THE USE OF PCA TO OVERCOME THE MULTICOLLINEARITY IN ARIMAX MODEL FOR FORECASTING THE RUPIAH-USD EXCHANGE RATE

(Thesis)

By

NAJLA KHANSA ALIFAH
2217031011

DEPARTMENT OF MATHEMATICS
FACULTY OF MATHEMATICS AND NATURAL SCIENCES
UNIVERSITY OF LAMPUNG
BANDAR LAMPUNG
2026

ABSTRACK

THE USE OF PCA TO OVERCOME THE MULTICOLLINEARITY IN ARIMAX MODEL FOR FORECASTING THE RUPIAH-USD EXCHANGE RATE

By

NAJLA KHANSA ALIFAH

Time series forecasting is used to estimate the future value of a variable based on its historical patterns. In practice, a variable is not only influenced by its past values but also by external factors, leading to the use of the Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) model. However, the inclusion of multiple exogenous variables may lead to multicollinearity, which can reduce the reliability of the model. This study aims to apply Principal Component Analysis (PCA) to address multicollinearity in the ARIMAX model for forecasting the Rupiah exchange rate against the United States Dollar for the period January to December 2025. The data used are monthly secondary data from January 2012 to December 2024, with exogenous variables including money supply, interest rate, inflation, exports, and imports. The results show that PCA produces two principal components that explain 85.55% of the total data variability. The best model obtained is ARIMAX(0,1,2) with natural logarithm transformation. The forecasting results indicate that the Rupiah exchange rate fluctuates but tends to depreciate, with forecasting errors in each period ranging from 0.15546% to 3.716329% and a Mean Absolute Percentage Error (MAPE) of 1.263363%, indicating that the model has a high level of accuracy.

Keywords: ARIMAX, Rupiah Exchange Rate, Forecasting, Multicollinearity, Principal Component Analysis.

ABSTRAK

PENGGUNAAN PCA UNTUK MENGATASI MULTIKOLINEARITAS DALAM MODEL ARIMAX UNTUK PERAMALAN NILAI TUKAR RUPIAH-USD

Oleh

NAJLA KHANSA ALIFAH

Peramalan deret waktu digunakan untuk memperkirakan nilai suatu variabel di masa mendatang berdasarkan pola historisnya. Dalam praktiknya, variabel tidak hanya dipengaruhi oleh nilai masa lalu, tetapi juga oleh faktor eksternal, sehingga digunakan model *Autoregressive Integrated Moving Average with Exogenous Variables* (ARIMAX). Namun, penggunaan beberapa variabel eksogen berpotensi menimbulkan multikolinearitas yang dapat menurunkan keandalan model. Penelitian ini bertujuan menerapkan *Principal Component Analysis* (PCA) untuk mengatasi multikolinearitas pada model ARIMAX dalam meramalkan nilai tukar rupiah terhadap dolar Amerika Serikat periode Januari hingga Desember 2025. Data yang digunakan merupakan data sekunder bulanan periode Januari 2012 hingga Desember 2024, dengan variabel eksogen berupa jumlah uang beredar, suku bunga (BI Rate), inflasi, ekspor, dan impor. Hasil penelitian menunjukkan bahwa PCA menghasilkan dua komponen utama yang mampu menjelaskan 85.55% keragaman data. Model terbaik adalah ARIMAX(0,1,2) dengan transformasi logaritma natural. Hasil peramalan menunjukkan bahwa nilai tukar rupiah berfluktuasi namun cenderung mengalami depresiasi, dengan tingkat kesalahan pada setiap periode berkisar antara 0.15546% hingga 3.716329% dan nilai *Mean Absolute Percentage Error* (MAPE) sebesar 1.263363%, sehingga model memiliki tingkat akurasi yang tinggi.

Keywords: ARIMAX, Multikolinearitas, Nilai Tukar Rupiah, Peramalan, *Principal Component Analysis*.

THE USE OF PCA TO OVERCOME THE MULTICOLLINEARITY IN ARIMAX MODEL FOR FORECASTING THE RUPIAH-USD EXCHANGE RATE

By

NAJLA KHANSA ALIFAH

Thesis

Submitted in Partial Fulfillment of the Requirements for the Degree of BACHELOR OF MATHEMATICS

Submitted to

Department of Mathematics
Faculty of Mathematics and Natural Sciences

DEPARTMENT OF MATHEMATICS
FACULTY OF MATHEMATICS AND NATURAL SCIENCES
UNIVERSITY OF LAMPUNG
BANDAR LAMPUNG
2026

Thesis Title	:	**THE USE OF PCA TO OVERCOME THE MULTICOLLINEARITY IN ARIMAX MODEL FOR FORECASTING THE RUPIAH-USD EXCHANGE RATE**
Student Name	:	*Najla Khansa Alifah*
Student ID Number	:	**2217031011**
Study Program	:	**Mathematics**
Faculty	:	**Mathematics and Natural Sciences**

APPROVED BY

1. Advisory Committee

Dr. Khoirin Nisa, S.Si., M.Si.
NIP. 197407262000032001

Misgiyati, S. Pd., M. Si.
NIP. 198509282023212032

2. Head of the Department of Mathematics

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 19740316 200501 1 001

VALIDATED BY

1. Examination Committee

Head	: **Dr. Khoirin Nisa, S.Si., M.Si.**	....
Secretary	: **Misgiyati, S. Pd., M. Si.**	....
Examiner Non-Advisor	: **Drs. Nusyirwan, M.Si.**	....

2. Dean of the Faculty of Mathematics and Natural Sciences

Dr. Eng. Heri Satria, S.Si., M.Si.
NIP. 19711001 200501 1 002

Date of Thesis Examination: **May 26, 2026**

STUDENT'S THESIS DECLARATION

The undersigned below:

Name	:	**Najla Khansa Alifah**
Student ID Number	:	**2217031011**
Study Program	:	**Mathematics**
Thesis Title	:	**The Use of PCA to Overcome the Multicollinearity in ARIMAX Model for Forecasting the Rupiah-USD Exchange Rate**

Hereby declares that this thesis is the result of my own work. If, at any time in the future, this thesis is proven to be plagiarized or written by another person, I am willing to accept sanctions in accordance with the applicable academic regulations.

Bandar Lampung, May 26, 2026

Author

SEPULUH RIBU RUPIAH 10000 METERAI TEMPEL B1467AOX038183437

Najla Khansa Alifah

NPM 2217031011

AUTHOR BIOGRAPHY

Najla Khansa Alifah was born in Bandar Lampung on June 14, 2004, and is the first daughter of two siblings.

The author completed kindergarten education at Lazuardi Haura Global Islamic School and graduated in 2010. Elementary education was completed at SDIT Permata Bunda II and graduated in 2016. Furthermore, the author continued junior high school education at SMPIT Daarul ‘Ilmi and graduated in 2019, then continued senior high school education at SMAN 2 Bandar Lampung and graduated in 2022.

In 2022, the author was accepted as a student in the Mathematics Study Program, Faculty of Mathematics and Natural Sciences, University of Lampung through the SNMPTN admission pathway. During her studies, the author actively participated in several committee activities, such as the Mathematics Department Anniversary (DINAMIKA) and MIPA Expo. In addition, the author carried out an Internship at Badan Perencanaan Pembangunan, Riset dan Inovasi Daerah (BAPPERIDA) of Bandar Lampung City in the Division of Regional Development Planning, Control, and Evaluation, and also conducted Community Service Program (KKN) in Kota Karang Subdistrict, Teluk Betung Timur District, Bandar Lampung.

WORDS OF INSPIRATION

"For indeed, with hardship comes ease. Indeed, with hardship comes ease."

(QS. Al-Insyirah: 5-6)

"It always seems impossible until it's done."

(Nelson Mandela)

"When life gives you lemons, make lemonade."

DEDICATION

All praise and gratitude be to Allah SWT for His mercy, blessings, and grace so that the author could complete this thesis. Every process until this stage could not be separated from the help, ease, and strength given by Allah SWT.
With sincere gratitude, this thesis is dedicated to:

Beloved Father, Mother, and Younger Brother

Thank you for all the love, prayers, support, care, and encouragement given to the author throughout this journey. Thank you for always being a source of strength and motivation so that the author could complete this thesis and reach this stage.

