DYNAMIC FACTOR MODEL ANALYSIS BASED ON PRINCIPAL COMPONENT EXTRACTED FACTOR FOR GRDP PREDICTION

(Undergraduate Thesis)

By

AGUSTINO SIMATUPANG
2217031157

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

ABSTRACT

DYNAMIC FACTOR MODEL ANALYSIS BASED ON PRINCIPAL COMPONENT EXTRACTED FACTOR FOR GRDP PREDICTION

By

Agustino Simatupang

Dynamic Factor Model (DFM) are multivariate forecasting models that represent the co-movements of observed variables through a small number of latent factors. DFMs effectively extract common information from highly correlated variables while reducing the influence of idiosyncratic noise, thereby improving forecasting accuracy. This study applies DFM to model and forecast the real and nominal Gross Regional Domestic Product (GRDP) of Lampung Province. The limitation of quarterly GRDP data was addressed through Denton temporal disaggregation to obtain monthly observations. Latent factors were extracted using Principal Component Analysis (PCA) and estimated within a state-space framework using the Kalman Filter under the Maximum Likelihood Estimation (MLE) framework. The AR(1), AR(2), and AR(3) specifications were evaluated using the Log-Likelihood, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), the Ljung–Box test, and the Mean Absolute Percentage Error (MAPE). The first principal component explained approximately 99% of the total data variation. The DFM-AR(3) model achieved the best performance, reconstructing the original quarterly GRDP series with MAPE values of 1.43% for real GRDP and 1.52% for nominal GRDP. These results demonstrate that DFM is an effective approach for modeling and forecasting GRDP using a limited number of highly correlated variables.

Keywords: Dynamic Factor Model, Principal Component Analysis, Denton Method, Forecasting, GRDP

ABSTRAK

ANALISIS DYNAMIC FACTOR MODEL BERDASARKAN FAKTOR EKSTRAKSI PRINCIPAL COMPONENT UNTUK PREDIKSI PDRB

By

Agustino Simatupang

Dynamic Factor Model (DFM) merupakan model peramalan multivariat yang merepresentasikan pergerakan bersama antarvariabel melalui sejumlah kecil faktor laten. DFM unggul dalam mengekstraksi informasi bersama dari variabel yang memiliki korelasi tinggi serta mengurangi pengaruh *idiosyncratic noise*, sehingga meningkatkan akurasi peramalan. Penelitian ini bertujuan menerapkan DFM untuk memodelkan dan meramalkan Produk Domestik Regional Bruto (PDRB) riil dan nominal Provinsi Lampung. Keterbatasan data PDRB triwulanan diatasi melalui disagregasi temporal menggunakan metode Denton untuk memperoleh data bulanan. Selanjutnya, faktor laten diekstraksi menggunakan *Principal Component Analysis* (PCA) dan diestimasi dalam kerangka *state-space* menggunakan *Kalman Filter* dengan *Maximum Likelihood Estimation* (MLE). Spesifikasi AR(1), AR(2), dan AR(3) dibandingkan menggunakan *Log-Likelihood*, *Akaike Information Criterion* (AIC), *Bayesian Information Criterion* (BIC), uji Ljung–Box, dan *Mean Absolute Percentage Error* (MAPE). Hasil menunjukkan bahwa komponen utama pertama menjelaskan sekitar 99% variasi data. Model DFM-AR(3) memberikan kinerja terbaik serta menghasilkan akurasi tertinggi dalam merekonstruksi data PDRB triwulanan, dengan nilai MAPE sebesar 1,43% untuk PDRB riil dan 1,52% untuk PDRB nominal. Temuan ini menunjukkan bahwa DFM merupakan pendekatan yang efektif untuk memodelkan dan meramalkan dinamika PDRB menggunakan variabel yang berkorelasi tinggi.

Keywords: *Dynamic Factor Model, Principal Component Analysis,* Metode *Denton,* Peramalan*,* PDRB

DYNAMIC FACTOR MODEL ANALYSIS BASED ON PRINCIPAL COMPONENT EXTRACTED FACTOR FOR GRDP PREDICTION

AGUSTINO SIMATUPANG

Undergraduate Thesis

As One of the Requirements to Obtain the Degree

BACHELOR OF MATHEMATICS

In the

Department of Mathematics

Faculty of Mathematics and Natural Sciences

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

Undergraduate Thesis Title	:	**DYNAMIC FACTOR MODEL ANALYSIS BASED ON PRINCIPAL COMPONENT EXTRACTED FACTOR FOR GRDP PREDICTION**
Student Name	:	*Agustino Simatupang*
Student ID	:	**2217031157**
Study Program	:	**Mathematics**
Faculty	:	**Faculty of Mathematics and Natural Sciences**

APPROVED BY

1. Advisory Committee

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

Dr. Edwin Russel, S.E., M.Sc.
NIP. 198406192024061001

2. Head of the Department of Mathematics

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 197403162005011001

VALIDATED BY

1. Examination Committee

Chair	:	**Dr. Khoirin Nisa, S.Si., M.Si.**	....
Secretary	:	**Dr. Edwin Russel, S.E., M.Sc.**	....
Non-Advisor Examiner	:	**Widiarti, S.Si, M.Si.**	....

2. Dean of the Faculty of Mathematics and Natural Sciences

KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS LAMPUNG
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

Dr. Eng. Heri Satria, S.Si. M.Si.
NIP. 197110012005011002

Date of Undergraduate Thesis Defense: **June 03, 2026**

UNDERGRADUATE THESIS STATEMENT

The undersigned below:

Name	:	**Agustino Simatupang**
Student ID	:	**2217031157**
Department	:	**Mathematics**
Title of Undergraduate Thesis	:	**DYNAMIC FACTOR MODEL ANALYSIS BASED ON PRINCIPAL COMPONENT EXTRACTED FACTOR FOR GRDP PREDICTION**

I hereby declare that this undergraduate thesis is my own original work. If it is later proven that this undergraduate thesis is a copy or made by someone else, I am willing to accept sanctions according to the applicable academic regulations.

