BAREKENG: Journal of Mathematics and Its Applications
December 2025 Volume 19 Issue 4 Page 2301–2316
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss4pp2301-2316

APPLICATION OF THE PRINCIPAL COMPONENT ANALYSIS-VECTOR AUTOREGRESSIVE INTEGRATED (PCA-VARI) MODEL TO FORECASTING ECONOMIC GROWTH IN INDONESIA

Aulia Rahman Al Madani[1], Sandrina Najwa[2], Budi Nurani Ruchjana[3*]

[1,2]*Magister of Applied Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran*
[3]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran*
Jln. Raya Bandung Sumedang, km 21, Jatinangor, Sumedang, 45363, Indonesia

*Corresponding author's e-mail: * **budi.nurani@unpad.ac.id***

Article History:
Received: 3rd November 2024
Revised: 28th January 2025
Accepted: 27th May 2025
Available online: 1st September 2025

Keywords:
Economic Growth;
Forecasting;
Macroeconomic Indicator;
PCA-VARI;
Time Series.

ABSTRACT

Indonesia's economic growth has undergone significant fluctuations in recent years, driven by global shocks such as the 2020 COVID-19 pandemic, the 2013 taper tantrum, and the 2022 global energy crisis. These events underscore the urgent need for more accurate and robust forecasting models to support economic stability and policymaking. This study applies the Principal Component Analysis-Vector Autoregressive Integrated (PCA-VARI) model to forecast economic growth in Indonesia. PCA reduces seven economic variables into two principal components for ten years (2012-2022). The results show that the first component (PC_1) shows the highest correlation with the variables of Money Supply, BI Rate, and Foreign Exchange Reserves, which reflect monetary policy and financial stability. Meanwhile, the second component (PC_2) is highly correlated to the GDP Index, Exchange Rate, and Inflation variables, which reflect macroeconomic conditions. VARI, as a non-stationary multivariate time series model, is used to model the relationship between these components, with the third-order lag selected as the optimal lag based on the Akaike Information Criterion (AIC), Hannan-Quinn Criterion (HQ), and Final Prediction Error (FPE) values. The results show that the PCA-VARI(3) model is able to provide highly accurate forecasting with a MAPE of 1.21% for PC_1 and 1.34% for PC_2, and has met all the necessary model assumptions.

How to cite this article:

A. R. A. Madani, S. Najwa and B. N. Ruchjana., "APPLICATION OF THE PRINCIPAL COMPONENT ANALYSIS-VECTOR AUTOREGRESSIVE INTEGRATED (PCA-VARI) MODEL TO ECONOMIC GROWTH FORECASTING IN INDONESIA," *BAREKENG: J. Math. & App.*, vol. 19, no. 4, pp. 2301-2316, December, 2025.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*
Research Article · Open Access

1. INTRODUCTION

Indonesia's economy has undergone significant changes in recent years, with various macroeconomic factors affecting its growth rate [1]. Amidst slowing economic growth in the world, Indonesia's is still high, reaching 5.01% during 2022 and is expected to be in the range of 4.5-5.3% in 2023 [2]. Despite being one of the fastest growing economies in the world, economic growth in Indonesia has not exceeded 6.0% since 2012 [3]. This growth was influenced by various factors such as increased infrastructure investment, strong export performance, accommodative fiscal and monetary policies [4][5][6].

This study applies the Principal Component Analysis-Vector Autoregressive Integrated (PCA-VARI) model to better understand the factors driving Indonesia's economic growth and forecast its future growth rate. Principal Component Analysis (PCA) is a statistical technique widely used to reduce the dimensionality of complex datasets while preserving essential information and the main variability contained in the data [7]. PCA simplifies complex datasets, facilitating clearer interpretations and more efficient modeling. The Vector Autoregressive (VAR) model is commonly employed to forecast multivariate time series data when the data are stationary. However, if the data are non-stationary, the Vector Autoregressive Integrated (VARI) model is used instead, capturing dynamic linear correlations among variables and making it suitable for prediction and interpretation purposes [8].

The combination of PCA and VAR has been applied in various previous studies. In the analysis of the interaction between environmental, economic, and social in the context of sustainable development in China, PCA was used to construct the development index of each subsystem, which was then analyzed with a VAR model to understand the dynamics of the relationship between subsystems [9]. In another study, PCA-VAR was used to evaluate development sustainability through analyzing the interaction between economic development, social, and environmental conditions. PCA helps identify key indices in each aspect, while VAR facilitates understanding of the dynamic impact of changes in each subsystem [10]. In research related to climate data in the West Java region, PCA was utilized to reduce climate variables into two main components, while VARI was used to forecast the dynamics and interactions between these components [11].

Based on the above explanation, this study applies the PCA-VARI approach with a different focus from previous studies. PCA is used to simplify the complexity of economic variables, while VARI is focused on forecasting economic growth. This approach not only overcomes the challenges of high-dimensional and non-stationary economic data, but also provides insights that can be used by policymakers in formulating more effective strategies to promote sustainable economic development in Indonesia.

2. RESEARCH METHODS

2.1 Research Data

The data used is sourced from economic growth data in Indonesia from 2012-2022 with monthly time intervals. The data is obtained from secondary data from the BPS - Statistics Indonesia, Bank Indonesia and the Federal Reserve Bank. Table 1 shows the variables used in this study.

Table 1. Research Variables

Variables	Variables Definition	Unit	Source
Money Supply (X_1)	Money supply expressed in M2	Billion Rupiah	BPS - Statistics Indonesia
Interest Rate (X_2)	BI Rate	Percentage	BPS - Statistics Indonesia
Economic Activity (X_3)	GDP Index	Index	Federal Reserve Bank
Exchange Rate (X_4)	Monthly average IDR per USD	Rupiah	Bank Indonesia
Inflation (X_5)	Annual inflation rate (yoy)	Percentage	Bank Indonesia
Foreign Exchange Reserves (X_6)	Assets held by the central bank in foreign currency	Million USD	Bank Indonesia

Variables	Variables Definition	Unit	Source
Stock Transactions and Index (X_7)	Value of stock transactions on the stock exchange	Billion Rupiah	BPS - Statistics Indonesia

***Data source**: (BPS - Statistics Indonesia, Bank Indonesia and the Federal Reserve Bank.)*

2.2 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a statistical method commonly used to test and diagnose correlated variables. The PCA method is used for data that has many variables and has a correlation between the variables. PCA calculation is seen from the eigenvalue [12]. If given random $X = [X_1, X_2, \cdots, X_p]$ which has a covariance matrix $\sum$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots > \lambda_p \geq 0$ then the PCA can be represented using the following **Equation (1) [12]**:

$$\begin{gathered} Y_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \\ Y_2 = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \\ \vdots \\ Y_p = a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p \end{gathered} \tag{1}$$

where $Y_1, Y_2, \cdots, Y_p$ are the principal components, each formed by a linear combination of the original variables $X_1, X_2, \cdots, X_p$. The coefficients a_{ij} represent the contribution of each variable X_j to the i^{th} principal component Y_i. These coefficients are derived from the eigenvectors of the covariance matrix. The goal of PCA is to transform correlated variables into uncorrelated components that capture the maximum variance.