Supervisors and Examiner

Thank you for the knowledge, guidance, advice, and support provided during the process of completing this thesis.

Beloved Friends

Thank you for the support, care, assistance, and togetherness shared with the author throughout this journey.

Beloved Alma Mater, University of Lampung

ACKNOWLEDGEMENTS

All praise and gratitude be to Allah SWT for His mercy, grace, and ease so that the author could complete the thesis entitled “The Use of PCA to Overcome the Multicollinearity in ARIMAX Model for Forecasting the Rupiah-USD Exchange Rate” as one of the requirements for obtaining a Bachelor’s degree in Mathematics at the Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Lampung.

During the preparation of this thesis, many parties provided assistance, support, guidance, and prayers. Therefore, the author would like to express sincere gratitude to:

1. Mrs. Dr. Khoirin Nisa, S.Si., M.Si., as the first supervisor who has taken the time to provide knowledge, direction, suggestions, motivation, and guidance to the author with great patience.
2. Mrs. Misgiyati, S.Pd., M.Si., as the second supervisor who has provided guidance, attention, suggestions, and support to the author.
3. Mrs. Dra. Dorrah Azis, M.Si., as the previous second supervisor who has provided direction and suggestions to the author.
4. Mr. Drs. Nusyirwan, M.Si., as the examiner who has provided suggestions and input for the improvement of this thesis.
5. Mr. Prof. Dr. Mustofa, M.A., Ph.D., as the academic advisor who has provided direction and guidance to the author during the academic years.
6. Mr. Dr. Aang Nuryaman, S.Si., M.Si., as the Head of the Mathematics Department, Faculty of Mathematics and Natural Sciences, University of Lampung.

7. Mr. Dr. Eng. Heri Satria, S.Si., M.Si., as the Dean of the Faculty of Mathematics and Natural Sciences, University of Lampung.
8. All Lecturers and Staff of the Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Lampung, who have provided knowledge, assistance, and valuable experiences to the author during the study period.
9. The author's beloved parents, Father and Mother, who always provided prayers, love, support, attention, and endless encouragement to the author in every process that was undertaken.
10. The author's beloved younger brother, who always provided attention, support, and encouragement to the author.
11. Beloved friends who always provided support, assistance, encouragement, and togetherness to the author throughout this journey.
12. All parties who cannot be mentioned one by one who have helped and provided support to the author in completing this thesis.
13. Lastly, thank you to myself for being able to endure, strive, and complete the entire academic process and the writing of this thesis until the end. Although the process was not always easy and there are still many shortcomings, thank you for always trying your best to reach this stage.

The author realizes that this thesis still has shortcomings. Therefore, the author expects constructive criticism and suggestions for future improvement. Hopefully, this thesis can provide benefits to readers and parties in need.

Bandar Lampung, May 26, 2026

Najla Khansa Alifah

TABLE OF CONTENTS

LIST OF FIGURE

LIST OF TABLE

I. INTRODUCTION

1.1 Background and Issues

Time series forecasting is a branch of predictive analytics used to estimate the value of a variable in a future period based on observations or measurements from previous periods. In the context of time series, data are collected and arranged chronological, so the objective of forecasting is to generate an estimate of future values based on the historical patterns that have been formed (Kolambe & Arora, 2024).

Classical statistical models commonly used in time series forecasting include the Autoregressive (AR) model, Moving Average (MA) model, their combination known as the Autoregressive Moving Average (ARMA) model, and its extension, the Autoregressive Integrated Moving Average (ARIMA) model. The AR model assumes that the value of a variable in the current period is influenced by its values in previous periods, whereas the MA model considers the influence of errors from previous periods. The ARMA model combines these two approaches. The AR, MA, and ARMA models are employed to model stationary data. Meanwhile, the ARIMA model represents an extension of ARMA by incorporating a differencing process to address non-stationarity in time series data (Box et al., 2015).

In practice, a time series is not always influenced solely by its own historical values, but may also be affected by other relevant variables. Therefore, the Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) model was developed as an extension of ARIMA that incorporates one or more additional time series as exogenous variables. The ARIMAX model extends ARIMA by adding the linear influence of exogenous variables on the stationary response series. By taking these external factors into account, the model has the potential to enhance the comprehensiveness and accuracy of the predictions (Somyanonthanakul *et al.*, 2022).

Several studies have applied the ARIMAX method in time series forecasting. Yang et al. (2023) employed the ARIMAX model to forecast pulmonary tuberculosis (PTB) cases in China by incorporating the Baidu Search Index as an exogenous variable. The study compared the performance of the ARIMA and ARIMAX models, and the evaluation results indicated that the ARIMAX model demonstrated superior performance, with an AIC value of 2761.58 and a MAPE of 5.34%, compared to the ARIMA model, which yielded an AIC value of 2804.41 and a MAPE of 13.19%.

Another study conducted by Salsabila & Oktaviarina (2024) applied the ARIMAX model to forecast the Gross Regional Domestic Product (GRDP) of East Java Province by incorporating exports and imports variables as exogenous variables. The best model obtained was ARIMAX(0,2,1), with an AIC value of 892.62 and good forecasting accuracy on the testing data, indicated by a MAPE value of 1.76%.

In addition, Othman et al. (2025) applied the ARIMAX model to forecast OPEC oil prices by incorporating exogenous variables including OPEC production, global demand and supply, transportation costs, the GDP of OPEC and non-OPEC countries, and the world population. In the modeling process, multicollinearity testing was conducted on the exogenous variables using the Variance Inflation Factor (VIF), where several variables exhibited VIF values greater than 10, indicating a high correlation among the explanatory variables. To address this issue, the researchers applied wavelet decomposition, which effectively reduced the VIF values and mitigated the problem of multicollinearity. The results of the study concluded that the ARIMAX model with an ARIMA(2,1,1) specification provided the best forecasting performance, with a MAPE of 1.86%, and demonstrated stable and consistent results based on Monte Carlo simulations.

Based on several previous studies, the ARIMAX model demonstrates good forecasting performance by incorporating exogenous variables in the modeling process. However, the use of more than one exogenous variable may potentially lead to multicollinearity issues. Multicollinearity refers to a condition in which there is a high correlation among independent variables within a model, which can undermine the reliability of the resulting model. Therefore, in ARIMAX models involving multiple exogenous variables, testing for and addressing multicollinearity becomes an important aspect (Othman et al., 2025).

Principal Component Analysis (PCA) is one of the approaches that can be used to address multicollinearity issues. PCA aims to reduce data dimensionality by simplifying a set of observed variables. This process is carried out by transforming the original independent variables into a new set of variables that are uncorrelated with one another, known as principal components (Chairunnisa et al., 2025).

Several studies have applied the PCA method to address multicollinearity issues. A study conducted by Alphonsus & Raji (2019) demonstrated that PCA can be effectively used to handle multicollinearity while simultaneously reducing the dimensionality of morphological data in Bunaji cows. From 16 morphological variables exhibiting high correlations, PCA successfully reduced these variables into four principal components that are mutually uncorrelated and able to explain 90.45% of the total data variation.

In addition, in a study conducted by Chairunnisa et al. (2025), PCA was applied to address multicollinearity issues in multiple linear regression on Human Development Index (HDI) data in Kalimantan Island. Of the five independent variables employed, two exhibited Variance Inflation Factor (VIF) values greater than 10, indicating the presence of multicollinearity. To address this issue, PCA was applied and resulted in two principal components selected based on a cumulative variance proportion of 89% of the total data variation.

The exchange rate plays an important role in the global economy as it influences international trade, investment, interest rates, and inflation. Its unpredictable fluctuations create uncertainty for policymakers and market participants, thereby necessitating accurate modeling and forecasting. As a financial time series, the exchange rate also exhibits complex characteristics, including autocorrelation and linkages with macroeconomic factors (Yerima et al., 2025). In the global economy, the international exchange rate standard is still predominantly dominated by the United States currency, namely the United States dollar, which also serves as the primary currency used by Indonesia in most international transactions (Manihuruk et al., 2023).

Several studies have analyzed the factors influencing the rupiah exchange rate against the United States dollar. Carissa & Khoirudin (2020) examined the effects of money supply, interest rates, inflation, and imports on the rupiah exchange rate using a multiple linear regression approach. The results indicated that, simultaneously, all four variables had a significant effect on the rupiah exchange rate, with a coefficient of determination of 75.3%. Partially, money supply, interest rates, and imports were found to have a positive and significant effect on the rupiah exchange rate, whereas inflation did not have a significant effect.