Bandar Lampung, June 3, 2026

Author

10000

METERAI TEMPEL

D0518AOX063630005

Agustino Simatupang

AUTHOR BIOGRAPHY

"*Every dream is a seed waiting to be nurtured, and persistence is the water that turns it into reality*," these are the words that define the author. He is Agustino Simatupang, born in Panjang, on 16 August 2004. He completed his elementary education at Swadhipa Elementary School, continued his junior secondary education at State Junior High School 1 Natar, and his senior secondary education at Swadhipa Senior High School with a concentration in Mathematics and Natural Sciences. Driven by his interest in mathematics and analytical thinking, he pursued his undergraduate studies in the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Lampung, and graduated in 2026 with the degree of Bachelor of Mathematics.

Throughout his undergraduate years, Agustino actively participated in various academic, organizational, and research activities to broaden both his knowledge and practical skills. He completed his Field Work Practice at the General Bureau of the Regional Secretariat of Lampung Province and an Independent Internship at the Financial Services Authority (OJK) of Lampung Province, where he gained valuable experience in data management, statistical analysis, public services, and quantitative research.

Beyond his academic activities, Agustino has been actively involved in teaching since 2022 as a Private Mathematics Tutor for elementary, junior high, senior high school students, and Teaching Assistant for the Multivariate Analysis course at Universitas Lampung. He further strengthened his experience through research collaborations, organizational involvement, community service, and several interdisciplinary projects related to data analysis and mathematical applications.

His commitment to continuous learning is reflected in his participation in national and international academic programs, including the International Student Mobility (Inbound Programme) at the University of Malaya, Malaysia, the Student Entrepreneur Program (PMW) of Universitas Lampung, as well as several national scientific writing and innovation competitions. Beyond that, Agustino is deeply committed to social impact. As an active member of the Generations with a Planning (Genre) organization under the auspices of the National Population and Family Planning Board (BKKBN), he has played a pivotal role in facilitating and empowering adolescents across Indonesia, including in Lampung, Jakarta, and Bekasi. His dedication to social causes also led him to participate in major national events ADUJAKNAS in Tanjung Pinang and Bintan, among others, further refining his leadership and communication skills.

His undergraduate thesis, entitled "Dynamic Factor Model Analysis Based on Principal Component Extracted Factor for GRDP Prediction", reflects his interest in applying mathematical and statistical methods to solve real-world problems. He believes that learning is a lifelong journey and remains committed to continuously expanding his knowledge, embracing new challenges, and developing innovative data-driven solutions that create meaningful impacts for society.

MOTTO

*"I believe that **no dream is too big** for anyone. Every child deserves the chance to dream, and **no limitation** should define what is possible. With willingness, perseverance, and prayer, even **the impossible can become reality**."*

"Ora et Labora. Pray before your work and work faithfully toward what you have prayed for."

"Dream it, pursue it, and let it become your reality."

DEDICATION

To God Almighty, my deepest gratitude for His endless grace, mercy, and guidance. By His blessings alone, I was given the strength and perseverance to complete this undergraduate thesis. I also dedicated my thanks to:

My Beloved Mom

Thank you for your unconditional love, sacrifices, prayers, and unwavering support. Thank you for believing in my dreams, respecting my choices, and standing by my side through every challenge.

Advisors and Examiner

Thank you for your invaluable guidance, patience, constructive feedback, and encouragement. Your kindness, understanding, and insightful advice have not only strengthened this undergraduate thesis but also enriched my academic and personal growth.

My Family and Friends

Thank you for your continuous support, prayers, encouragement, and kindness. Your presence and willingness to help have made this journey much easier and more meaningful.

Everyone Who Has Been Part of This Journey

Thank you to everyone who has contributed, directly or indirectly, to the completion of this undergraduate thesis. Every form of support and encouragement has been sincerely appreciated.

ACKNOWLEDGMENTS

Praise and gratitude be to Almighty God for His abundant grace, mercy, and unwavering guidance, through which I was able to complete this undergraduate thesis entitled “Dynamic Factor Model Analysis Based on Principal Component Extracted Factor for GRDP Prediction” as one of the requirements for obtaining the degree of Bachelor of Mathematics at the Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Lampung.

The completion of this thesis would not have been possible without the invaluable support, encouragement, guidance, and prayers of many individuals. Therefore, I would like to express my deepest gratitude to:

1. Dr. Khoirin Nisa, S.Si., M.Si., my principal advisor, for her invaluable guidance, patience, encouragement, and continuous support throughout every stage of this research. Her insightful advice, thoughtful discussions, and genuine concern have inspired me not only to become a better researcher but also to grow as an individual.
2. Dr. Edwin Russel, S.E., M.Sc., my co-advisor, for his constructive suggestions, valuable insights, generous guidance, and continuous encouragement throughout the completion of this thesis.
3. Widiarti, S.Si., M.Si., my examiner, for her thoughtful comments, constructive criticism, and valuable recommendations that greatly improved the quality of this thesis.
4. Siti Laelatul Chasanah, S.Pd., M.Si., my academic advisor, for her guidance, advice, and encouragement throughout my academic journey at University of Lampung.
5. Dr. Aang Nuryaman, S.Si., M.Si., Head of the Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Lampung.