PCA aims to reduce data and eliminate fewer dominant factors without reducing the original data variables, the steps taken for PCA are as follows [13]:

1. Calculate variance and covariance matrices
2. Calculate the eigenvalue of the covariance matrix
3. Determining the value of the Principal Component proposition.
4. Calculating factor loading from eigenvalues.

The Kaiser Meyer Olkin (KMO) test is a statistical method used before determining the main components of the KMO correlation matrix. KMO is used to test the feasibility of factor analysis with the KMO test scale ranging from 0 to 1, where for KMO values greater than 0.5, factor analysis is feasible to use and vice versa if the KMO value is less than 0.5 it is not feasible to use. Therefore, for the factor analysis process to be considered feasible, the KMO value must be greater than 0.5. The KMO is expressed using **Equation (2) [14]**:

$$\text{KMO} = \frac{\sum_i^n \sum_{j\neq 1}^n r_{ij}^2}{\sum_i^n \sum_{j\neq 1}^n r_{ij}^2 + \sum_i^n \sum_{j\neq 1}^n a_{ij}^2} \tag{2}$$

with $i = 1,2,3, \ldots$, and $j = 1,2,3, \ldots$

r_{ij}^2 : simple correlation coefficient of variables i and j.
a_{ij}^2 : partial correlation coefficient of variables i and j.

2.3 Vector Autoregressive (VAR)

Vector Autoregressive (VAR) is one of the methods used for time series analysis involving several endogenous variables explained by lag [15]. The VAR model is used on stationary data. The general form of the VAR model with order p, denoted as VAR(p) is given in **Equation (3)**, with the corresponding vector and matrix representation provided in **Equation (4) [16]**:

$$Z_t = \phi_0 + \Phi_1 Z_{t-1} + \cdots + \Phi_p Z_{t-p} + a_t \tag{3}$$

with

$$Z_t = \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ \vdots \\ Z_{k,t} \end{bmatrix}, Z_{t-i} = \begin{bmatrix} Z_{1,t-i} \\ Z_{2,t-i} \\ \vdots \\ Z_{k,t-i} \end{bmatrix}, \phi_0 = \begin{bmatrix} \phi_{10} \\ \phi_{20} \\ \vdots \\ \phi_{k0} \end{bmatrix}, \phi_p = \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1k} \\ \phi_{12} & \phi_{22} & \cdots & \phi_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{1k} & \phi_{k2} & \cdots & \phi_{kk} \end{bmatrix}, a_t = \begin{bmatrix} a_{1,t} \\ a_{2,t} \\ \vdots \\ a_{k,t} \end{bmatrix} \tag{4}$$

where Z_t is a $k \times 1$ vector at time t, Z_{t-i} refers to a $k \times 1$ vector with $i = 1,2,\cdots,p$, the notation ϕ_0 represents a constant vector of size $k \times 1$, the notation $\phi_1, \phi_2, \cdots, \phi_p$ is a VAR matrix up to p with dimension $k \times k$ and a_t is a $k \times 1$ vector at time t, assumed to be normally distributed $a_t \sim N(0, \sigma^2)$. The advantages of the VAR method are as follows **[15]**:

1. Does not differentiate between independent and dependent variables.
2. Simple estimation using the Ordinary Least Squares (OLS) method.
3. Forecasting results using the VAR method are often better in many cases compared to complex simultaneous equation models.

2.4 Vector Autoregressive Integrated (VARI)

The VARI model is an extension of the Autoregressive Integrated (ARI) model. If the data used is non-stationary, the differencing process is applied **[17]**. The differencing is performed d times until the data becomes stationary, resulting in a VARI(p, d) model with k variables. If $p = 1$ and $d = 1$, the VARI(1,1) model can be expressed in matrix form as shown in **Equation (5)**:

$$\begin{bmatrix} \dot{Z}_{1,t} \\ \dot{Z}_{2,t} \\ \vdots \\ \dot{Z}_{k,t} \end{bmatrix} - \begin{bmatrix} \dot{Z}_{1,t-1} \\ \dot{Z}_{2,t-1} \\ \vdots \\ \dot{Z}_{k,t-1} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1k} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{k1} & \phi_{k2} & \cdots & \phi_{kk} \end{bmatrix} \left(\begin{bmatrix} \dot{Z}_{1,t-1} \\ \dot{Z}_{2,t-1} \\ \vdots \\ \dot{Z}_{k,t-1} \end{bmatrix} - \begin{bmatrix} \dot{Z}_{1,t-2} \\ \dot{Z}_{2,t-2} \\ \vdots \\ \dot{Z}_{k,t-2} \end{bmatrix} \right) + \begin{bmatrix} a_{1,t} \\ a_{2,t} \\ \vdots \\ a_{k,t} \end{bmatrix} \tag{5}$$

or

$$\dot{Z}_t - \dot{Z}_{t-1} = \Phi_1 (\dot{Z}_{t-1} - \dot{Z}_{t-2}) + a_t \tag{6}$$

Assuming that the error term follows a normal distribution, such that $a_t \sim N(0, \sigma^2 I)$ and $Y_t = \Delta Z_t = \dot{Z}_t - \dot{Z}_{t-1}$ Then, **Equation (7)** can be written as follows:

$$Y_t = \Phi_1 Y_{t-1} + a_t \tag{7}$$

2.5 Model Identification and Specification

In the VARI model there are assumptions that must be met, namely that the time series data used must be stationary, stationary data does not have seasonal patterns, trends and constant variance. To overcome stationarity can use the Augmented Dickey Fuller (ADF) test with $\gamma = 0$, the test is formulated in **Equation (8) [18]**:

$$\Delta Z_t = \alpha + \beta t + \gamma Z_{t-1} + \delta_1 \Delta Z_{t-1} + \delta_2 \Delta Z_{t-2} + \cdots + \delta_p \Delta Z_{t-p} \tag{8}$$

where Z_t is time series data, which can be determined by linear regression of ΔZ_t to t and Z_{t-1} by testing $\gamma \neq 0$, with Δ denoting the first-difference operator ($\Delta Z_t = Z_t - Z_{t-1}$). If it does not have a unit root, it can be said to be stationary **[19]**. Then the error rate (α) is the critical value, with test statistics in **Equation (9)**:

$$\varsigma_{test} = \frac{\hat{\phi} - 1}{SE(\hat{\phi})} \tag{9}$$

where $SE(\hat{\phi})$ denotes as estimator of standard deviation of the OLS estimator $\hat{\phi}$. If the p-value obtained is less than 0.05 then H_0 is rejected. If the data used is not stationary in variance, it can use the Box-Cox transformation test, where λ is a power parameter that determines the type of transformation applied to the data, that has a variance from -5 to 5, whose optimal value will produce the best approach to a normal distribution curve. The transformation of y is given by **Equation (10)**:

$$y(\lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & if\ \lambda \neq 0 \\ \ln y\ , & if\ \lambda = 0 \end{cases} \tag{10}$$