Furthermore, a study conducted by Manihuruk et al. (2023) analyzed the effects of exports, imports, and money supply on the rupiah exchange rate against the United States dollar using a multiple linear regression method. The results indicated that, simultaneously, all three variables had a significant effect on the rupiah exchange rate, with a coefficient of determination of 88.85%. Partially, exports had a negative and significant effect on the rupiah exchange rate, money supply had a positive and significant effect, whereas imports did not have a significant effect.

Based on the aforementioned background, this study aims to apply the Principal Component Analysis method to address multicollinearity issues in the ARIMAX model for forecasting the rupiah exchange rate against the United States dollar for the period of January to December 2025 by incorporating exogenous variables including money supply, interest rates, inflation, exports, and imports.

1.2 Research Objectives

This study aims to apply the Principal Component Analysis method to address multicollinearity in the ARIMAX model for forecasting the rupiah exchange rate against the United States dollar for the period from January to December 2025.

1.3 Research Benefits

This study is expected to provide benefits in the form of recommendations or considerations for the government and financial institutions in formulating policies related to the stability of the rupiah exchange rate against the United States dollar, so that the policies implemented can be more targeted and effective. In addition, this study is also expected to enhance scientific insight and knowledge for both the author and readers regarding the application of the Principal Component Analysis method in addressing multicollinearity in the ARIMAX model.

II. LITERATURE REVIEW

2.1 Review Matrix

2.1.1 Definition of Matrix

A matrix is defined as a collection of numbers arranged in rows and columns forming a rectangular array. The numbers that form a matrix are called entries or elements of the matrix, and they are written inside brackets, either ( ) or [ ]. In general, a matrix is denoted by a capital letter, for example $\boldsymbol{A}$, while the entry of the matrix $\boldsymbol{A}$ in the i-th row and j-th column is denoted by a_{ij}. If matrix $\boldsymbol{A}$ consists of m rows and n columns, then the order of the matrix is written as $m \times n$. Matrix $\boldsymbol{A}$ can be expressed as follows (Anton & Rorres, 2013):

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

which can equivalently be written as:

$$\boldsymbol{A} = [a_{ij}]$$

where $i = 1,2,3, \ldots, m$ and $j = 1,2,3, \ldots, n$.

2.1.2 Transpose of Matrix

Suppose matrix $\boldsymbol{A}$ has order $m \times n$. The transpose of matrix, $\boldsymbol{A}$ denoted by $\boldsymbol{A}^T$**,** is obtained by interchanging the rows and columns of matrix $\boldsymbol{A}$. That is, each row of matrix $\boldsymbol{A}$ becomes a column of matrix $\boldsymbol{A}^T$, and each column of matrix $\boldsymbol{A}$ becomes a row of matrix $\boldsymbol{A}^T$. Thus, the order of matrix $\boldsymbol{A}^T$ is $n \times m$ (Anton & Rorres, 2013).

$$\boldsymbol{A}^T = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}.$$

If a matrix $\boldsymbol{A}$ is transposed twice, the result is the matrix itself, as shown by the following property:

$$(\boldsymbol{A}^T)^T = \boldsymbol{A}.$$

2.1.3 Determinant of Matrix

Suppose $\boldsymbol{A}$ is a square matrix of order $n \times n$. The determinant of $\boldsymbol{A}$ can be computed by multiplying the entries in a row or a column of matrix $\boldsymbol{A}$ by their corresponding cofactors and summing the resulting products. The resulting scalar value is called the determinant of matrix $\boldsymbol{A}$, and the process is known as cofactor expansion of matrix $\boldsymbol{A}$. The determinant of matrix $\boldsymbol{A}$ can be calculated through cofactor expansion along any row or any column of the matrix. In general, the expansion can be written as follows (Anton & Rorres, 2013):

- Cofactor expansion along the j-th column:

$$det(\boldsymbol{A}) = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}$$

- Cofactor expansion along the i-th row:

$$det(\boldsymbol{A}) = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}$$

where:

a_{ij} = the entry of matrix $\boldsymbol{A}$ in the i-th row and j-th column

C_{ij} = the cofactor of entry a_{ij}, defined as $C_{ij} = (-1)^{i+j}M_{ij}$, and M_{ij} is the minor of entry a_{ij}, namely the determinant of the submatrix obtained by deleting the i-th row and j-th column of matrix $\boldsymbol{A}$.

2.1.4 Invers of Matriks

Suppose $\boldsymbol{A}$ s a square matrix of order $n \times n$. Matrix $\boldsymbol{A}$ has an inverse, denoted by $\boldsymbol{A}^{-1}$, if it satisfies the following property (Anton & Rorres, 2013):

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n$$

where $\boldsymbol{I}_n$ is the identity matrix of order $n \times n$.

The inverse of a matrix can be obtained using the adjoint method, with the formula given as follows (Indriati, 2019):

$$\boldsymbol{A}^{-1} = \frac{1}{det(\boldsymbol{A})} Adj(\boldsymbol{A})$$

where $Adj(\boldsymbol{A})$ denotes the adjoint matrix of $\boldsymbol{A}$.

2.1.5 Eigenvalues and Eigenvectors

Suppose $\boldsymbol{A}$ is a square matrix of order $n \times n$. A scalar λ is called an eigenvalue of matrix $\boldsymbol{A}$, and a nonzero vector $\boldsymbol{x} \in \mathbb{R}^n$ is called an eigenvector corresponding to λ, if it satisfies the equation:

$$\boldsymbol{Ax} = \lambda \boldsymbol{x} \qquad (2.1).$$

Equation (2.1) can be rewritten as:

$$\boldsymbol{Ax} = \lambda \boldsymbol{Ix}$$

thus, the equivalent form is:

$$(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = 0 \qquad (2.2).$$

For Equation (2.2) to have a nontrivial solution $(\boldsymbol{x} \neq 0)$, the coefficient matrix $\lambda \boldsymbol{I} - \boldsymbol{A}$ must be singular, that is, its determinant must be equal to zero. Hence, we obtain:

$$det(\lambda \boldsymbol{I} - \boldsymbol{A}) = 0 \qquad (2.3).$$

Equation (2.3) is called the characteristic equation of matrix $\boldsymbol{A}$. This characteristic equation produces a polynomial of degree n in λ, known as the characteristic polynomial of $\boldsymbol{A}$. The roots of the characteristic polynomial are the eigenvalues of matrix $\boldsymbol{A}$. Since a polynomial of degree n has at most n roots, a matrix of order $n \times n$ has at most n eigenvalues. The eigenvalues obtained are then substituted into Equation (2.2) to determine the corresponding eigenvectors (Anton & Rorres, 2013).

2.2 Time Series and Forecasting

A time series is a set of observations of a variable arranged chronologically at equally spaced time intervals, such as daily, monthly, or yearly intervals (Montgomery et al., 2015). One of the main characteristics of time series data is the presence of dependence among observations that are close in time, therefore time series analysis focuses on modeling this dependence structure (Box *et al*., 2015).

According to Makridakis et al. (1983), there are four types of time series data patterns, namely:

1. Horizontal pattern occurs when the data values fluctuate around a constant mean.

Figure 1. Horizontal Pattern (Makridakis et al., 1983).

2. Seasonal pattern occurs when a time series is influenced by seasonal factors such as quarterly, monthly, or weekly periods.

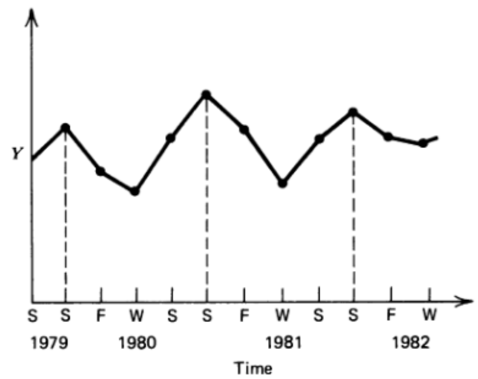

Figure 2. Seasonal pattern (Makridakis et al., 1983).

3. Cyclical pattern occurs when the data are influenced by long-term economic fluctuations, such as those associated with business cycles.