6. Dr. Eng. Heri Satria, S.Si., M.Si., Dean of the Faculty of Mathematics and Natural Sciences, University of Lampung.
7. All lecturers and administrative staff of the Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Lampung, for their dedication, knowledge, support, and invaluable experiences throughout my undergraduate studies.
8. My beloved mother, my greatest source of strength and inspiration. Thank you for believing in me even during the moments when I struggled to believe in myself. Your unconditional love, endless prayers, unwavering sacrifices, and constant encouragement have carried me through every challenge and made this achievement possible. This accomplishment belongs to you as much as it belongs to me.
9. My beloved family, especially my nephews, for filling my life with joy, laughter, and renewed spirit. Thank you for your love, prayers, encouragement, and for always reminding me that there is happiness beyond every challenge.
10. My dearest friends, who have become my second family throughout this journey. Thank you for being my safe place to share stories, for your willingness to help without hesitation, for celebrating every achievement, comforting every disappointment, and continuously encouraging me to pursue every dream and aspiration.
11. Everyone who has supported, encouraged, and prayed for me throughout this journey, although I am unable to mention each of you individually. Every kindness, no matter how small, has contributed to the completion of this thesis and will always be remembered with sincere gratitude.
12. Finally, to myself. Thank you for choosing to keep moving forward even when the journey felt overwhelming. Thank you for holding on when giving up seemed easier, for choosing faith over fear when hope appeared distant, and for continuing to believe that every effort would eventually find its purpose. Thank you for embracing every challenge, every setback, every unanswered question, and every lesson along the way. Most importantly, thank you for remaining true to yourself and for having the courage to pursue the dreams that once existed

only as words written in the pages of hope. This thesis stands as a reminder that perseverance, faith, and consistency can transform aspirations into reality.

I sincerely hope that this undergraduate thesis will contribute to the advancement of knowledge, particularly in the fields of applied mathematics, statistics, econometrics, and time series forecasting, and provide valuable references for future research.

Bandar Lampung, June 3, 2026

Agustino Simatupang

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table Page

I. INTRODUCTION

1.1 Background and Problems Statement

The Dynamic Factor Model (DFM) is a robust statistical model that assumes the driving forces behind the co-movement of various economic variables. Due to its ability to detect short-term trends and cyclical patterns, this model is commonly used in nowcasting and forecasting indicators such as Gross Regional Domestic Product (GRDP) (Kuck & Schweikert, 2021). Generally, the methodological approach to applying DFM involves three main approaches: Maximum Likelihood Estimation (MLE) using state-space forms, followed by the Bayesian Approach for high–dimensional data and the Two-Step Estimations suggested in Stock & Watson (2002). By utilizing Principal Component Analysis (PCA) and the Kalman Filter, this ultimate technique delivers the necessary computational power to obtain economic signals from intricate regional datasets.

In addition to frequency constraints, the often vast and complex structure of macroeconomic datasets demands that these models be applied by researchers. PCA is an essential simplification method that reduces data complexity by reducing multiple interrelated variables into a few crucial components that represent most of the data variation (Juanhua & Donglin, 2025). To extract the key factors for dynamic modeling, PCA is often used as the foundational stage of DFM analysis by mitigating the "curse of dimensionality" (Chen & Fan, 2023).

Measuring the economic growth rate of a region is possible by using production, expenditure and income methods, with GRDP being one of its primary indicators along with GDP. From the added value of the production sector to consumption and investment in the community, these three approaches offer a comprehensive view of economic activity. Inflation, interest rates, investment, exports-imports and household consumption are complexly interrelated and influence the dynamics of GRDP (Supriyatna et al., 2024).

However, the primary challenge in regional economic analysis, particularly in Lampung Province, lies in the limited frequency of official data. While GRDP is a vital indicator of economic health, it is only published on a quarterly basis. This low frequency often obscures short-term economic shocks and rapid cyclical changes that occur month-to-month. To address this information gap, temporal disaggregation using the Proportional Denton Method is essential (Fonzo & Marini, 2012). By utilizing a high-frequency indicator, this method ensures that monthly estimates remain mathematically consistent with official quarterly figures, providing a more granular dataset for precise analysis.

Traditional econometric frameworks such as Ordinary Least Squares (OLS), AutoRegressive Integrated Moving Average with Explanatory Variables (ARIMAX), and Vector Autoregression (VAR) are not always effective in solving high-dimensional problems. The instability of coefficient estimates, in addition to the inflation factor (VIF) being inflating and multicollinearity in these methods making individual predictor effects difficult to isolate statistically (Kim, 2019). In particular, the limitations of standard VAR models in terms of dimensionality make them problematic (Chan et al., 2022), due to which an extensive parameter space often hinders performance. Furthermore, even when high-frequency data is available, macroeconomic indicators such as Real and Nominal GRDP often treated as isolated series, ignoring the fact that they are manifestations of the same underlying economic engine.

Prior research has increasingly embraced the Stock & Watson (2002) framework, which integrates PCA and DFM to reveal the latent structure of interrelated economic data. Instead of forecasting variables in isolation, this study uses PCA to extract a single latent factor as a common unseen driver representing the core economic pulse of Lampung Province. By filtering out idiosyncratic noise, the Two-Step Estimation utilizing both PCA and the Kalman Filter ensures more stable, synchronized, and robust forecasts. This methodological synergy allows for the collective capture of 'co-movement' across variables while simultaneously addressing the inherent challenges of high-dimensional data structures.

In the context of Indonesia, study by Tarsidin et al. (2018) and Ringo & Monika (2022) emphasize that although DFM effectively manages regional GRDP forecasting, this method must be adjusted to the availability and volatility of regional data. Ultimately, PCA addresses multicollinearity, enabling DFM to capture dynamic relationships in GRDP (O’Keeffe & Petrova, 2025). The integration of these elements significantly improves forecasting accuracy, especially when dealing with high-dimensional data within the limitations of time series.

1.2 Research Questions

1. How effective is Principal Component Analysis (PCA) in representing the joint dynamics between Real and Nominal GRDP in Lampung Province through selected principal components?
2. Is the Dynamic Factor Model (DFM) still able to produce accurate and stable forecast estimates for GRDP, even when built on a dataset with limited dimensions (low-dimensional)?