The level of accuracy calculated from the lag value that produces the $p - value$ is done before forming the VARI model. To assess the feasibility of the model this time using the Akaike Information Criterion (AIC), Schwarz Criterion (SC), Hannan-Quinn Criterion (HQ), and Final Prediction Error (FPE) calculation for several independent variables p, as shown in **Equation (11)** to **Equation (14) [20]**:

$$AIC(p) = \ln|\hat{\varsigma}(p)| + \frac{2pk^2}{T} \tag{11}$$

$$SC(p) = \ln|\hat{\varsigma}(p)| + \frac{pk^2\ln(T)}{T} \tag{12}$$

$$HQ(p) = \ln|\hat{\varsigma}(p)| + \frac{2pk^2\ln(\ln(T))}{T} \tag{13}$$

$$FPE(p) = |\hat{\varsigma}(p)| \times \left(\frac{T + pk + 1}{T - pk - 1}\right)^k \tag{14}$$

where $\hat{\varsigma}(p) = \frac{\sum_{t=1}^{T}\hat{\varepsilon}_t(\hat{\varepsilon}_t)}{T}$ is the residual covariance matrix, T is the number of observations, $\hat{\varepsilon}_t$ is the error value, k is the number of parameters in the model, and p is the number of observations at time t.

2.6 Parameter Estimation

Parameter estimation for the VARI($\mathbf{1}$) model can be performed using the Ordinary Least Squares (OLS) method with the vectorized formulation of the system given in **Equation (15) [16]**. To obtain parameter estimates, the sum of squared errors function is determined by transforming the matrix into vector form. Thus, the parameter estimation for VARI($\boldsymbol{p}, \boldsymbol{d}$) with 2 variables is as follows **[21]**:

$$y = vec(Y) = \begin{bmatrix} Y_{1,1} \\ Y_{1,2} \\ \vdots \\ Y_{1,n} \\ Y_{2,1} \\ Y_{2,2} \\ \vdots \\ Y_{2,n} \end{bmatrix}, \phi = \begin{bmatrix} \phi_{10} \\ \phi_{11} \\ \phi_{12} \\ \phi_{20} \\ \phi_{21} \\ \phi_{22} \end{bmatrix}, a = \begin{bmatrix} a_{1,1} \\ a_{1,2} \\ \vdots \\ a_{1,n} \\ a_{2,1} \\ a_{2,2} \\ \vdots \\ a_{2,n} \end{bmatrix}, W = \begin{bmatrix} 1 & Y_{1,1-1} & Y_{2,1-1} \\ 1 & Y_{1,2-1} & Y_{2,2-1} \\ \vdots & \vdots & \vdots \\ 1 & Y_{1,n-1} & Y_{2,n-1} \end{bmatrix} \tag{15}$$

The expanded matrix form of $\boldsymbol{v} = \mathbf{I} \otimes \mathbf{W} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \otimes \begin{bmatrix} \mathbf{1} & \boldsymbol{Y}_{1,1-1} & \boldsymbol{Y}_{2,1-1} \\ \mathbf{1} & \boldsymbol{Y}_{1,2-1} & \boldsymbol{Y}_{2,2-1} \\ \vdots & \vdots & \vdots \\ \mathbf{1} & \boldsymbol{Y}_{1,n-1} & \boldsymbol{Y}_{2,n-1} \end{bmatrix}$, which facilitates the vectorization process, is illustrated in **Equation (16)**

$$v = \begin{bmatrix} 1 & Y_{1,1-1} & Y_{2,1-1} & 0 & 0 & 0 \\ 1 & Y_{1,2-1} & Y_{2,2-1} & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & Y_{1,n-1} & Y_{2,n-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & Y_{1,1-1} & Y_{2,1-1} \\ 0 & 0 & 0 & 1 & Y_{1,2-1} & Y_{2,2-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 & Y_{1,n-1} & Y_{2,n-1} \end{bmatrix} \tag{16}$$

Thus, **Equation (7)** can be rewritten as **Equation (17)**:

$$\begin{gathered} Vec(Y) = Vec(\phi W) + Vec(a) \\ y = v\phi + a \end{gathered} \tag{17}$$

The sum of squared errors is represented in **Equation (18)**:

$$\begin{aligned} \boldsymbol{S} &= \sum_{i=1}^{2}\sum_{j=1}^{n} \boldsymbol{a}_{ij}^{2} \\ &= \boldsymbol{a}^{T}\boldsymbol{a} \\ &= (\boldsymbol{y} - \boldsymbol{v\phi})^{T}(\boldsymbol{y} - \boldsymbol{v\phi}) \\ &= \boldsymbol{y}^{T}\boldsymbol{y} - \boldsymbol{y}^{T}\boldsymbol{v\phi} - \boldsymbol{y}\boldsymbol{v}^{T}\boldsymbol{\phi}^{T} + \boldsymbol{v}^{T}\boldsymbol{\phi}^{T}\boldsymbol{v\phi} \\ &= \boldsymbol{y}^{T}\boldsymbol{y} - 2\boldsymbol{y}^{T}\boldsymbol{v\phi} + \boldsymbol{v}^{T}\boldsymbol{\phi}^{T}\boldsymbol{v\phi} \end{aligned} \tag{18}$$

Then, we take the partial derivative with respect to the parameter $\boldsymbol{\phi}$:

$$\begin{aligned} \frac{\partial S}{\partial \phi} &= \frac{\partial(y^{T}y - 2y^{T}v\phi + v^{T}\phi^{T}v\phi)}{\partial \phi} \\ &= -2y^{T}v + 2\phi^{T}v^{T}v \\ 2\phi^{T}v^{T}v &= 2y^{T}v \\ \phi^{T}v^{T}v(v^{T}v)^{-1} &= y^{T}v(v^{T}v)^{-1} \\ \phi^{T}I &= y^{T}v(v^{T}v)^{-1} \\ (\phi^{T}I)^{T} &= (y^{T}v(v^{T}v)^{-1})^{T} \\ \phi &= (v^{T}v)^{-1}v^{T}y \end{aligned} \tag{19}$$

Equation (19) represents the OLS parameter estimate. If the second derivative is positive, then the second derivative of **Equation (7)** with respect to $\boldsymbol{\phi}^{T}$ is as follows:

$$\begin{aligned} \frac{\boldsymbol{\partial^{2} S}}{\boldsymbol{\partial\phi\partial\phi^{T}}} &= \frac{\boldsymbol{\partial}\left(-2\boldsymbol{y}^{T}\boldsymbol{v} + 2\boldsymbol{\phi}^{T}\boldsymbol{v}^{T}\boldsymbol{v}\right)}{\boldsymbol{\partial\phi^{T}}} \\ &= 2\boldsymbol{v}^{T}\boldsymbol{v} \end{aligned} \tag{20}$$

Since the second derivative of the sum of squared errors with respect to $\boldsymbol{\phi}^{T}$**,** as given in **Equation (20)** is a positive number, we obtain the OLS parameter estimate provided in **Equation (19)**.

