0.583	Not significant in the short run.

Rainfall : NOB (-1)	-0.909	-1.833	0.069	Not significant in the short run.
Rainfall : BTR (-1)	-0.898	-1.236	0.219	Not significant in the short run.
NOB: Intercept	-0.001	-0.112	0.911	Not significant.
NOB : ECT	0.042	3.982	0.000	* Significant in the long run.
NOB : Palm Oil Production (-1)	-0.298	-1.017	0.311	Not significant in the short run.
NOB : Rainfall (-1)	0.178	2.871	0.000	* Significant in the short run.
NOB : NOB (-1)	0.343	1.150	0.253	Not significant in the short run.
NOB : BTR (-1)	0.075	0.171	0.864	Not significant in the short run.
BTR : Intercept	0.026	1.731	0.086	Not significant.
BTR : Palm Oil Production (-1)	-0.0147	-0.189	0.851	Not significant in the short run.
BTR : Rainfall (-1)	0.007	0.444	0.658	Not significant in the short run.
BTR : NOB (-1)	0.0246	0.311	0.757	Not significant in the short run.
BTR : BTR (-1)	-0.022	-0.191	0.849	Not significant in the short run.

The significant long-run results of the VECM estimation can be interpreted as follows:

- Hypotheses:
  H₀: $\beta_{ECT} = 0$ (no long-run relationship in palm oil production).
  H₁: $\beta_{ECT} \neq 0$ (a long-run relationship exists in palm oil production).
- Significance Level: $\alpha = 5\% = 0.05$
- Test Statistic:

$$t = \frac{\hat{\beta}}{SE(\hat{\beta})}$$

- Critical Region:
  1. If the t-statistic > t-table ($t_{\alpha,df} = t_{0.05,116}$), then H_0 is rejected, indicating a long-run relationship.
  2. If the t-statistic < t-table ($t_{\alpha,df} = t_{0.05,116}$), then H_0 is not rejected, indicating no long-run relationship.
- Decision:
  a. Palm oil production: ECT
     The t-statistic (4.814) > t-table (1.981), thus H_0 is rejected or H_1 is accepted.
  b. Rainfall: ECT
     The t-statistic (5.068) > t-table (1.981), thus H_0 is rejected or H_1 is accepted.
  c. NOB: ECT
     The t-statistic (3.982) > t-table (1.981), thus H_0 is rejected or H_1 is accepted.
- Conclusion:
  At the 5% significance level, there is sufficient evidence to reject H_0 since the t-statistics for Palm Oil Production: ECT, Rainfall: ECT, and NOB: ECT are greater than the critical value (1,981). The positive ECT coefficients for these three variables indicate the presence of a long-run equilibrium relationship, where palm oil production has a positive long-term effect of 0.052, rainfall of 0.090, and NOB of 0.042. This implies that in the long run, each variable contributes positively to restoring equilibrium whenever short-term fluctuations occur.

The significant short-run results of the VECM estimation can be interpreted as follows:

- Hypotheses:
  H_0: $\beta_{CH(-1)} = 0$ (previous rainfall has no short-run relationship with palm oil production and NOB).
  H_1: $\beta_{CH(-1)} \neq 0$ (previous rainfall has a short-run relationship with palm oil production and NOB).
- Significance Level: $\alpha = 5\% = 0.05$

- Test Statistic:

$$t = \frac{\hat{\beta}}{SE(\hat{\beta})}$$

- Critical Region:
  1. If the t-statistic > t-table ($t_{\alpha,df} = t_{0.05,116}$), then H_0 is rejected, indicating a short-run relationship.
  2. If the t-statistic < t-table ($t_{\alpha,df} = t_{0.05,116}$), then H_0 is not rejected, indicating no short-run relationship.
- Decision:
  a. Palm oil production: Rainfall(-1)
     The t-statistic (3.216) > t-table (1.981), thus H_0 is rejected or H_1 is accepted.
  b. NOB: Rainfall(-1)
     The t-statistic (2.871) > t-table (1.981), thus H_0 is rejected or H_1 is accepted.
- Conclusion:
  At the 5% significance level, there is sufficient evidence to conclude that previous rainfall has a significant short-run effect on both palm oil production and NOB. An increase of 1 mm in rainfall during the previous period results in an estimated increase of 0.203 kg in current palm oil production and 0,178 kg in current NOB. This indicates that rainfall variability in the short term plays a vital role in influencing palm oil productivity and fruit bunch development.