1.3 Research Objectives

1. Analyzing the ability of PCA to extract crucial information from real and nominal GRDP variables to mitigate data redundancy due to high multicollinearity.
2. Developing and evaluating the performance of GRDP forecasting models in capturing economic signals using the integration of PCA and DFM under limited dimensional data conditions.

1.4 Research Benefits

This study provides an empirical tool for the Lampung Provincial Government and stakeholders in forecasting GRDP values through a model that is resistant to data volatility and idiosyncratic disturbances. For researchers, this study enriches econometric literature by providing a methodological reference regarding the application of PCA and DFM integration on low-dimensional data or in regional economic analysis with limited indicators.

II. LITERATURE REVIEW

2.1 Time Series and Forecasting

Time series analysis is a statistical method used to understand data collected sequentially over time. Its main purpose is to identify historical patterns and predict future values. The main dynamics observed in time series data include trends (general direction), seasonality (regular fluctuations), cycles (longer and irregular fluctuations), and random variations, the latter of which is particularly important to address as it represents noise that can obscure the accuracy of predictions (Kyo et al., 2024). This analysis is vital in economics for forecasting critical indicators such as inflation, interest rates, and commodity prices, which are routinely recorded to anticipate future market behavior (Indah et al., 2023).

While univariate methods focus on a single variable, multivariate time series analysis considers multiple variables to model the dynamic relationships between them over time, which is essential for improving forecasting performance. This approach is important in economics, where structural relationships between interacting indicators such as inflation, consumption and investment, are analyzed to generate more comprehensive and accurate forecasts than those produced by isolated univariate models (Iftikhar et al., 2025).

A fundamental approach in multivariate time series forecasting is the Vector Autoregression (VAR) system, which models the dynamic interactions among a group of variables by allowing each variable to depend on its own past values as well as the past values of all other variables in the system.

This approach is formally expressed through the following general equation:

$$\mathbf{X_t} = \boldsymbol{\alpha} + \boldsymbol{\Sigma\Phi_1 X_{t-i}} + \boldsymbol{\epsilon_t} \tag{2.1}$$

where:

$\mathbf{X}_t$: the 2 × 1 vectors of observed variables,

$\boldsymbol{\alpha}$: intercept vector,

$\boldsymbol{\Phi_i}$: parameter matrix 2 × 2 for each lag i = 1, 2, …, p,

$\mathbf{X}_{t-i}$: the vector of observed variables at lag-i, and

$\boldsymbol{\epsilon}_t$: error vector at time t.

However, as the number of variables in VAR increases, structural complexity often leads to overparameterization, thereby reducing forecasting efficiency. Dynamic Factor Models (DFM) have emerged as a superior solution by modeling the co-movements of multiple variables through a small set of dynamic latent factors, making them highly effective for high-dimensional forecasting. Before constructing a DFM, Principal Component Analysis (PCA) is often performed to reduce dimensions, extracting underlying patterns into principal components that serve as more detailed inputs for the forecasting model. The strategic combination of multivariate time series analysis, DFM, and PCA allows researchers to capture complex dynamics more efficiently, significantly improving the accuracy of forecasting for macroeconomic aggregates such as Gross Regional Domestic Product (GRDP) (O'Keeffe & Petrova, 2025).

2.1.1 Stationarity

Stationarity is an important condition in time series analysis. A time series is said to be stationary if it has a constant mean, variance, and intertemporal covariance (Silva et al., 2021).

A commonly used test is the Augmented Dickey-Fuller (ADF) test, which is formulated as follows:

$$\Delta X_t = \alpha + \beta + \delta X_{t-1} + \sum_{i=1}^{m} \gamma_i \Delta X_{t-i} + \varepsilon_t \tag{2.2}$$

where:

ΔX_t : first derivative of X,

α : constant value,

β : trend regression coefficient,

δ : lag regression coefficient,

γ_i : lag regression coefficient of X,

ε : error,

m : numbers of lag, and

t : time.

The ADF test is conducted with the null hypothesis (H_0) stating that the data has a unit root, indicating non-stationarity, while the alternative hypothesis (H_1) suggests that the data is stationary. The decision region is determined by comparing the p-value with the significance level (α = 0.05); if the *p-value* is less than 0.05, H_0 is rejected, confirming the data is stationary. Alternatively, stationarity is achieved if the ADF test statistic is more negative than the MacKinnon critical values, allowing the data to be used for further dynamic factor modeling.

2.1.2 Differencing

In time series analysis, differencing is defined as an important data preprocessing technique for transforming non-stationary data into stationary data. Non-stationary data, whose mean or variance changes over time due to trends or seasonality, can cause invalid forecasting problems. The differencing process works by calculating the difference between the current period's observation and the previous period's

observation, thereby effectively eliminating linear trends and stabilizing the data mean to make it suitable for linear econometric models such as DFM.
Mathematically, the first-order differencing formula is expressed as follows:

$$\Delta X_t = X_t - X_{t-1} \tag{2.3}$$

where:

ΔX_t : the differenced value at time *t*, representing the change from the previous period

X_t : actual data value at the current period *t*, and

X_{t-1} : actual data value at the prior period $(t - 1)$.

If the time series data still shows a strong trend and remains non-stationary even after applying first-order differentiation, second-order differentiation techniques must be applied. This technique is defined as a differentiation process applied to the results of first-order differentiation, not to the actual data itself. Theoretically, second-order differentiation is very effective for eliminating quadratic trends and stabilizing complex data variance, ensuring that the data meets the strict stationarity assumptions required for PCA and DFM modeling (Ryan et al., 2025).
Mathematically, the second-order differencing formula is expressed as the difference between two consecutive first-order differences:

$$\Delta^2 X_t = \Delta X_t - \Delta X_{t-1} \qquad (2.4).$$

In a technically expanded form, the above formula can be further developed based on the actual data to facilitate direct computation:

$$\Delta^2 X_t = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) \tag{2.5}$$

where:

$\Delta^2 X_t$: second-order differenced value at time t,

ΔX_t : first-order differencing result at time t, and

X_{t-2} : actual data value at the previous two periods $(t - 2)$.