2.7 Diagnostic Test

A normality test is conducted to assess the normality of the residual data. The test used is the Kolmogorov-Smirnov test is defined in **Equation (21)**, with the following hypotheses:

H_0 : $a_t \sim N(0, \sigma_{at}^2)$ (the residuals are normally distributed)
H_1: $a_t \neq N(0, \sigma_{at}^2)$ (the residuals are not normally distributed)

with statistics test:

$$D = \max|F_0(a_t) - S(a_t)| \tag{21}$$

where $F_0(a_t)$ is the cumulative distribution function of the normal distribution, and $S(a_t)$ is the cumulative distribution function of the data. If the value of $D_{calculated} > D_{table}$ or $p - value < \alpha$, then H_0 is rejected, indicating that the residuals are not normally distributed.

Meanwhile, the Ljung-Box test statistic as formulated in **Equation (22)** is a diagnostic method that uses all residual ACF samples as a whole to test the null hypothesis. The hypotheses used are as follows:

H_0: $\rho_1 = \rho_2 = \cdots = \rho_k = 0$ (there is no correlation among the residuals)
H_1: $\rho_k \neq 0; k = 1,2,3,..,n$ (there is correlation among the residuals)

with statistics test:

$$Q = n(n+2) \sum_{k=1}^{m} (n-k)^{-1} \hat{\rho}_k^2 \tag{22}$$

where n is the number of residuals, k is the lag, and $\hat{\rho}_k$ is the autocorrelation between residuals. If $p - value > \alpha$, then H_0 is not rejected, indicating that the residuals in the multivariate model satisfy the white noise assumption.

The causality test is used to determine the causal relationship between variables in the Vector Autoregressive Integrated (VARI) model. Assuming there are two variables, X and Y, X is considered to Granger-cause y if past values of X provide information that improves the prediction of Y **[22]**. The Granger causality test is formulated as shown in **Equation (23)** and **Equation (24)**:

$$Y_t = \sum_{i=1}^{n} \alpha_i Y_{i-1} + \sum_{i=1}^{n} \beta_i X_{i-1} + a_{1t} \tag{23}$$

$$X_t = \sum_{i=1}^{n} \gamma_i X_{i=1} + \sum_{i=1}^{n} \tau_i Y_{i-1} + a_{2t} \tag{24}$$

where X_t and Y_t represent the current values of the variables, $\alpha_i, \beta_i, \gamma_i, \tau_i$ are model parameters to be estimated, while a_{1t} and a_{2t} are white noise errors.

The causality between variables is tested using the F-test statistic, formulated as follows **Equation (25) [23]**:

$$F_{calculated} = \frac{(RSS_R - RSS_{UR})/p}{RSS_{UR}/(n-b)} \tag{25}$$

where RSS is residual sum of squares, p represent the lag length, n represents the number of observations, and b represents the number of parameters estimated in the unrestricted model. Suppose Y is the dependent variable, the restricted model is then obtained by regressing Y on all its lagged values, excluding the lagged values of X as independent variables. The form of the restricted model is given in **Equation (26)**:

$$Y_t = \sum_{i=1}^{n} \alpha_i Y_{i-1} + e_{1t} \tag{26}$$

There are two hypotheses used when conducting the Granger causality test, as follows:

H_0: X is not the cause of Granger Y.
H_1: X is the cause of Granger Y.
H_0: Y is not the cause of Granger X.
H_1: Y is the cause of Granger X.

If the calculated F value $> F_{(\mathrm{p,n-b})}$ or if the $p - value < \alpha$, then H_0 is rejected, indicating that the conditions for causality are met, and the best VARI model can be selected.

2.8 Forecasting Criteria

One of the forecasting criteria that can be used to assess the accuracy of a model in predicting actual events is the Mean Absolute Percentage Error (MAPE). The lower the MAPE value, the closer the forecasting results are to the actual outcomes. MAPE is defined as follows in **Equation (27) [24]**:

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{Z_{i,t} - \hat{Z}_{i,t}}{Z_{i,t}} \right| \times 100\% \tag{27}$$

with $Z_{i,t}$ is actual value at time t location i, $\hat{Z}_{i,t}$ is forecasted value at time t location i and n is number of time series observations.

The scale for evaluating the accuracy of the forecast based on the MAPE value is as follows **[25]**:

Table 2. Forecast Accuracy Criteria Based on MAPE Value

MAPE	Description
$< 10\%$	Highly accurate forecast
$10\% - 20\%$	Good forecast
$20\% - 50\%$	Reasonable/acceptable forecast
$> 50\%$	Inaccurate forecast

2.9 Research Flow

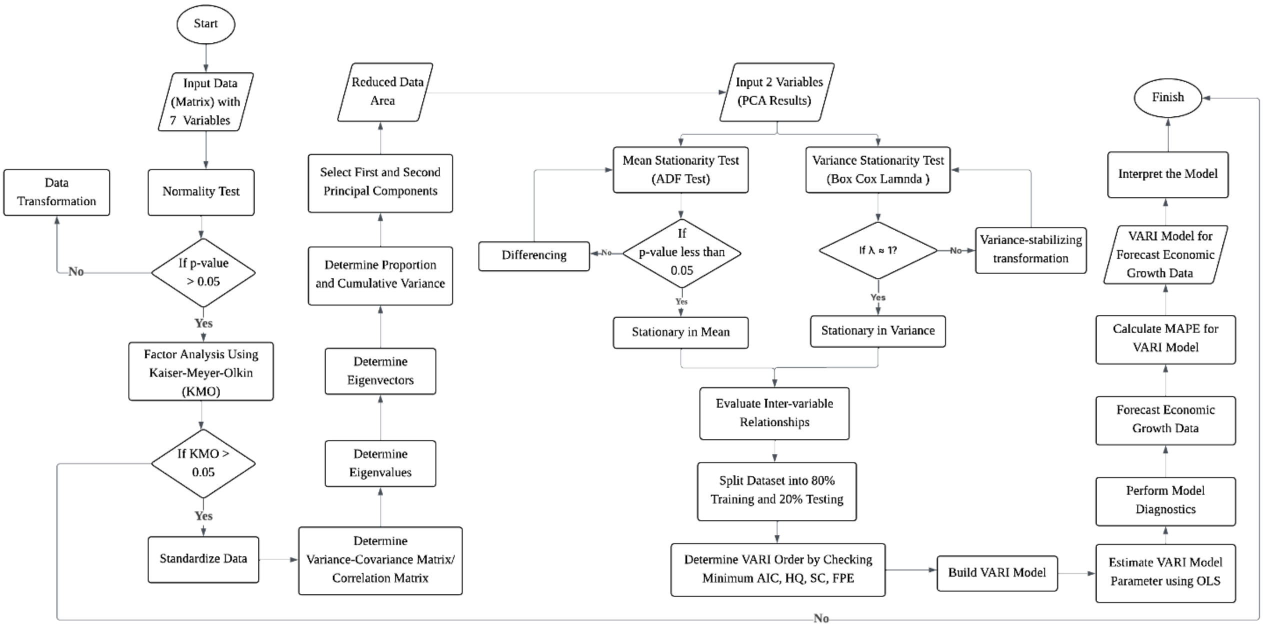

Figure 1. **PCA-VARI Flowchart**

The analysis steps conducted in this study are comprehensively outlined and visually represented in **Figure 1**. This flowchart provides a clear and systematic depiction of the methodological process, guiding through each phase of the analysis, from data preparation to the final interpretation of results. The data analysis was carried out using R Studio, which facilitated the implementation of the Principal Component Analysis (PCA) and Vector Autoregressive Integrated (VARI) model for economic growth forecasting in Indonesia."