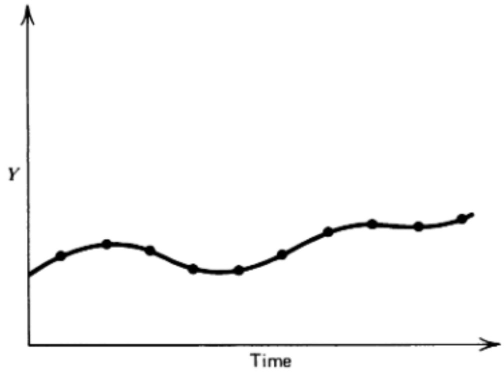

Figure 3. Cyclical pattern (Makridakis et al., 1983).

4. Trend pattern occurs when there is a long-term tendency of increase or decrease in the data values.

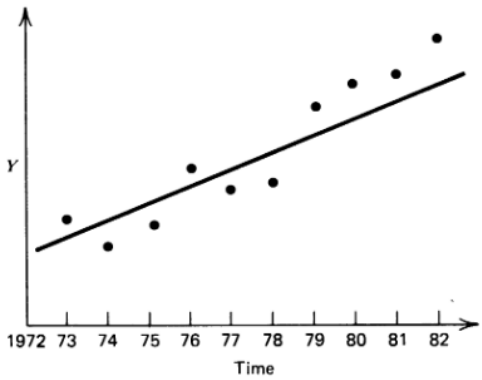

Figure 4. Trend pattern (Makridakis et al., 1983).

Forecasting is the process of estimating future events or the value of a variable in upcoming periods. Based on the time horizon, forecasting is classified into short term, medium term, and long term forecasting. Short term forecasting covers several periods ahead, such as daily, weekly, or monthly forecasts. Medium term forecasting generally covers a period of one to two years ahead, while long term forecasting may extend beyond that period to several years into the future (Montgomery et al., 2015).

In general, forecasting methods are divided into two categories, namely qualitative and quantitative methods. Qualitative methods are subjective and based on expert judgement. In contrast, quantitative methods utilize historical data analyzed through statistical models to capture patterns in the data and project them into future periods (Montgomery et al., 2015).

2.3 Multicollinearity

Multicollinearity is a condition in which two or more independent variables have a strong linear relationship with one another (Chan et al., 2022). Multicollinearity can be detected by calculating the Variance Inflation Factor (VIF) value. If the VIF value exceeds 10, it can be concluded that multicollinearity is present (Gujarati & Porter, 2009). According to Gujarati & Porter (2009), the formula for calculatingVIF is expressed as follows:

$$VIF_j = \frac{1}{1 - R_j^2} \tag{2.4}$$

where:

R_j^2 = coefficient of determination obtained from the regression of the j-th independent variable on all other independent variables.

2.4 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a multivariate analysis technique that constructs a number of linear combinations of variables to obtain principal components that are mutually uncorrelated. The first principal component is the linear combination with maximum variance, while each subsequent principal component is the linear combination with maximum variance subject to being orthogonal to the preceding components. This method is used to reduce the dimensionality of the data while retaining most of the variation present in the original dataset (Rencher, 2002).

The steps in implementing PCA are as follows:

1. Data standardization

   PCA is sensitive to differences in variable scales. Therefore, if the variables have different units of measurement, standardization must first be performed. Differences in scale can affect the variance and consequently influence the formation of principal components. Standardization is carried out by centering and scaling each variable as follows (Jolliffe & Cadima, 2016):

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \tag{2.5}$$

   where:

   z_{ij} = standardized value

   x_{ij} = original value of the i-th observation of the j-th variable

   $\bar{x}_j$ = mean of the j-th variable

   s_j = standard deviation of the j-th variable.

   After standardization, PCA is performed using the correlation matrix.

2. Constructing the covariance or correlation matrix

   PCA is based on either the covariance matrix or the correlation matrix, depending on the scale of measurement of the data. If the variables are measured in the same units, the covariance matrix is used and is formulated as:

$$\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})^T$$

   where Y_i represents the i-th observation vector and $\bar{Y}$ denotes the mean vector.

If the variables are measured on different scales, the correlation matrix is used. The correlation between the j-th and k-th variables is expressed as:

$$r_{jk} = \frac{s_{jk}}{s_j s_k}.$$

The correlation matrix can also be obtained from the covariance matrix through:

$$\boldsymbol{R} = \boldsymbol{D}_s^{-1} \boldsymbol{S} \boldsymbol{D}_s^{-1}$$

where $\boldsymbol{D}_s$ is a diagonal matrix containing the standard deviations of each variable (Rencher, 2002).

3. Computing eigenvalues and eigenvectors

After the covariance or correlation matrix has been constructed, the next step is to compute the eigenvalues and eigenvectors through the following equations:

$$\boldsymbol{S}\boldsymbol{a}_j = \lambda_j \boldsymbol{a}_j$$

or

$$\boldsymbol{R}\boldsymbol{a}_j = \lambda_j \boldsymbol{a}_j$$

where λ_j is the eigenvalue and $\boldsymbol{a}_j$ is the corresponding eigenvector. The eigenvalue represents the amount of variance explained by the j-th principal component, while the eigenvector indicates the direction of the component in the multidimensional space. The eigenvalues are arranged from the largest to the smallest so that the first component explains the greatest variance in the data.

The proportion of variance explained by the j-th principal component is calculated by comparing the j-th eigenvalue to the sum of all eigenvalues, given by:

$$\frac{\lambda_j}{\sum_{k=1}^{p} \lambda_p} \qquad (2.6).$$

Since the sum of all eigenvalues is equal to the total variance of the data, this ratio indicates the relative contribution of each principal component to the overall variation in the dataset (Rencher, 2002).

4. Determining the number of principal components

   Determining the number of principal components is an important step in PCA, as it determines the extent to which dimensionality reduction can be achieved without losing meaningful information. The commonly used criteria for determining the number of components are as follows (Rencher, 2002):

   1. Retaining components until the cumulative proportion of explained variance reaches a certain level, for example 80% of the total variance.
   2. Retaining components whose eigenvalues are greater than the average eigenvalue, that is,

$$\lambda_j > \frac{\sum_{i=1}^{p} \lambda_i}{p}.$$

      In PCA based on the correlation matrix, the average eigenvalue is equal to 1, so this criterion is known as the rule $\lambda > 1$.

5. Forming the principal components

   Each principal component is expressed as a linear combination of the variables analyzed (Rencher, 2002):

$$PC_j = a_{j1}X_1 + a_{j2}X_2 + \cdots + a_{jp}X_p = \boldsymbol{a}_j^T\boldsymbol{X} \qquad (2.7)$$

where:

PC_j	=	j-th principal component
a_{jp}	=	loading of the p-th variable on the j-th component
$X_1, \ldots, X_p$	=	variables used in the analysis, which may be the original variables (in PCA based on the covariance matrix) or standardized variables (in PCA based on the correlation matrix).

2.5 Stasionarity

Stationarity is one of the characteristics that must be considered in time series analysis. A time series is said to be stationary if it has a constant mean and variance over time, and if the covariance between two periods depends only on the length of the lag and not on the actual time at which it is computed (Gujarati & Porter, 2009). Nonstationary data generally contain a unit root or a deterministic trend. The presence of a unit root indicates that the effect of a shock is permanent, causing the data to follow a random walk pattern and making it difficult to forecast (Rifai & Zhahirulhaq, 2024).

In time series analysis, one of the steps to achieve variance stationarity is through data transformation. One commonly used transformation is the Box-Cox transformation, which was introduced by Box and Cox in 1964. The Box-Cox transformation is expressed as follows (Cryer & Chan, 2008):

$$g(Y) = \begin{cases} \dfrac{Y^{\lambda} - 1}{\lambda}, & \lambda \neq 0 \\ logY, & \lambda = 0 \end{cases}$$

the parameter λ determines the form of transformation applied to the data.

To test for stationarity in the mean, the Augmented Dickey-Fuller (ADF) test is commonly used to detect the presence of a unit root in time series data (Gujarati & Porter, 2009). In general, the ADF test model can be expressed as follows:

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum_{i=1}^{m} \alpha_i \, \Delta Y_{t-i} + \varepsilon_t.$$

The steps in conducting the ADF test are as follows (Petrică et al., 2017).