2.2 Temporal Disaggregation: The Denton Method

Temporal disaggregation is a statistical technique used to convert low-frequency time series data into high-frequency data. In this study, the Denton method is employed to transform the GRDP data of Lampung Province from quarterly to monthly frequency.

The Denton method, specifically the Proportional Denton variant, aims to minimize the relative movement between the estimated series (X_t) and an indicator variable (I_t), while maintaining consistency with the original low-frequency data (O_T). Mathematically, this is achieved by minimizing the squared difference of the ratios between the estimated data and the indicator through the following objective function:

$$\min_{X_t} \sum_{t=2}^{n} \left[\frac{X_t}{I_t} - \frac{X_{t-1}}{I_{t-1}}\right]^2 \quad (2.6),$$

subject to the constraint that the sum of the estimated data within a low-frequency period must equal the original observed value:

$$\sum_{t \in T} X_t = O_T \quad (2.7).$$

The primary advantage of this method over simple interpolation techniques (such as linear interpolation) is its ability to avoid the "step problem" unnatural data jumps at transition points between periods while preserving the economic dynamics represented by the indicator variable (Fonzo & Marini, 2012).

2.3 Dynamic Factor Model

The Dynamic Factor Model is a statistical model used to summarize information from a large number of time series variables into several latent (unobserved) factors that follow a dynamic process. Its main purpose is to simplify high-dimensional data into a concise representation that remains informative. DFM assumes that most of the co-movement in the data can be explained by a small number of unobserved latent factors, which facilitates analysis and forecasting (Hu & Yao, 2022). The general DFM equation can be written as follows:

$$\mathbf{X_t} = \mathbf{\Lambda F}_t + \mathbf{e}_t \tag{2.8}$$

and

$$\mathbf{F}_t = \mathbf{\Sigma \Phi F}_{t-i} + \mathbf{u}_t \tag{2.9}$$

where:

$\mathbf{X}_t$: observation vector at time t,

$\mathbf{F}_t$: linear combination of principal components as regressors,

$\mathbf{\Lambda}$: loading matrix from factor to observation,

$\mathbf{e}_t$: residual,

$\mathbf{\Phi}$: coefficient matrix, and

$\mathbf{u}_t$: disturbance from the factor process.

2.4 Model Estimation

According to Stock & Watson (2002), the DFM can be estimated using three main approaches: the Maximum Likelihood (ML) method, Bayesian methods, or the Two-Step Estimation approach. In the context of high-dimensional macroeconomic data, the Two-Step Estimation method is the most computationally efficient and widely utilized choice as it cleanly separates the process of factor identification from the modeling of temporal dynamics.

The Two-Step Estimation method is executed through two essential and interdependent stages, the extraction of static factors using PCA and the modeling of factor dynamics using the Kalman Filter.

2.4.1 Stage One: Static Factor Extraction using PCA

Principal Component Analysis is a dimension reduction technique used to identify and extract the main components from correlated data. In the context of multivariate time series, PCA helps summarize information from many variables into several principal components that can be used as inputs in DFM models, thereby simplifying data complexity and improving model efficiency. This method is widely used in economics, particularly in macroeconomic and financial data analysis, as it can overcome the significant challenges of analyzing multivariate data (Frost, 2022).

PCA is performed using the following sequential steps:

1. Data Standardization

   Scaling variables to an equal magnitude is a crucial step, particulary prior to implementing PCA. PCA is very sensitive to data scale because its purpose is to maximize variance, so if variables have different scales, variables with larger scales will dominate the first principal component and ignore the contribution of other variables that may be equally important. Therefore, data standardization is an important step to ensure that each variable contributes proportionally to the principal components, resulting in a more accurate representation of the underlying data structure (Forkman et al., 2019).

   In general, there are several data standardization methods that can be used, namely Z-Score standardization (Standard Scaling), Min-Max Scaling, Robust Scaling (Median and IQR), Unit Vector Scaling, and other methods. The method commonly used for PCA and DFM analysis is z-score standardization. Data standardization using z-scores can be calculated using the following formula in Eq. (2.10).

$$Z = \frac{X - \bar{X}}{s} \tag{2.10}$$

where:

Z : standardized value,

X : actual value (observation i),

$\bar{X}$: average of all data, and

s : standard deviation of data.

2. Constructing the Covariance or Correlations Matrix

The covariance matrix (**Σ**) and the correlation matrix (**R**) constitute the fundamental mathematical basis for eigendecomposition in PCA. The covariance matrix is a symmetric square matrix $p \times p$ that encapsulates the variance of variables along its main diagonal and the covariance between variable pairs in its off-diagonal elements. Generally, the Covariance Matrix **Σ** is computed from mean-centered data $\mathbf{X}_c = \mathbf{X} - \bar{\mathbf{X}}$, where $\bar{\mathbf{X}}$ denotes the mean matrix, utilizing the following matrix formula:

$$\boldsymbol{\Sigma} = \frac{1}{n-1}\mathbf{X}_c^{\mathrm{T}}\mathbf{X}_c \tag{2.11}$$

The covariance matrix in Eq. (2.11) can be illustrated using a 2×2 bivariate covariance matrix, where each element represents either the variance of an individual variable or the covariance between two variables. In this case, the off-diagonal element σ_{12}, which denotes the covariance between variables x_1 and x_2, is computed as:

$$\sigma_{12} = Cov(x_1, x_2) = \frac{1}{n-1}\sum_{i=1}^{n}(x_{i,1} - \bar{x}_1)\,(x_{i,2} - \bar{x}_2) \tag{2.12}$$

where σ_{12} represents the covariance between variables x_1 and x_2. Likewise, if $x_1 = x_2$, the element represents the variance of variable $x_1(Var(x_1))$. The matrix **Σ** is typically employed when all variables share identical units of measurement. However, in instances of heterogeneous data, such as economic indicators where Gross Regional Domestic Product (GRDP) is measured in trillions while inflation is expressed in percentages the use of **Σ** may cause the

PCA results to be disproportionately dominated by variables with the largest scales (Dien et al., 2005).