3. RESULTS AND DISCUSSION

3.1 Descriptive Statistics

Table 3 presents the descriptive statistics of the research data. The research data has been transformed using logarithms due to the left-skewed tendency in the data distribution. After the transformation, the data shows a tendency toward a normal distribution, as indicated by the probability values obtained from the Jarque-Bera statistical test.

Table 3. **Descriptive Statistics**

Variable	Mean	Median	Max.	Min.	Std. Dev.	Skew.	Kurt.	Jarque-Bera	Prob.
X_1	5,243,964	5,198,863	8,528,022	2,849,796	1,503,460	0.296	-0.921	6.3364	0.0420
$\ln X_1$	15.43	14.46	15.96	14.86	0.29	-0.137	-0.998	5.5798	0.0614
X_2	5.489	5.62	7.75	3.50	1.39	0.144	-1.236	8.5161	0.0141
$\ln X_2$	1.670	1.727	2.048	1.253	0.26	-0.160	-1.189	8.0033	0.0182
X_3	99.82	99.98	101.73	95.38	1.93	-1.364	2.835	88.968	<0.001
$\ln X_3$	4.603	4.605	4.622	4.558	0.01	-1.414	2.994	97.417	<0.001
X_4	13,071	13,542	15,805	9,014	1,758.64	-0.990	-0.130	22.109	<0.001
$\ln X_4$	9.468	9.514	9.668	9.107	0.14	-1.176	0.193	31.465	<0.001
X_5	4.129	3.59	8.79	1.32	1.93	0.689	-0.305	11.071	0.0039
$\ln X_5$	1.3055	1.2781	2.1736	0.2776	0.48	-0.276	-0.512	2.9609	0.2275
X_6	119,633	119,076	146,870	92,671	13,304.9	0.096	-0.992	5.298	0.0707
$\ln X_6$	11.69	11.69	11.90	11.44	0.11	-0.076	-0.952	4.795	0.0909

Variable	Mean	Median	Max.	Min.	Std. Dev.	Skew.	Kurt.	Jarque-Bera	Prob.
X_7	171,006	153,322	410,267	71,033	73,407.5	1.188	0.744	35.353	<0.001
$\ln X_7$	11.97	11.94	12.92	11.17	0.39	0.456	-0.536	6.0646	0.0482

In the description of the research data, it is observed that the average amount of money supply (X_1) is 5,243,964 billion rupiah. The lowest amount of X_1 was recorded in February 2012 at 2,849,796 billion rupiah, while the highest amount was reached in December 2022, with 8,528,022 billion rupiah. The time series data for X_1 shows an upward trend, while other data exhibit fluctuating trends. For instance, the time series data for the interest rate (X_2) shows stability during several observation periods because the interest rate is set by Bank Indonesia as a monetary policy instrument. The peak X_2 occurred at 7.75% from November 2014 to January 2015, while the lowest rate was recorded at 3.50% from February 2021 to July 2022. The economic activity (X_3) reached its highest peak at 101.73 in December 2019 to January 2020, while the lowest was at 95.38 in May 2020. Meanwhile, the highest exchange rate (X_4) was 15,805 rupiah in April 2020, while the lowest was 9,014 rupiah in February 2012. The inflation pattern shows a correlation that is nearly identical to the rupiah exchange rate. The inflation (X_5) occurred in August 2013 at 8.79%, while the lowest X_5 was recorded in August 2020 at 1.32%. The average foreign exchange reserves (X_6) amounted to 119,633 million USD. The highest X_6 were recorded in September 2021 at 146,870 million USD, while the lowest was 92,671 million USD in July 2013. On the other hand, the average stock transaction and index (X_7) was 171,006 billion rupiah. The highest X_7 occurred in January 2021 with a value of 410,267 billion rupiah, while the lowest was recorded in September 2012 at 71,033 billion rupiah.

3.2 Application of Principal Component Analysis (PCA) for Economic Growth Data

The initial data processing using PCA was conducted to evaluate the suitability of economic growth data in Indonesia. The results of this test showed a KMO value of 0.77, indicating that the data meets the suitability criteria, with a KMO value above 0.50. Therefore, the PCA modeling can proceed with the confidence that the data used is adequate for PCA.

The next step is to test the homogeneity of the variance-covariance matrix using Bartlett's Test. The test results show a p-value < 0.001, which is less than the significance level of 0.05. This indicates that the variance-covariance matrix is heterogeneous. Therefore, for PCA analysis, the correlation matrix is used.

Table 4. Eigenvalue Estimation

Component	Cumulative Proportion	Eigenvalue
1	0.6737321	4.71612452
2	0.8138922	0.98112069
3	0.90519392	0.63911223
4	0.95766555	0.36730142
5	0.98353285	0.18107106
6	0.99694072	0.09385515
7	1.00000000	0.02141493

Table 4 shows that after forming the correlation matrix on the data, the eigenvalues of the matrix were obtained. The First Component (PC_1) provides a cumulative proportion of 67.37%, while the Second Component (PC_2) provides a cumulative proportion of 81.38%. With a total cumulative proportion of over 80%, it can be concluded that PC_1 and PC_2 are sufficient to explain the entire dataset.

Table 5. **Eigenvector Estimation**

Variables	PC_1	PC_2
$\ln X_1$	0.4309224	-0.28695408
$\ln X_2$	-0.4037830	-0.19455204
$\ln X_3$	-0.2224136	-0.74159735
$\ln X_4$	0.3503654	-0.40620484
$\ln X_5$	-0.3766986	-0.32618424
$\ln X_6$	0.4303839	-0.01475317
$\ln X_7$	0.3897555	-0.24129658

The PCA analysis for economic growth data in Indonesia is presented in **Table 5**. Considering the calculated PC_1, it can be concluded that the highest correlation for the First Component (PC_1) is found in the money supply (X_1), interest rate (X_2), foreign exchange reserves (X_6) and stock transactions and index (X_7). These four variables are factors related to monetary policy and financial stability of a country. Considering the calculated PC_2, it can be concluded that the highest correlation for the Second Component (PC_2) is found in the economy activity (X_3), exchange rate (X_4), and inflation (X_5). These three variables are factors related to the economic conditions of a country.