1. Hypotheses:
    - $H_0: \delta \geq 0$, indicating that the data contain a unit root (non-stationary).

- $H_1: \delta < 0$, indicating that the data do not contain a unit root (stationary).

2. Significance level:

   $\alpha = 5\% = 0.05$.

3. Test statistic:

   The ADF test uses the t statistic, which is written as.

$$t_\delta = \frac{\hat{\delta}}{Se(\hat{\delta})}$$

   where:

   $\hat{\delta}$ = estimated coefficient

   $Se(\hat{\delta})$ = standard error of $\hat{\delta}$.

4. Decision rule:

   - Reject H_0 if the calculated t statistic is less than the Dickey-Fuller critical value or if the $p - value < \alpha$.
   - Fail to reject H_0 if the calculated t-statistic is greater than the Dickey-Fuller critical value or if the $p - value > \alpha$.

5. Decision.
6. Conclusion.

2.6 Differencing

Differencing is a method used to transform time series data into a more stationary form. The first differencing of a time series is expressed as follows (Montgomery et al., 2015):

$$\nabla Y_t = Y_t - Y_{t-1}$$

using the backshift operator B, defined as $BY_t = Y_{t-1}$, the first differencing can be written as:

$$\nabla Y_t = (1 - B)Y_t.$$

If the data are still not stationary after the first differencing, a second differencing can be applied as follows:

$$\nabla^2 Y_t = (1 - B)^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2}.$$

In general, the d-th order differencing is defined as:

$$\nabla^d Y_t = (1 - B)^d Y_t$$

with $B^d Y_t = Y_{t-d}$.

Differencing is used to eliminate trends or nonstationary components in the data. In the ARIMA model, the integrated component (d) indicates the number of differencing operations required to make the time series stationary. This process aims to address the presence of a unit root so that the resulting model becomes more stable and produces better forecasts (Rifai & Zhahirulhaq, 2024).

2.7 Autocorrelation Function and Partial Autocorrelation Function

In time series analysis, the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) are statistical tools used to identify patterns of temporal dependence and to assist in the model identification process. The patterns formed in the ACF and PACF plots are used to determine the order of the Autoregressive (AR) and Moving Average (MA) models, where the PACF helps identify the order p of the AR component, while the ACF helps identify the order q of the MA component (Cryer & Chan, 2008).

The Autocorrelation Function (ACF) is used to measure the correlation between an observation Y_t and Y_{t+k} at lag k. The sample ACF is expressed as follows (Wei, 2006):

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k}(Y_t - \bar{Y})\,(Y_{t+k} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}.$$

Meanwhile, the Partial Autocorrelation Function (PACF) measures the correlation between Y_t and Y_{t+k} after removing the effects of the intervening observations, namely $Y_{t+1}, Y_{t+2}, \dots, Y_{t+k-1}$. The sample PACF is expressed as follows (Wei, 2006):

$$\hat{\phi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^{k}\hat{\phi}_{kj}\,\hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^{k}\hat{\phi}_{kj}\,\hat{\rho}_j}$$

and

$$\hat{\phi}_{k+1,j} = \hat{\phi}_{kj} - \hat{\phi}_{k+1,k+1}\hat{\phi}_{k,k+1-j}.$$

2.8 Time Series Models

In time series analysis, there are various commonly used models, including the Autoregressive (AR) model, the Moving Average (MA) model, the Autoregressive Moving Average (ARMA) model, and the Autoregressive Integrated Moving Average (ARIMA) model.

2.8.1 Autoregressive (AR) Model

The Autoregressive (AR) model is a model for stationary time series data that assumes the value of a variable in the current period is influenced by its past values. The following is the autoregressive model of order p (Box et al., 2015).

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$

where:

Y_t = data at time period t

$Y_{t-1}, \dots, Y_{t-p}$ = data at time periods t-1, ... , t-p

$\phi_1, \dots, \phi_p$ = AR model parameters

ε_t = error term at time t.

The AR(p) model can also be expressed using the backshift operator B,where $B^k Y_t = Y_{t-k}$. Using this operator, the AR(p) model becomes:

$$Y_t = \phi_1 B Y_t + \phi_2 B^2 Y_t + \cdots + \phi_p B^p Y_t + \varepsilon_t$$
$$\left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right) Y_t = \varepsilon_t$$
$$\phi_p(B) Y_t = \varepsilon_t.$$

2.8.2 Moving Average (MA) Model

The Moving Average (MA) model is a time series model used for stationary data, which explains that the value of a variable in the current period is influenced by the error terms from the current and previous periods. The following is the moving average model of order q (Box et al., 2015).

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

where:

Y_t = data at time period t

$\theta_1, \dots, \theta_q$ = MA model parameters

ε_t = error term at time t

$\varepsilon_{t-1}, \dots, \varepsilon_{t-q}$ = error terms at times t-1, ... , t-q.

The MA(q) model can also be expressed using the backshift operator B. Using this operator, the MA(q) model becomes:

$$Y_t = \varepsilon_t - \theta_1 B\varepsilon_t - \theta_2 B^2\varepsilon_t - \cdots - \theta_q B^q\varepsilon_t$$
$$Y_t = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right)\varepsilon_t$$
$$Y_t = \theta_q(B)\varepsilon_t.$$

2.8.3 Autoregressive Moving Average (ARMA) Model

The Autoregressive Moving Average (ARMA) model is a combination of the AR model of order p and the MA model of order q, which is used to analyze and model stationary time series data. The ARMA model explains that the current value of a variable is influenced by its past values as well as by past error terms. The following is the ARMA(p, q) model (Box et al., 2015).

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q}$$

where:

Y_t	= data at time period *t*
$Y_{t-1}, \dots, Y_{t-p}$	= data at time period *t-1, ... , t-p*
$\phi_1, \dots, \phi_p$	= AR model parameters
$\theta_1, \dots, \theta_q$	= MA model parameters
ε_t	= error term at time *t*
$\varepsilon_{t-1}, \dots, \varepsilon_{t-q}$	= error terms at times *t-1, ... , t-q*.

The ARMA(p, q) model can also be expressed using the backshift operator B. Using this operator, the ARMA(p, q) model becomes:

$$Y_t = \phi_1 BY_t + \phi_2 B^2 Y_t + \cdots + \phi_p B^p Y_t + \varepsilon_t - \theta_1 B\varepsilon_t - \theta_2 B^2\varepsilon_t - \cdots - \theta_q B^q\varepsilon_t$$
$$\left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right)Y_t = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right)\varepsilon_t$$
$$\phi_p(B)Y_t = \theta_q(B)\varepsilon_t.$$

2.8.4 Autoregressive Integrated Moving Average (ARIMA) Model

The Autoregressive Integrated Moving Average (ARIMA) model is an extension of the ARMA model that incorporates a differencing component d to transform non-stationary time series into stationary ones. Using the backshift operator B, the ARIMA (p, d, q) model is expressed as follows (Box et al., 2015):

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B)\varepsilon_t$$

where:

Y_t	= data at time period t
$\phi_p(B)$	= AR parameter of order p
$\theta_q(B)$	= MA parameter of order q
$(1-B)^d$	= d-th order differencing
ε_t	= error term at time t.

2.9 Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) Model

The Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) model is an extension of the ARIMA model that incorporates one or more additional time series as exogenous variables. The ARIMAX(p, d, q) model extends ARIMA(p, d, q) by adding the linear effects of exogenous variables on the response series that has been made stationary. By taking these external factors into account, the model can improve the completeness and accuracy of the forecasts (Somyanonthanakul et al., 2022). Using the backshift operator B, the ARIMAX(p, d, q) model is expressed as follows (Novianda, 2025):

$$\phi_p(B)(1-B)^d Y_t = \mu + \sum_{i=1}^{k} \alpha_i X_{i,t} + \theta_q(B)\varepsilon_t \tag{2.8}$$

where:

Y_t	= data at time period *t*
$\phi_p(B)$	= AR parameter of order *p*
$\theta_q(B)$	= MA parameter of order *q*
$(1-B)^d$	= d-th order differencing
μ	= intercept
α_i	= coefficient of the *i*-th exogenous variable
$X_{i,t}$	= the *i*-th exogenous variable at time *t*
ε_t	= error term at time *t*.