Consequently, within the context of DFM, the correlation matrix **R** may be utilized as a robust alternative to the covariance matrix. Theoretically, the Correlation Matrix is equivalent to a covariance matrix computed from standardized data **Z**, where each variable possesses a mean of zero and a standard deviation of one.

The correlation coefficient ρ is defined as the covariance divided by the product of the standard deviations of the two variables $\sigma_1\sigma_2$:

$$\rho = Corr\ (x_1, x_2) = \frac{Cov(x_1, x_2)}{\sigma_1\sigma_2} \qquad (2.13).$$

By performing initial data standardization, the resulting correlation matrix ***R*** exhibits values of **1** on the main diagonal (representing self-correlation) and correlation coefficients ρ_j on the off-diagonal elements. The utilization of ***R*** effectively mitigates scale bias and ensures that all variables contribute to the principal components based on the strength of their relative linear relationships rather than the magnitude of their units. Thus, the correlation matrix serves as a methodologically sound basis for eigendecomposition in PCA when dealing with input variables of diverse magnitudes (Choi & Yang, 2022).

$$\mathbf{\Sigma} = \begin{pmatrix} Var(x_1) & Cov(x_1, x_2) \\ Cov(x_2, x_1) & Var(x_2) \end{pmatrix} \xrightarrow{\text{standardization}} \mathbf{R} = \begin{pmatrix} 1 & Corr(x_1, x_2) \\ Corr(x_2, x_1) & 1 \end{pmatrix} \ (2.14).$$

3. Calculating of Eigenvalues and Eigenvectors

   The core computational process of PCA involves the decomposition of the covariance or correlation matrix to extract the underlying variance structure of the data through the computation of eigenvalues and eigenvectors. Theoretically, eigenvalues λ represent scalars that quantify the magnitude of data variance explained by their corresponding eigenvectors; a larger eigenvalue indicates a more significant contribution of the respective principal component in preserving the global information of the dataset (Frost, 2022).

Mathematically, the eigenvalue λ holds a direct relationship with the variance of the principal component. The PCA process is fundamentally an optimization problem designed to find the vector $\mathbf{a}$ that maximizes the variance of the projected data ($Var(\mathbf{Xa}) = \mathbf{a}^T \sum \mathbf{a}$), subject to the constraint that $\mathbf{a}$ is a unit vector ($\mathbf{a}^T \mathbf{a} = 1$). Solving this constrained optimization problem using the Lagrange Multiplier method naturally yields the characteristic equation, $\sum \boldsymbol{a} = \lambda \mathbf{a}$). Substituting this result back into the maximized variance equation, $Var(\mathbf{Xa}) = \mathbf{a}^T(\lambda \mathbf{a}) = \lambda(\mathbf{a}^T \mathbf{a})$, proves that since $\mathbf{a}^T \mathbf{a} = 1$, the maximum variance explained by the component is exactly equal to its eigenvalue λ. These values are derived by solving the characteristic equation of the matrix determinant as follows:

$$\det(\boldsymbol{\Sigma} - \lambda \mathbf{I}) = 0 \tag{2.15}$$

where $\boldsymbol{\Sigma}$ denotes the covariance matrix, λ represents the target eigenvalue, and $\mathbf{I}$ is the identity matrix of identical dimensions. Upon determining the set of scalars λ, the subsequent step involves identifying the eigenvectors $\mathbf{v}$, defined as non-zero directional vectors within the multidimensional space that establish the new axes (principal components) for data projection. These vectors are computed by substituting the obtained eigenvalues back into the homogeneous linear equation system:

$$(\boldsymbol{\Sigma} - \lambda \mathbf{I})\mathbf{v} = \mathbf{0} \tag{2.16}$$

where $\mathbf{v}$ denotes a non-zero column vector and $\mathbf{0}$ is zero vector. Through this transformation, eigenvectors function as linear projection axes for the actual variables, facilitating efficient dimensionality reduction without compromising the dominant patterns or the intrinsic significance of the actual data structure (Frost, 2022).

4. Determining the Number of Principal Components

   The crucial stage in PCA is determining the optimal number of components r that must be retained. This process is essential because only these r selected static factors will be carried forward as input into the dynamic Kalman Filter model. Researchers can select the optimal value for r by considering several criteria, such as the Kaiser-Guttman Criterion, which recommends the retention of components with an eigenvalue λ greater than one $\lambda > 1$, or the Cumulative Proportion of Variance criterion, which requires the selected components to cumulatively explain a significant percentage of the total data variance (e.g., $\geq$ 80%). Another commonly used approach is the Scree Plot Analysis, which visually identifies the elbow or inflection point on the eigenvalue graph, marking the boundary between significant components and noise. However, the primary drawback of the *Scree Plot* is its subjective nature, as the determination of the *elbow* can often be ambiguous and vary among researchers (Qin & Dunia, 2000).