Table 6. **Loadings Estimation**

Variables	PC_1	PC_2
$\ln X_1$	0.431	0.287
$\ln X_2$	-0.404	0.195
$\ln X_3$	-0.222	0.742
$\ln X_4$	0.350	0.406
$\ln X_5$	-0.377	0.326
$\ln X_6$	0.430	0.014
$\ln X_7$	0.390	0.241

Table 6 shows the loadings used to form the principal component formulas. Loadings are coefficients that indicate the relationship between the original variables and the principal components produced by PCA. Although the eigenvalue for foreign exchange reserves ($\ln X_6$) is present (-0.01), its corresponding loading is near zero. This suggests that while ($\ln X_6$) explains some variance, it has minimal impact on the principal components. The small loading likely indicates a weak contribution, possibly due to high correlation with other variables, which reduces its influence in the PCA results. **Equation (28)** represents a linear combination between the first principal component and other variables as follows:

$$PC_1 = 0.431\, ln\, X_1 - 0.404\, ln\, X_2 - 0.222\, ln\, X_3 + 0.350\, ln\, X_4 - 0.377\, ln\, X_5 + 0.430\, ln\, X_6 + 0.390\, ln\, X_7 \tag{28}$$

Equation (29) represents a linear combination between the second principal component and other variables as follows:

$$PC_2 = 0.287\, ln\, X_1 + 0.195\, ln\, X_2 + 0.742\, ln\, X_3 + 0.406 \ln X_4 + 0.326 \ln X_5 + 0.241 \ln X_7 \tag{29}$$

3.3 Stationarity Testing and Model Selection in VARI Model

The Augmented Dickey-Fuller (ADF) test and the Box-Cox test were used in this study to examine the stationarity of the mean and variance of PC_1 and PC_2, represented as $Z_{1,t}$ and $Z_{2,t}$ respectively. Prior to any transformation, the ADF test results for both $Z_{1,t}$ and $Z_{2,t}$ indicated high p-values (0.5261 and 0.6956), which are greater than $\alpha(0.05)$. Therefore, the null hypothesis of a unit root could not be rejected, indicating that the data were non-stationary in mean. Moreover, the initial Box-Cox lambda (λ) values were found to be approximately –0.9999 for both components, suggesting that the data also lacked variance stationarity as recorded in **Table 7**.

Table 7. **Pre-transformation ADF and Box-Cox Tests**

$Z_{1,t}$			$Z_{2,t}$		
ADF	***p-value***	**Box-Cox**	**ADF**	***p-value***	**Box-Cox**
-3.9833	0.01228	0.9405164	-4.0431	0.01	0.9405164

To address these issues, a negative logarithmic transformation was applied to stabilize the variance, followed by first-order differencing to eliminate the unit root in the mean. After these procedures, the stationarity tests were performed again. The ADF test results show values $< \alpha(0.05)$ after $Z_{1,t}$ and $Z_{2,t}$ are differenced once. It is important to note that the critical values of the ADF and Box-Cox tests vary depending on the data assumptions, as recorded in **Table 8**. Thus, the validation tests support the estimated results. The ADF test results $< \alpha(0.05)$ and Box-Cox ≈ 1 indicate that the economic growth data is stationary in both mean and variance.

Table 8. **Post-transformation ADF and Box-Cox Tests**

$Z_{1,t}$			$Z_{2,t}$		
ADF	***p-value***	**Box-Cox**	**ADF**	***p-value***	**Box-Cox**
-3.9833	0.01228	0.9405164	-4.0431	0.01	0.9405164

The optimum lag was obtained through stationary variables in the PCA process by taking the first-order difference up to the fifth difference. Ultimately, the minimum third-order value of these variables became the input for the VARI model after differencing by 1. It should be noted that the third-order lag was selected as the optimal lag for the VARI model using the Akaike Information Criterion (AIC), Hannan-Quinn Criterion (HQ), Schwarz Criterion (SC), and Final Prediction Error (FPE).

Table 9. **Lag Order Selection Criteria in the VARI Model**

	VARI(1)	**VARI(2)**	**VARI(3)**	**VARI(4)**	**VARI(5)**
AIC(n)	-9.81363484	-10.0668909	**-10.1321340**	-10.0710232	-10.0338830
HQ(n)	-9.51841425	-9.72949602	**-9.75256467**	-9.64927959	-9.56996493
SC(n)	-9.08418719	**-9.23323652**	-9.19427273	-9.02895521	-8.88760811
FPE(n)	0.000054904	0.000042699	**0.000040097**	0.000042752	0.000044534

3.4 PCA-VARI Model

The next step is to estimate the parameters of the selected PCA-VARI(3) model. The estimation results are shown in **Table 10**. The asterisks indicate the significance of the estimated model coefficients. As shown in **Table 9** and **Table 10**, the lag order of the PCA-VARI model to be used is the third lag order, and the estimation is consistent with the predictions at the model significance level of $\alpha = 0.05$.

Table 10. **Estimation of Coefficients for the PCA-VARI(3) Model**

Variables	$\dot{\boldsymbol{Z}}_{1,t}$	$\dot{\boldsymbol{Z}}_{2,t}$
$\dot{Z}_{1,t-1}$	$\hat{\phi}^1_{1,1} = 0.92334$***	$\hat{\phi}^1_{2,1} = -0.33767$***
$\dot{Z}_{1,t-2}$	$\hat{\phi}^1_{1,2} = -0.16503$	$\hat{\phi}^1_{2,2} = 0.21156$*
$\dot{Z}_{1,t-3}$	$\hat{\phi}^1_{1,3} = 0.23515$	$\hat{\phi}^1_{2,3} = 0.11703$
$\dot{Z}_{2,t-1}$	$\hat{\phi}^2_{1,1} = -0.62780$ **	$\hat{\phi}^2_{2,1} = 0.76094$***
$\dot{Z}_{2,t-2}$	$\hat{\phi}^2_{1,2} = 0.49506$	$\hat{\phi}^2_{2,2} = -0.06810$
$\dot{Z}_{2,t-3}$	$\hat{\phi}^2_{1,3} = 0.08704$	$\hat{\phi}^2_{2,3} = 0.19179$*

Note: The asterisks ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.1 levels, respectively.

Based on **Table 10**, $\dot{Z}_{1,t}$ reflects the first principal component of the economic growth data in Indonesia within the PCA-VARI model. The interpretation of the model indicates that a one-fold increase in money supply (X_1), interest rate (X_2), foreign exchange reserves (X_6) and stock transactions and index (X_7) (representing the four strong correlations) in the previous month results in a 0.92334-fold increase in the factors affecting monetary policy and financial stability in the current month. Conversely, a one-fold decrease in the economy activity (X_3), exchange rate (X_4), and inflation (X_5) (representing the three strong correlations) in the previous month results in a 0.62780-fold increase in the factors affecting monetary policy and financial stability in the current month.

Notation $\dot{Z}_{2,t}$ serves as the input for the PCA-VARI model from the second principal component of Indonesia's economic growth data. The interpretation of the model suggests that a one-fold decrease in money supply (X_1), interest rate (X_2), foreign exchange reserves (X_6) and stock transactions and index (X_7) (representing the four strong correlations) in the previous month results in a 0.33767-fold decrease in the economy activity (X_3), exchange rate (X_4), and inflation (X_5) in the current month. Conversely, a one-fold increase in the economy activity (X_3), exchange rate (X_4), and inflation (X_5) in the previous month results in a 0.76094-fold increase in the economy activity (X_3), exchange rate (X_4), and inflation (X_5) in the current month.

The interpretation of this model indicates that the principal component variables influence each other through the second component variable. This suggests that in the VARI process, the performance observed in the two principal components demonstrates a strong correlation between the factors affecting monetary policy and financial stability (money supply, interest rate, foreign exchange reserves and stock transactions and index) and the factors influencing economic conditions (economy activity, exchange rate, and inflation). These two factors tend to be influenced by variables that mutually affect each other and are strongly linked to the predictions at a model significance level of $\alpha < 0.05$.