2.10 Parameter Estimation

There are several methods that can be used to estimate the parameters of time series models, including the Method of Moments, Maximum Likelihood Estimation (MLE), and Ordinary Least Squares (OLS) (Wei, 2006). In this study, the method selected for parameter estimation is MLE.

As an example, the ARMA(p,q) model is given as follows (Shumway & Stoffer, 2017):

$$Y_t = \mu + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

with $\varepsilon_t \sim iid\ N(0, \sigma^2)$.

For the general ARMA(p,q) model, let the parameter vector be:

$$\beta = \left(\mu, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q\right)'.$$

For the general ARMA model, the likelihood is more conveniently expressed in terms of innovations, namely the one-step-ahead prediction error, defined as $Y_t - Y_t^{t-1}$.

The likelihood function can be written as the product of conditional distributions:

$$L(\beta, \sigma^2) = \prod_{t=1}^{n} f(Y_t | Y_{t-1}, \dots, Y_1).$$

The conditional distribution is Gaussian with mean Y_t^{t-1} and variance $P_t^{t-1} = \sigma^2 r_t$.

Thus, the likelihood function can be expressed as:

$$L(\beta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \prod_{t=1}^{n} r_t(\beta)^{-1/2} exp\left[-\frac{S(\beta)}{2\sigma^2}\right]$$

where

$$S(\beta) = \sum_{t=1}^{n} \frac{\left(Y_t - Y_t^{t-1}(\beta)\right)^2}{r_t(\beta)}.$$

Based on the obtained likelihood function, the log-likelihood function can be expressed as follows:

$$\ln L(\beta, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2}\sum_{t=1}^{n} \ln r_t(\beta) - \frac{S(\beta)}{2\sigma^2}.$$

By differentiating the log-likelihood with respect to σ^2 and setting it equal to zero, the following result is obtained:

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{S(\beta)}{2\sigma^4} = 0.$$

Multiplying both sides by $2\sigma^4$:

$$-n\sigma^2 + S(\beta) = 0.$$

Thus, the maximum likelihood estimator for the error variance is obtained as:

$$\hat{\sigma}^2 = \frac{S(\hat{\beta})}{n}.$$

Substituting this estimator back into the log-likelihood yields the concentrated likelihood:

$$\ell(\beta) = ln[n^{-1}S(\beta)] + n^{-1}\sum_{t=1}^{n} lnr_t(\beta).$$

The estimate $\hat{\beta}$ is obtained by minimizing the function $\ell(\beta)$ with respect to β. In general, numerical optimization procedures are used to obtain the estimator and its standard error. One commonly used procedure is the Newton-Raphson algorithm, which is iterated until convergence is achieved.

2.11 Parameter Significance Test

The parameters in a time series model that have been estimated need to be tested for significance to determine whether they have a significant effect within the model. If a parameter is not significant, it is considered not to provide a meaningful contribution to the constructed model. The hypotheses for the parameter significance test are as follows (Montgomery et al., 2015).

1. Hypotheses:
   - $H_0: \beta_i = 0$ (the model parameter coefficients are not significant).
   - $H_1: \beta_i \neq 0$ (the model parameter coefficients are significant).
2. Significance level:
   $\alpha = 5\% = 0.05$.

3. Test statistic:

$$t = \frac{\widehat{\beta_i}}{Se(\widehat{\beta_i})}$$

where:

t	=	test statistic value
$\widehat{\beta_i}$	=	estimated coefficient of the i-th parameter
$Se(\widehat{\beta_i})$	=	standard error of $\widehat{\beta_i}$.

4. Decision rule:
   - Reject H_0 if $|t_{calculated}| > t_{\alpha/2,n-p}$ or if the $p - value < \alpha$.
   - Fail to reject H_0 if $|t_{calculated}| < t_{\alpha/2,n-p}$ or if the $p - value > \alpha$.
5. Decision.
6. Conclusion.

2.12 Residual Diagnostics

Residuals in a time series model are defined as the differences between the actual observed values and the values estimated by the model. Residual analysis is used to evaluate whether the model has adequately captured the patterns present in the data (Rifai & Zhahirulhaq, 2024). One of the important assumptions in time series analysis is that the residuals follow a white noise process, meaning that they are uncorrelated, have a mean equal to zero, and have constant variance (Panjaitan et al., 2018). The white noise property can be tested using the Ljung-Box statistic, with the following procedure (Ljung & Box, 1978).

1. Hypotheses:
   - $H_0: \rho_1 = \rho_2 = \cdots = \rho_m = 0$ (the residuals are white noise).
   - H_1: At least one $\rho_k \neq 0$ (the residuals are not white noise).
2. Significance level:

   $\alpha = 5\% = 0.05$.

3. Test statistic:

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

where:

n = number of observations

$\hat{\rho}_k$ = autocorrelation at lag k

m = number of lags.

4. Decision rule:
   - Reject H_0 if $Q > \chi^2_{(m-p-q)}$ or if the $p-value < \alpha$.
   - Fail to reject H_0 if $Q < \chi^2_{(m-p-q)}$ or if the $p-value > \alpha$.
5. Decision.
6. Conclusion.

2.13 Model Evaluation

In this study, the selection of the best model is based on the Akaike Information Criterion (AIC), while forecasting performance is evaluated using the Mean Absolute Percentage Error (MAPE). The formulas for AIC and MAPE are presented as follows:

1. Akaike Information Criterion (AIC)

   The Akaike Information Criterion (AIC) is used to compare several competing models. The AIC value is calculated using the following formula (Othman et al., 2025):

$$AIC = 2k - 2\ln(\hat{L})$$

where:

k = number of parameters estimated in the model

$\hat{L}$ = maximum likelihood of the model.

In model selection, the model with the smallest AIC value is chosen as the best model among the candidate models that adequately fit the data (Hyndman & Khandakar, 2008).

2. Mean Absolute Percentage Error (MAPE)

The Mean Absolute Percentage Error (MAPE) is a percentage-based metric that measures the average absolute percentage difference between the predicted values and the actual values. This metric is used to evaluate forecasting accuracy based on the observed actual values (Kolambe & Arora, 2024). The formula for MAPE is given as follows (Makridakis et al., 1983):

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \times 100\%$$

where:

n = total number of observations

Y_t = actual value at time t

$\hat{Y}_t$ = forecasted value at time t.

The criteria for interpreting MAPE values are presented in Table 1 (Erdianto, 2023):

Table 1. MAPE value interpretation

MAPE Value	**Forecast Accuracy Interpretation**
< 10%	High forecasting accuracy
10% - 20%	Good forecasting accuracy
20% - 50%	Acceptable forecasting accuracy
> 50%	Poor forecasting accuracy

2.14 Rupiah Exchange Rate against the United States Dollar

The exchange rate is the price of one currency in terms of another currency. The exchange rate may fluctuate in the form of depreciation or appreciation (Daleno et al., 2023). The exchange rate plays a crucial role in supporting international transaction activities. In addition, it is often used as one of the indicators to assess a country's economic condition (Sutandi et al., 2021).

In the global economy, the international exchange rate standard is still predominantly dominated by the United States currency, namely the United States dollar, which also serves as the primary currency used by Indonesia in most international transactions (Manihuruk et al., 2023). In studies conducted by Carissa & Khoirudin (2020) and Manihuruk et al. (2023), the exchange rate of the Indonesian rupiah against the United States dollar was analyzed by considering several macroeconomic factors, namely money supply, interest rates, inflation, exports, and imports.

According to Bank Indonesia (n.d.), the money supply is divided into two, namely in a narrow sense (M1) and in a broad sense (M2). M1 includes currency held by the public and demand deposits (rupiah denominated current accounts), while M2 includes M1, quasi money such as savings, time deposits in rupiah and foreign currency, as well as foreign currency demand deposits, and securities issued by the monetary system held by the domestic private sector with a remaining maturity of up to one year.

The interest rate (BI Rate) is the policy rate set by Bank Indonesia as a benchmark for financial institutions in determining the interest rates offered to customers, including lending and deposit rates (Bank Indonesia, n.d.).

Inflation is a condition characterized by a general and continuous increase in the prices of goods and services over a certain period of time. Price increases that occur only in one or two types of goods cannot be classified as inflation, unless such increases are widespread and lead to price increases in other goods (Bank Indonesia, n.d.).