Once the value of r has been decided based on the chosen criteria, the final step is performing the projection to obtain the final estimation of the static factors $\mathbf{F}_t$. This projection is accomplished by multiplying the standardized data matrix $\mathbf{X}_\mathrm{t}$ by the loading matrix $\mathbf{\Lambda}$ composed of the r selected eigenvectors, according to the following PCA equation:

$$\mathbf{F}_\mathrm{t} = \mathbf{X}_\mathbf{t}\mathbf{\Lambda} \tag{2.17}$$

where $\mathbf{F}_\mathrm{t}$ is the vector of selected static factors $T \times r$, $\mathbf{X}_\mathrm{t}$ is the standardized indicator variables matrix $T \times N$, and $\mathbf{\Lambda}$ is the loading matrix consisting of the r eigenvectors corresponding to the r largest eigenvalues $N \times r$. These selected static factors $\mathbf{F}_\mathrm{t}$ are subsequently used in the second stage of DFM estimation (Song & Shin, 2018).

5. Forming Principal Component

PCA is most accurately understood as a linear decomposition asserting that the observed data **X** is reconstructed from two primary sources: the common factors **F**, which account for the shared variance across all variables, and the idiosyncratic error $\boldsymbol{e}$, which represents unique variance or noise specific to each variable. Structurally, the data matrix **X** is reconstructed by multiplying the factor matrix **F** by the transpose of the loading matrix $\mathbf{\Lambda}^\mathrm{T}$. These loadings serve as weights defining the relationship between the latent factors and the observed variables (Lever et al., 2017). The goal of this decomposition is dimensionality reduction, yielding a smaller number of static factors $\mathbf{F}$ that capture the dominant patterns for subsequent input into the dynamic model.

The general matrix form used to represent this linear decomposition is:

$$\mathbf{X}_{\mathrm{N\times P}} = \mathbf{F}_{\mathrm{N\times K}}\, \mathbf{\Lambda}^{\mathrm{T}}_{\mathrm{K\times P}} + \mathbf{e}_{\mathrm{N\times P}} \tag{2.18}$$

where $\mathbf{X}$ is the Observed Data Matrix, $\mathbf{F}$ is the extracted Static Factor Matrix, $\mathbf{\Lambda}$ is the Loading Matrix, and $\mathbf{e}$ is the Idiosyncratic Error Matrix (Cynthia & Onyinyechi, 2019).

2.4.2 Stage Two: Dynamic Factor Modeling using Kalman Filter

Following the extraction of static factors using PCA, the DFM Two-Step Estimation proceeds by modeling the temporal dynamics of these factors using the Kalman Filter (KF). The KF acts as an optimal recursive algorithm designed to estimate unobserved state variables in this case, the factors. By processing noisy data sequentially through a State-Space system, the algorithm separates the true signal (the factors) from the noise (error) at every time step to ensure accurate estimation (Doz et al., 2011).

A. The State-Space Structure

   The DFM system estimated by the Kalman Filter is defined by two fundamental equations: the Measurement Equation and the Transition Equation (Kulikova et al., 2020).

   1. Measurement Equation: This equation links the observed variables $\mathbf{X}_{\mathrm{t}}$ to the extracted static factors $\mathbf{F}_{\mathrm{t}}$ and the specific error $\mathbf{e}_{\mathrm{t}}$. This is essentially the output of the PCA embedded back into the model in Eq. (2.8) which explains the observed variables $\mathbf{X}_{\mathrm{t}}$ at time t as a function of the unobserved static factors $\mathbf{F}_{\mathrm{t}}$ scaled by the loadings matrix $\mathbf{\Lambda}$ plus a specific error $\mathbf{e}_{\mathrm{t}}$.

2. Transition Equation: This equation defines the temporal evolution of the factors, which is typically modeled as a Vector Autoregression (VAR) of order p. This is the rule governing how the state variable changes over time in Eq. (2.9) which models the evolution of the factors $\mathbf{F}_t$ over time using a VAR of order p, where $\mathbf{F}_{t-i}$ are the factors at past time steps $t - i$ weighted by the autoregressive matrix $\boldsymbol{\Phi}_i$, and u_t is the error.

B. Optimality and Function

- Recursive Function system: The KF operates in a continuous cycle of prediction and updating. It forecasts the factor status $\mathbf{F}_{t|t-1}$, then uses the actual observation $\mathbf{X}_t$ to refine this forecast, yielding the optimal estimate $\mathbf{F}_{t|t}$ (Zhang & Ding, 2020).
- Optimality Best Linear Unbiased Estimator (BLUE): The KF is proven to be the BLUE. This proof guarantees that the resulting factor estimates have the minimal estimation error covariance matrix $\mathbf{P}_t$, provided that both system error $\mathbf{u}_t$ and measurement error $\mathbf{e}_t$ are normally distributed. This ensures the accuracy and stability of the DFM model (Kulikova et al., 2020).

2.5 Model Evaluation

The objective of model evaluation is to measure the accuracy and performance of the constructed DFM in projecting GRDP values for Lampung. Evaluation is conducted by comparing the forecasted values generated by the model $\widehat{X_t}$ with the true actual values $X_t$within the out-of-sample period (data that was not utilized for model estimation). The following indicators are commonly employed in econometric studies to quantify forecasting errors:

2.5.1 Goodness of Fit Test

DFM models are often evaluated using information criteria to balance data fit with model complexity (the number of estimated parameters). The evaluation begins with the Log-Likelihood function. The Log-Likelihood function for the state-space model is calculated using the innovations derived from the Kalman Filter, which serves as the foundation for both AIC and BIC.