The prediction model for the first principal component of Indonesia's economic growth data in the PCA-VARI model is presented in **Equation (30)**, while the model for the second principal component is described in **Equation (31)**.

$$\Delta Z_{1,t} = 0.923\Delta Z_{1,t-1} - 0.628\Delta Z_{2,t-1} - 0.165\Delta Z_{1,t-2} + 0.495\Delta Z_{2,t-2} + 0.235\Delta Z_{1,t-3} + 0.087\Delta Z_{2,t-3} \tag{30}$$

$$\Delta Z_{2,t} = -0.338\Delta Z_{1,t-1} + 0.761\Delta Z_{2,t-1} + 0.212\Delta Z_{1,t-2} - 0.068\Delta Z_{2,t-2} + 0.117\Delta Z_{1,t-3} + 0.192\Delta Z_{2,t-3} \tag{31}$$

Table 11. Diagnostic Tests for Coefficient Estimates in the PCA-VARI(3) Model

Diagnostics Tests	Variables	F-Test	*p-value*
Granger	$Z_{1,t}$	16.4	0.000089
	$Z_{2,t}$	8.33	0.0046
Ljung-Box	$Z_{1,t}$		0.83
	$Z_{2,t}$		0.96
Kolmogorov-Smirnov	$Z_{1,t}$		0.71
	$Z_{2,t}$		0.29

Verification of the obtained model to ensure the significance of multivariate modeling was conducted by performing diagnostic tests to examine the frequency distribution of residuals using the Kolmogorov-Smirnov normality test and the Ljung-Box white noise test. Table 11 presents the diagnostic test results for the PCA-VARI(3) model.

In Table 11, it can be seen that in the Granger Causality test for $\boldsymbol{Z_{1,t}}$, the p-value is 0.000089 (<0.05), leading to the rejection of the null hypothesis, and it can be concluded that $\boldsymbol{Z_{1,t}}$ passes the model diagnostic test. Similarly, for $\boldsymbol{Z_{2,t}}$, the p-value is 0.0046 (<0.05), also leading to the rejection of the null hypothesis, and it can be concluded that $\boldsymbol{Z_{2,t}}$ passes the model diagnostic test. The white noise test using Ljung-Box yielded a $p - value$ close to 1, while the normality test using Kolmogorov-Smirnov resulted in a p-value > 0.05. Therefore, the obtained PCA-VARI(3) model passes the multivariate diagnostic tests.

3.5 Prediction using PCA-VARI Model

The next step is to predict the factors affecting monetary policy and financial stability (PC_1) and the factors affecting economic conditions (PC_2). Before making predictions, the data is split into training and testing datasets with a proportion of 80% and 20%, respectively. Afterward, predictions are made using the PCA-VARI(3) model, and the prediction results are shown in Figure 2.

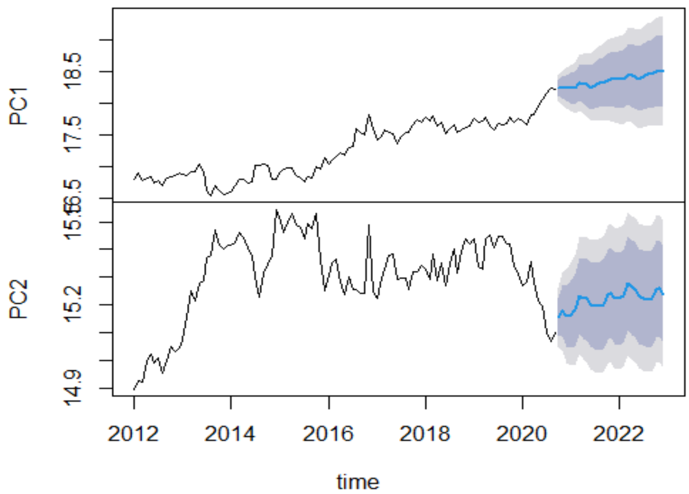

Figure 2. Plot of Predictions for PC_1 and PC_2

The MAPE value for PC_1 was 1.21%, while the MAPE value for PC_2 was 1.34%, both calculated based on the testing dataset. The MAPE value < 10% indicates that the prediction results for the PCA-VARI(3) model on Indonesia's economic growth data are highly accurate.

4. CONCLUSION

This study demonstrates that the Principal Component Analysis-Vector Autoregressive Integrated (PCA-VARI) model is effective in forecasting economic growth in Indonesia with high accuracy. Data from 2012-

2022 were used for the analysis, and PCA successfully identified two principal components sufficient to explain most of the data variability, with PC_1 related to monetary policy and financial stability (money supply, interest rates, foreign exchange reserves, and stock transactions) and PC_2 related to economic conditions (economic activity, exchange rate, and inflation). The VARI model with an optimal lag of the third order revealed significant interactions among the principal variables, with prediction results showing a Mean Absolute Percentage Error (MAPE) of 1.21% for PC_1 and 1.34% for PC_2. These results suggest that short-term economic fluctuations are closely linked to monetary and financial factors, while structural economic trends are influenced by broader economic conditions. Understanding these dynamics provides a foundation for designing policy interventions that are responsive to both short-term volatility and long-term development goals. Based on these findings, it is recommended that the government integrate predictive models such as PCA-VARI into macroeconomic planning. Strengthening the coordination between monetary and fiscal policies, maintaining exchange rate stability, controlling inflation, and encouraging domestic economic activity are key strategies to enhance economic resilience and support sustainable growth.

AUTHOR CONTRIBUTIONS

Aulia Rahma Al Madani: Conceptualization, Formal analysis, Methodology, Software, Writing - original draft, Writing - review and editing. Sandrina Najwa: Writing - review and editing, Validation, Writing - original draft, Visualization, Data curation. Budi Nurani Ruchjana: Supervision, Funding acquisition. All authors discussed the results and contributed to the final manuscript.

FUNDING STATEMENT

This research has been funded by Academic Leadership Grant with contract number: 1425/UN6.3.1/PT.00/2024.

ACKNOWLEDGMENT

The authors would like to thank the Rector of Padjadjaran University, Director of DRPM Padjadjaran University, Center for Modeling and Computation Studies FMIPA Padjadjaran University for providing research funds for the dissemination of research results of lecturers and students.

CONFLICT OF INTEREST

The authors declare no conflicts of interest to report study.

REFERENCES

[1] C. Seftarita, Ferayanti, Fitriyani, and A. Diana, "DO PANDEMIC COVID 19 AND BUSINESS CYCLE INFLUENCE THE INDONESIA COMPOSITE INDEX?," in *E3S Web of Conferences*, EDP Sciences, Jan. 2022.doi: https://doi.org/10.1051/e3sconf/202234005004

[2] Badan Pusat Statistik, BERITA RESMI STATISTIK: PERTUMBUHAN EKONOMI INDONESIA TRIWULAN IV-2022. Jakarta: Badan Pusat Statistik, 2022.