Exports refer to the activity of sending goods and services from within a country to abroad, whether for commercial or non-commercial purposes, including goods processed abroad and then brought back into the country legally. Meanwhile, imports refer to the activity of bringing goods and services from abroad into a country carried out by the residents of that country, which results in the outflow of foreign currency (Kementerian Perdagangan Republik Indonesia, n.d.).

III. RESEARCH METHODOLOGY

3.1 Place and Time for Research

This study was conducted during the even semester of the 2025/2026 academic year at the Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Lampung.

3.2 Data for Research

The data used in this study are secondary data in the form of monthly time series data covering the period from January 2012 to December 2024, resulting in a total of 156 observations. These data were obtained from Badan Pusat Statistik (BPS), Bank Indonesia (BI), and Satu Data Perdagangan Kementerian Perdagangan Republik Indonesia. Table 2 presents the variables used in this study.

Table 2. Research variables

Variable	Description	Unit	Source
Y	Rupiah-USD exchange rate	IDR/USD	https://satudata.kemendag.go.id/data-informasi/perdagangan-dalam-negeri/nilai-tukar
X_1	Money supply (M2)	Billion IDR	https://www.bps.go.id/id/statistics-table/2/MTIzIzI=/uang-beredar--milyar-rupiah-.html

Variable	Description	Unit	Source
X_2	Interest rate (BI rate)	%	https://www.bps.go.id/id/statistics-table/2/Mzc5IzI=/bi-rate.html
X_3	Inflation	%	https://www.bi.go.id/id/statistik/indikator/data-inflasi.aspx
X_4	Export	Million USD	https://satudata.kemendag.go.id/data-informasi/perdagangan-luar-negeri/ekspor-impor
X_5	Import	Million USD	https://satudata.kemendag.go.id/data-informasi/perdagangan-luar-negeri/ekspor-impor

3.3 Research Method

In this study, the rupiah exchange rate against the United States dollar is forecast using the ARIMAX model. Prior to model estimation, Principal Component Analysis is applied to the exogenous variables to address multicollinearity. All stages of the analysis are conducted computationally using RStudio software. The steps undertaken in this study are as follows:

1) Conducting descriptive analysis of the research variables.
2) Examining the presence of multicollinearity among the exogenous variables using the VIF test.
3) Applying PCA to the exogenous variables.
4) Constructing a new dataset based on the principal components obtained from PCA.
5) Testing the stationarity of the dependent variable, using the Box-Cox test to examine variance stationarity and the ADF test to assess stationarity in the mean.
6) Estimating the ARIMAX model using the newly constructed dataset derived from the principal components as exogenous variables. Model identification is conducted using the analysis of ACF and PACF plots.

7) Determining the best model based on the AIC and MAPE values.
8) Conducting diagnostic tests on the residuals of the selected model using the Ljung-Box test.
9) Forecasting the rupiah exchange rate against the United States dollar for the period January 2025 to December 2025.
10) Interpreting the research results.

Figure 5 illustrates the research flowchart for forecasting the rupiah exchange rate against the United States dollar using PCA and the ARIMAX model.

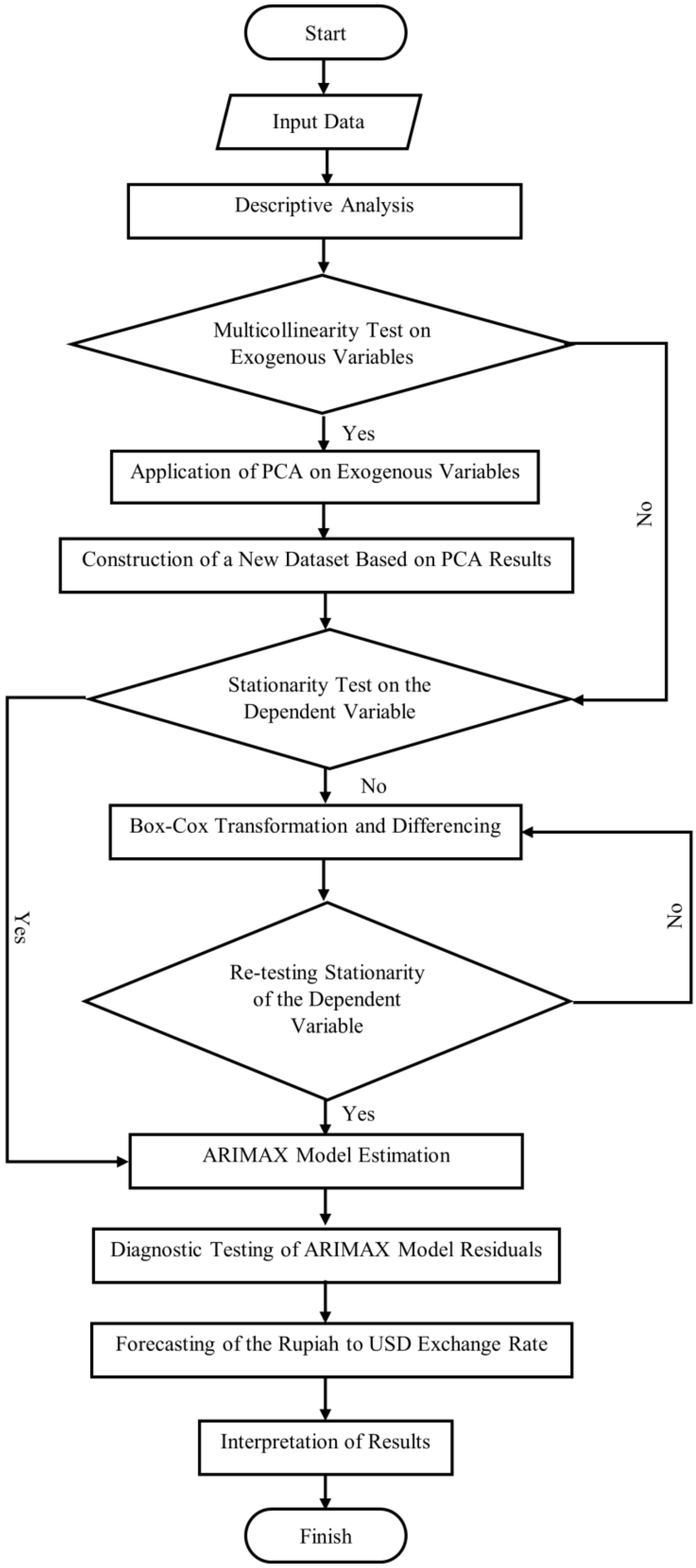

Figure 5. Research Flowchart

V. CONCLUSION

5.1 Conclusions

Based on the results and discussion presented in the previous chapters, the following conclusions are drawn.

1. Principal Component Analysis was employed to address multicollinearity among the exogenous variables, yielding two principal components that explain 85.55% of the total variance.
2. The ARIMAX model that is most appropriate for forecasting the rupiah exchange rate against the United States dollar is the ARIMAX(0,1,2), with the model equation expressed as follows:

$$(1-B)ln(Y_t) = 0.004 - 0.00881Z_{1,t} + 0.016683Z_{2,t} + 0.018115Z_{3,t} - 0.0005Z_{4,t} + 0.002865Z_{5,t} + (1 - 0.1674B - 0.1879B^2)\varepsilon_t$$

3. The forecasting results of the rupiah exchange rate against the United States dollar using the ARIMAX(0,1,2) model for the period January to December 2025 show a mean value of IDR 16320.1 per USD, with a minimum forecast value of IDR 15820.37 per USD and a maximum forecast value of IDR 16734.54 per USD.
4. The comparison between the forecasted values and the actual values for the period January to December 2025 indicates that the ARIMAX(0,1,2) model is able to follow the exchange rate movement pattern well, with relatively small forecasting errors (APE), ranging from 0.15546% to 3.716329%. Overall, the MAPE of 1.263363% indicates that the model has a high level of forecasting accuracy for the period.

5.2 Recommendations

Future research is recommended to expand and vary the exogenous variables, both from domestic and external factors, in order to enrich the information used in the modeling process and potentially improve forecasting performance. An initial analysis of the relationship between the dependent variable and the predictors may also be conducted prior to the application of PCA and ARIMAX modeling to obtain an understanding of the direction and strength of the relationships among variables. In this study, PCA was primarily employed to address multicollinearity among the exogenous variables. In addition to the PCA approach, alternative methods such as Ridge Regression, Lasso Regression, or Partial Least Squares (PLS) may be considered to handle multicollinearity and their performance may be compared with PCA.