4.9. VECM Model

In the previous stage, the estimation of the Vector Error Correction Model (VECM) was conducted, addressing the research objectives through the obtained estimation results. However, to complement the discussion, the following section presents the mathematical form of the VECM equations. These equations aim to illustrate mathematically how the variables in this study interact both in the short term through the lagged variables and in the long term through the Error Correction Term (ECT). The general form of the VECM can be presented as shown in Equation (3). Based on the estimation results, the model obtained is a VECM (1), with the following equations constructed:

$$\Delta Y_{1,t} = 0.012 + 0.052ECT_{t-1} - 0.431Y_{1,t-1} + 0.203Y_{2,t-1} + 0.446Y_{3,t-1} + 0.221Y_{4,t-1} + \varepsilon_{1,t} \quad (17)$$

$$\Delta Y_{2,t} = 0.025 + 0.009ECT_{t-1} + 0.854Y_{1,t-1} - 0.057Y_{2,t-1} - 0.909Y_{3,t-1} - 0.898Y_{4,t-1} + \varepsilon_{2,t} \quad (18)$$

$$\Delta Y_{3,t} = -0.006 + 0.042ECT_{t-1} - 0.298Y_{1,t-1} + 0.178Y_{2,t-1} + 0.343Y_{3,t-1} + 0.075Y_{4,t-1} + \varepsilon_{4,t} \quad (19)$$

$$\Delta Y_{4,t} = 0.026 + 0.004ECT_{t-1} - 0.015Y_{1,t-1} + 0.007Y_{2,t-1} + 0.025Y_{3,t-1} - 0.022Y_{4,t-1} + \varepsilon_{3,t} \quad (20)$$

Description:
ECT : Error Correlation Term.
$\Delta Y_{1,t}$: Palm Oil Production.
$\Delta Y_{2,t}$: Rainfall.
$\Delta Y_{3,t}$: NOB.
$\Delta Y_{4,t}$: BTR.

Based on the estimation results, all variables have positive ECT values, indicating a positive effect of the Error Correction Term (ECT) on long-term adjustment. The ECT coefficient for production output (0.052) shows that about 5,2% of long-term disequilibrium is corrected each period, while rainfall (0.009) and BTR (0.004) adjust more slowly, and NOB (0.042) adjusts relatively faster.

The lagged production output variable generally has a negative effect, except in the rainfall equation (0.854), where it positively affects rainfall, though insignificantly. Lagged rainfall mostly shows a positive relationship with other variables such as production output (0.203) and NOB, but negatively affects itself (−0.057). Only the coefficients for production output and NOB are statistically significant.

Lagged NOB generally has positive coefficients in most equations, suggesting a direct relationship, except in the rainfall equation (−0.909), where it shows an inverse relationship. However, these effects are statistically insignificant. The lagged BTR variable displays mixed results: positive in the production output (0.221) and NOB (0.075) equations, but negative in rainfall (−0.898) and BTR (−0.022). None of these effects are statistically significant, indicating that previous BTR values do not significantly influence the variables in the model.

4.10. Model Evaluation

The Root Mean Square Error (RMSE) is a model evaluation metric that measures accuracy by calculating the square root of the average squared difference between predicted and actual values. A smaller RMSE value (closer to zero) indicates higher accuracy. In this study, RMSE was used because other metrics, such as Mean Absolute Percentage Error (MAPE), are unsuitable for standardized data (Morley et al., 2018).

Table 9: Root Mean Square Error Output

Variable	RMSE Value
Palm Oil Production	0.970
Rainfall	1.334
NOB	0.962
BTR	0.965

Based on the RMSE results in Table 9, each variable shows a different level of prediction error. The rainfall variable has the highest RMSE value (1,334), indicating lower prediction accuracy, while the NOB variable has the lowest RMSE value (0,962), showing higher accuracy. Overall, the RMSE values for palm oil production, rainfall, NOB, and BTR are relatively low and close to zero, indicating good model performance. Thus, the VECM model is suitable for forecasting purposes.

5. Conclussion

Based on the analysis of the relationship between production output, rainfall, number of bunches (NOB), and bunches to ripening (BTR) using the Vector Error Correction Model (VECM), several conclusions can be drawn. Using historical data from 2015 to 2024 in the Sei Rokan plantation, the results show that production output, rainfall, and NOB exhibit a positive long-run relationship with the error correction term (ECT). This indicates that these variables move toward long-term equilibrium in the same direction, meaning that an increase in one variable tends to be followed by an increase in the others, reflecting the system's adjustment toward stability over time.