[3] R. Pardede and S. Zahro, "SAVING NOT SPENDING: INDONESIA'S DOMESTIC DEMAND PROBLEM," *Bull Indones Econ Stud*, vol. 53, no. 3, pp. 233–259, Sep. 2017.doi: https://doi.org/10.1080/00074918.2017.1434928

[4] OECD, *OECD Economic Outlook*. Paris: OECD Publishing, 2016.

[5] Asian Development Bank, *Meeting Asia's Infrastructure Needs.* Philippines: Asian Development Bank, 2017.

[6] J. Mota, A. Moreira, and A. Alves, "Impact of export promotion programs on export performance," *Economies*, vol. 9, no. 3, Sep. 2021.doi: https://doi.org/10.3390/economies9030127

[7] Y. Xie, T. Wang, J. Kim, K. Lee, and M. K. Jeong, "LEAST ANGLE SPARSE PRINCIPAL COMPONENT ANALYSIS FOR ULTRAHIGH DIMENSIONAL DATA," *Ann Oper Res*, 2024.doi: https://doi.org/10.1007/s10479-024-06428-0

[8] S. Wulandary, "VECTOR AUTOREGRESSIVE INTEGRATED (VARI) METHOD FOR FORECASTING THE NUMBER OF INTERNASIONAL VISITOR IN BATAM AND JAKARTA," *Jurnal Matematika, Statistika & Komputasi*, vol. 17, no. 1, pp. 94–108, 2020.doi: https://doi.org/10.20956/jmsk.v17i1.10536
[9] F. Tan and Z. Lu, "STUDY ON THE INTERACTION AND RELATION OF SOCIETY, ECONOMY AND ENVIRONMENT BASED ON PCA-VAR MODEL: AS A CASE STUDY OF THE BOHAI RIM REGION, CHINA," *Ecol Indic*, vol. 48, pp. 31–40, 2015.doi: https://doi.org/10.1016/j.ecolind.2014.07.036
[10] S. Mamipour, M. Yahoo, and S. Jalalvandi, "AN EMPIRICAL ANALYSIS OF THE RELATIONSHIP BETWEEN THE ENVIRONMENT, ECONOMY, AND SOCIETY: RESULTS OF A PCA-VAR MODEL FOR IRAN," *Ecol Indic*, vol. 102, pp. 760–769, Jul. 2019.doi: https://doi.org/10.1016/j.ecolind.2019.03.039
[11] D. Munandar, B. N. Ruchjana, and A. S. Abdullah, "PRINCIPAL COMPONENT ANALYSIS-VECTOR AUTOREGRESSIVE INTEGRATED (PCA-VARI) MODEL USING DATA MINING APPROACH TO CLIMATE DATA IN THE WEST JAVA REGION," BAREKENG: Jurnal Ilmu Matematika dan Terapan, vol. 16, no. 1, pp. 099–112, Mar. 2022.doi: https://doi.org/10.30598/barekengvol16iss1pp099-112
[12] R. A. Johnson and D. W. Wichern, *APPLIED MULTIVARIATE STATISTICAL ANALYSIS*, Sixth Edition. Upper Saddle River: Pearson Prentice Hall, 2007.
[13] I. T. Jolliffe, *Principal Component Analysis*, vol. Second Edition. 2002.
[14] N. A. I. A. Latif, I. M. Z. Abidin, N. Azaman, N. Jamaludin, and A. A. Mokhtar, "A FEATURE EXTRACTION TECHNIQUE BASED ON FACTOR ANALYSIS FOR PULSED EDDY CURRENT DEFECTS CATEGORIZATION," in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, 2019.
[15] D. N. Gujarati, *Basic Econometrics*, Fourth Edition. New York: McGraw-Hill, 2003.
[16] W. W. S. Wei, *Multivariate Time Series Analysis and Applications*. Upper Saddle River: Pearson Prentice Hall, 2019.
[17] Z. Putri, A. Yuniar, T. Toharudin, and B. Ruchjana, "MODEL VECTOR AUTOREGRESSIVE INTEGRATED (VARI) DAN PENERAPANNYA PADA DATA PERKEMBANGAN HARGA ECERAN BERAS DI TIGA IBU KOTA PROVINSI WILAYAH PULAU JAWA," *Pattimura Proceeding: Conference of Science and Technology*, vol. 2, no. 1, pp. 533–544, May 2022.doi: https://doi.org/10.30598/PattimuraSci.2021.KNMXX.533-544
[18] E. Paparoditis and D. N. Politis, "THE ASYMPTOTIC SIZE AND POWER OF THE AUGMENTED DICKEY–FULLER TEST FOR A UNIT ROOT," *Econom Rev*, vol. 37, no. 9, pp. 955–973, Oct. 2018.doi: https://doi.org/10.1080/00927872.2016.1178887
[19] Z. Guo, "RESEARCH ON THE AUGMENTED DICKEY-FULLER TEST FOR PREDICTING STOCK PRICES AND RETURNS," *Advances in Economics, Management and Political Sciences*, vol. 44, no. 1, pp. 101–106, Nov. 2023.doi: https://doi.org/10.54254/2754-1169/44/20232198
[20] F. Bashir and H. L. Wei, "HANDLING MISSING DATA IN MULTIVARIATE TIME SERIES USING A VECTOR AUTOREGRESSIVE MODEL-IMPUTATION (VAR-IM) ALGORITHM," *Neurocomputing*, vol. 276, pp. 23–30, Feb. 2018.doi: https://doi.org/10.1016/j.neucom.2017.03.097
[21] L L. Fitria Dewi, Y. Suparman, and I. Gede Nyoman Mindra Jaya, "IMPLEMENTATION OF THE AUTOREGRESSIVE INTEGRATED (VARI) VECTOR MODEL FOR STRESS-TESTING ANALYSIS OF INDONESIAN BANKING," *Journal of Social Research*, 2023.
[22] L. Barnett and A. K. Seth, "THE MVGC MULTIVARIATE GRANGER CAUSALITY TOOLBOX: A NEW APPROACH TO GRANGER-CAUSAL INFERENCE," *J Neurosci Methods*, vol. 223, pp. 50–68, Feb. 2014.doi: https://doi.org/10.1016/j.jneumeth.2013.10.018
[23] A. Shojaie and E. B. Fox, "GRANGER CAUSALITY: A REVIEW AND RECENT ADVANCES," *Annu Rev Stat Appl*, vol. 9, pp. 289–319, 2022.doi: https://doi.org/10.1146/annurev-statistics-040120-010930
[24] L. Zhou, P. Zhao, D. Wu, C. Cheng, and H. Huang, "TIME SERIES MODEL FOR FORECASTING THE NUMBER OF NEW ADMISSION INPATIENTS," *BMC Med Inform Decis Mak*, vol. 18, no. 1, Jun. 2018.doi: https://doi.org/10.1186/s12911-018-0616-8
[25] E. Vivas, H. Allende-Cid, and R. Salas, "A SYSTEMATIC REVIEW OF STATISTICAL AND MACHINE LEARNING METHODS FOR ELECTRICAL POWER FORECASTING WITH REPORTED MAPE SCORE," *Entropy*, vol. 22, no. 12, pp. 1–24, Dec. 2020.doi: https://doi.org/10.3390/e22121412