Furthermore, the data used in this study still exhibit extreme values (outliers) in both the dependent and exogenous variables. Therefore, future research may consider the use of more robust approaches, such as Robust Principal Component Analysis (Robust PCA) and Robust ARIMAX models, to ensure that the modeling results are more resistant to the influence of outliers and yield more stable estimates.

Stationarity testing in this study was focused on the dependent variable. Future research may consider examining the time series characteristics of the exogenous variables to obtain more comprehensive results.

REFERENCES

Alphonsus, C., & Raji, A. O. (2019). Application of Principal Component Analysis (PCA) for Correcting Multicollinearity and Dimension Reduction of Morphological Parameters in Bunaji Cows. *Nigerian Journal of Animal Science*, *21*(2), 1–8.

Anton, H., & Rorres, C. (2013). *Elementary Linear Algebra: Applications Version* (11th ed.). John Wiley & Sons.

Bank Indonesia. (n.d.). *Inflasi*. Retrieved April 16, 2026, from https://www.bi.go.id/id/fungsi-utama/moneter/inflasi/Default.aspx

Bank Indonesia. (n.d.). *Laporan*. Retrieved April 16, 2026, from https://www.bi.go.id/id/publikasi/laporan/default.aspx?kategori=perkembangan+uang+beredar&periode=

Bank Indonesia. (n.d.). *Memahami Suku Bunga Acuan BI: Kunci Ambil Keputusan Keuangan yang Tepat*. Retrieved April 19, 2026, from https://www.bi.go.id/id/publikasi/ruang-media/cerita-bi/Pages/data-suku-bunga-bi-saat-ini.aspx

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). John Wiley & Sons.

Carissa, N., & Khoirudin, R. (2020). The Factors Affecting the Rupiah Exchange Rate in Indonesia. *Jurnal Ekonomi Pembangunan*, *18*(1), 37–46.

Chairunnisa, N. H., Nohe, D. A., & Syaripuddin. (2025). Mengatasi Multikoliniearitas dalam Regresi Linier Berganda Menggunakan Principal Component Analysis. *EKSPONENSIAL*, *16*(1), 1–9.

Chan, J. Y. L., Leow, S. M. H., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z. W., & Chen, Y. L. (2022). Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review. *Mathematics*, *10*(8), 1283.

Cryer, J. D., & Chan, K. S. (2008). *Time series analysis: With applications in R* (2. ed.). Springer New York.

Daleno, V. C., Kumaat, R. J., & Tumangkeng, S. Y. L. (2023). Faktor-Faktor yang Mempengaruhi Fluktuasi Nilai Tukar Rupiah terhadap Dolar Amerika Serikat. *Jurnal Berkala Ilmiah Efisiensi*, *23*(6), 13–24.

Erdianto, M. A. (2023). Perancangan Model Peramalan Jangka Pendek Harga Komoditas Pertanian di Indonesia Menggunakan Machine Learning. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *3*(4), 338–346.

Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics* (5. ed). McGraw-Hill.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for *R. Journal of Statistical Software*, *27*(3), 1–22.

Indriati, K. (2019). *Matriks, Vektor, dan Program Linier*. Penerbit Unika Atma Jaya Jakarta.

Jolliffe, I. T., & Cadima, J. (2016). Principal Component Analysis: A Review and Recent Developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202.

Kementerian Perdagangan Republik Indonesia. (n.d.). *Total Ekspor Impor*. Satu Data Perdagangan Kemendag RI. Retrieved April 19, 2026, from https://satudata.kemendag.go.id/data-informasi/perdagangan-luar-negeri/ekspor-impor

Kolambe, M., & Arora, S. (2024). Forecasting the Future: A Comprehensive Review of Time Series Prediction Techniques. *Journal of Electrical Systems*, *20*(2s), 575–586.

Ljung, G. M., & Box, G. E. P. (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika*, *65*(2), 297–303.

Makridakis, S., Wheelwright, S. C., & McGee, V. E. (1983). *Forecasting: Methods and Applications* (2nd ed.). John Wiley & Sons.

Manihuruk, F. E., Silfani, D., Feby, Y., & Marbun, J. (2023). Analisis Pengaruh Ekspor, Impor, dan Jumlah Uang Beredar di Indonesia terhadap Kurs Rupiah/USD. *Jurnal Riset Ilmu Ekonomi*, *3*(2), 118–129.

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to Time Series Analysis and Forecasting* (2nd ed). John Wiley & Sons.

Novianda, B. (2025). Forecasting Inflation in Indonesia Using ARIMAX: The Role of Money Supply and Exchange Rate. *Equity: Jurnal Ekonomi*, *13*(2), 33–44.

Othman, S. K., Abdulkareem, B. M., & Badal, M. A. (2025). Advanced Arimax Modeling for Opec Oil Forecasting Using Multivariate Wavelet Techniques. *International Journal of Environmental Sciences*, *11*(15s), 2313–2326.

Panjaitan, H., Prahutama, A., & Sudarno. (2018). Peramalan Jumlah Penumpang Kereta Api Menggunakan Metode ARIMA, Intervensi, dan ARFIMA (Studi Kasus: Penumpang Kereta Api Kelas Lokal Ekonomi DAOP IV Semarang). *Jurnal Gaussian*, *7*(1), 96–109.

Petrică, A. C., Stancu, S., & Ghițulescu, V. (2017). Stationarity – The Central Concept in Time Series Analysis. *International Journal of Emerging Research in Management and Technology*, *6*(1), 6–16.

Rencher, A. C. (2002). *Methods of Multivariate Analysis* (2nd ed). John Wiley & Sons.

Rifai, N. A. K., & Zhahirulhaq, M. A. (2024). Forecasting Inflation in Indonesia Using the Autoregressive Integrated Moving Average Method. *Parameter: Journal of Statistics*, *4*(1), 37–45.

Salsabila, N., & Oktaviarina, A. (2024). Peramalan PDRB di Jawa Timur Menggunakan Model ARIMAX dengan Variabel Eksogen Ekspor-Impor. *MATHunesa: Jurnal Ilmiah Matematika*, *12*(1), 208–218.

Shumway, R. H., & Stoffer, D. S. (2017). *Time Series Analysis and Its Applications: With R Examples* (4th ed.). Springer.

Somyanonthanakul, R., Warin, K., Amasiri, W., Mairiang, K., Mingmalairak, C., Panichkitkosolkul, W., Silanun, K., Theeramunkong, T., Nitikraipot, S., & Suebnukarn, S. (2022). Forecasting COVID-19 Cases Using Time Series Modeling and Association Rule Mining. *BMC Medical Research Methodology*, *22*, 281.

Sutandi, Wibowo, S., Sutisna, N., Fung, T. S., & Januardi, L. (2021). Pengaruh Inflasi, Nilai Tukar (Kurs) Rupiah dan Tingkat Suku Bunga terhadap Indeks Harga Saham Gabungan (IHSG) di Bursa Efek Indonesia (BEI) Periode 2016-2020. *AKUNTOTEKNOLOGI : JURNAL ILMIAH AKUNTANSI DAN TEKNOLOGI*, *13*(2), 78–90.

Wei, W. W. S. (2006). *Time Series Analysis: Univariate and Multivariate Methods* (2nd ed.). Pearson Addison Wesley.

Yang, J., Zhou, J., Luo, T., Xie, Y., Wei, Y., Mai, H., Yang, Y., Cui, P., Ye, L., Liang, H., & Huang, J. (2023). Predicting Pulmonary Tuberculosis Incidence in China Using Baidu Search Index: An ARIMAX Model Approach. *Environmental Health and Preventive Medicine*, *28*, 68.

Yerima, D. P., Bawa, M. U., Adamu, H. Y., Maitoro, S. U., Idris, H. I., & Jibrin, Y. A. (2025). Modelling and Forecasting for the Dynamics of Financial Time Series of Exchange Rate Using ARIMAX and SARIMA Models. *International Journal of Innovative Mathematics, Statistics & Energy Policies*, *13*(3), 209–214.