In the short run, the previous period's rainfall has a significant effect on both production output and NOB. This implies that rainfall from the previous period contributes meaningfully to changes in current production and the number of bunches, suggesting that rainfall conditions play an essential role in determining short-term fluctuations in palm oil yield.Furthermore, the evaluation of the VECM model using the Root Mean Square Error (RMSE) demonstrates that the model has a low prediction error for all variables analyzed. This indicates that the VECM model performs well and is reliable for forecasting future values of production, rainfall, NOB, and BTR. Overall, the findings confirm that the VECM model is suitable for explaining both the short-term and long-term relationships among key factors influencing palm oil production in the Sei Rokan plantation.

References

Ambala, R. N., & Anarfo, E. B. (2022). A Vector Autoregression (VAR) analysis of corruption, economic growth, and foreign direct investment in Ghana. *Cogent Economics and Finance*, *10*(1).

BMKG. (2025). *Curah hujan*. BMKG. https://gaw-bariri.bmkg.go.id/index.php/karya-tulis-dan-artikel/gawsarium/262-curah-hujan

Granger, C. W. J., & Newbold, P. (1978). Spurious regressions in econometrics. *Journal of Econometrics*, *2*(2), 111–120.

Hanck, C., Arnold, M., Gerber, A., & Schmelzer, M. (2025). *Introduction to econometrics with R*. Stata Press. https://www.econometrics-with-r.org/14.6-llsuic.html

Hasibuan, H. A. (2020). Penentuan rendemen, mutu dan komposisi kimia , minyak sawit dan minyak inti sawit tandan buah segar bervariasi kematangan sebagai dasar untuk penetapan standar kematangan panen. *Jurnal Penelitian Kelapa Sawit*, *28*(3), 123–132.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, *12*(2–3), 231–254.

Juselius, K. (2019). *The Cointegrated VAR Model: Methodology and Applications*. Oxford University Press.

Krisdiarto, A. W., Sutiarso, L., & Widodo, K. H. (2017). Optimasi kualitas tandan buah segar kelapa sawit dalam proses panen-angkut menggunakan model dinamis. *Agritech*, *37*(1), 102.

Lubis, A. U. (2008). *Kelapa sawit (Elaeis guineensis Jacq.) di Indonesia*. Pusat Penelitian Marihat Bandar Kuala Pematang Siantar.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer.

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1999). *Forecasting method and applications* (2nd ed.). Erlangga.

Morley, S. K., Brito, T. V., & Welling, D. T. (2018). Measures of model performance based on the log accuracy ratio. *Space Weather*, *16*(1), 69–88.

PKT Group. (2024). *Cara menghitung berat janjang rata-rata (BJR) kelapa Sawit*. PKT Group. https://pkt-group.com/sawitnotif/ketahui-cara-menghitung-berat-janjang-rata-rata-bjr-kelapa-sawit/?utm_source=chatgpt.com

Prahutama, A., Ispriyanti, D., & Tiani Wahyu Utami, D. (2019). *Pemodelan sektor-sektor inflasi di Indonesia menggunakan Vector Autorgressive (VAR)*. *20*(1), 47–52.

PTPN IV. (2025). *Company profile*. PTPN IV. https://www.ptpn4.co.id/about

Qu, Z., & Perron, P. (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica*, *75*(2), 459–502.

Soekartawi. (2002). *Prinsip dasar ekonomi pertanian: Teori dan aplikasi* (2nd ed.). PT Raja Grafindo Persada.

Srinivas, T. A. S., Sravanthi, Y., Kumar, Y. V., & Srihith, I. V. D. (2024). Data standardization: Key to effective data integration. *Advanced Innovations in Computer Programming Languages*, *6*(1), 1–4.

Van Dao, D., Ly, H. B., Trinh, S. H., Le, T. T., & Pham, B. T. (2019). Artificial intelligence approaches for prediction of compressive strength of Geopolymer Concrete. *Materials*, *12*(6), 1–17.

Wei, W. W. (2006). *Time series analysis: Univariate and multaivariate methods*. Pearson Education.

Winarno, S., Usman, M., Warsono, Kurniasari, D., & Widiarti. (2021). Application of Vector Error Correction Model (VECM) and impulse response function for daily stock prices. *Journal of Physics: Conference Series*, *1751*(1–17).