- Log-Likelihood Function

$$\ln L(\theta) = -\frac{nT}{2}\ln(2\pi) - \frac{1}{2}\ \sum_{t=1}^{T}(\ln|\mathbf{F}_{\mathrm{t}}| + \mathbf{v}_{\mathrm{t}}'\mathbf{F}_{\mathrm{t}}^{-1}\mathbf{v}_{\mathrm{t}}) \qquad (2.19),$$

- Akaike Information Criterion (AIC)

$$\mathrm{AIC} = -2\ ln\ (L) + 2k \qquad (2.20),$$

- Bayesian Information Criterion (BIC)

$$\mathrm{BIC} = -2\ ln\ (L) + k\ ln\ (T) \qquad (2.21).$$

In these formulations, $\ln L\ (\theta)$ represents the maximum Log-Likelihood value derived from the Kalman Filter innovations, where $\mathbf{v}_{\mathrm{t}}$ is the innovation (prediction error) at time t and $\mathbf{F}_{\mathrm{t}}$ is its corresponding innovation covariance matrix. The complexity of the model is accounted for by k, which denotes the number of estimated parameters, while n and T signify the number of observed variables and the total number of observations, respectively. While both AIC and BIC aim for the smallest value to indicate a superior fit, BIC imposes a larger penalty for model complexity, especially in larger samples (Brewer et al., 2016).

2.5.2 Forecasting Performance

Several indicators are utilized to quantitatively measure the accuracy and reliability of a forecasting model's performance, which are:

- Root Mean Square Error (RMSE)

As an indicator that measures the square root of the average squared differences between the predicted and actual values, the Root Mean Square Error (RMSE) inherently imposes a higher penalty or weight on larger forecasting errors. A smaller RMSE value indicates superior forecasting capability (Karunasingha, 2022). The RMSE is calculated according to the formula:

$$\text{RMSE} = \sqrt{\frac{1}{h}\sum_{t=T+1}^{T+h}\left(X_t - \hat{X}_t\right)^2} \tag{2.22}$$

where:

h : forecast horizon (the number of periods being forecasted),

X_t : actual value (observation i), and

$\hat{X}_t$: forecasted (predicted) value at time t.

- Mean Absolute Error (MAPE)

Mean Absolute Percentage Error (MAPE) is a statistical metric used to evaluate the performance and accuracy of a forecasting model by measuring the average percentage of absolute errors between actual and predicted values. The primary advantage of MAPE lies in its highly intuitive interpretation, as it presents the error rate in percentage form, making it easier for stakeholders to understand the model's deviation regardless of the original data scale (Jeong, 2024).

Mathematically, this forecasting accuracy is calculated using the following equation:

$$\text{MAPE} = \frac{100\%}{n} \sum_{t-1}^{n} \left| \frac{X_t - \hat{X}_t}{X_t} \right| \tag{2.23}$$

where:

n : total number of periods of observed data points,

X_t : actual value at time t,

$\hat{X}_t$: forecast value at time t, and

$|X_t - \hat{X}_t|$: the direct difference between actual and forecast values.

2.5.3 Diagnostic Evaluation

- Ljung-Box Test

Diagnostic evaluation focuses on testing the statistical properties of the model's residuals (errors) to ensure that the underlying assumptions of the *Kalman Filter* (such as lack of serial correlation and normality) are met. Failure in diagnostics implies that factor inference might be invalid. To test whether the model residuals are *white noise* (possessing no serial correlation), the Ljung-Box Test is commonly employed. The Ljung-Box Q statistic is defined by the following equation:

$$Q = n(n+2) \sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{n-k} \tag{2.24}$$

where n is the number of observations, $\hat{\rho}_k$ is the residual autocorrelation coefficient at lag k, and h is the number of lags tested. If the resulting Q statistic is smaller than the X^2 critical value, the null hypothesis (that the residuals have no autocorrelation) is not rejected, indicating the DFM has adequately captured the time-series information (Eze et al., 2020).

The hypotheses tested in this procedure are formulated as follows:

- $H_0 : \rho_1 = \rho_2 = \cdots = \rho_h = 0$

  (The data are independently distributed; the residuals are white noise).

- $H_1 : \exists!\ \rho_k \neq 0$

  (The data are not independently distributed; the residuals exhibit serial correlation).

The test decision is determined by comparing the Qstatistic against the critical value from the Chi-Square (χ^2) distribution:

1. If $Q \leq \chi^2_{(1-\alpha,h)}$ or *p-value* > α : The null hypothesis H_0 is not rejected. This indicates that the residuals possess no serial correlation, suggesting that the DFM has adequately captured the time-series information.
2. If $Q > \chi^2_{(1-\alpha,h)}$ or *p-value* < α : The null hypothesis H_0 is rejected. This indicates that the residuals still contain autocorrelation patterns, implying that the model requires re-evaluation (e.g., by increasing the number of lags or including additional factors).

- Jarque-Bera Test

Furthermore, the Jarque-Bera Test is used to verify the normality of the residuals:

$$JB = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4}\right) \tag{2.25}$$

where S is the sample skewness and K is the sample kurtosis. The hypotheses and decision rules are:

H_0 : The residuals follow a normal distribution ($S = 0, K = 3$).

H_1 : The residuals do not follow a normal distribution.

The decision rule stipulates that H_0 is maintained if the *p-value* is greater than 0.05, confirming that the residuals exhibit a skewness of zero and a kurtosis of three. If the *p-value* is less than 0.05, the null hypothesis is rejected, suggesting that the model's errors are non-normally distributed, which may stem from outliers or structural breaks in the data (Aslam et al., 2021).

2.6 Gross Regional Domestic Product

Gross Regional Domestic Product (GRDP) is a measure used to describe the total value of final goods and services produced by all economic activities in a region within a certain period of time. GRDP reflects the level of economic activity and is an important indicator in assessing regional welfare and economic growth (Rahmi et al., 2025). In general, GRDP is obtained by adding up the gross value added generated by various economic sectors, such as agriculture, manufacturing, trade, construction, transportation, and services. GRDP data is widely used in economic research and development planning because it provides a comprehensive picture of the economic structure, productivity, and competitiveness of a region. In addition, GRDP is also often used to make predictions or projections of economic growth, as changes in its value reflect the dynamics of interrelated economic sectors.

Several variables that affect GRDP include the level of investment that drives business expansion, the number of workers that indicates production capacity, government spending that reflects fiscal policy and public expenditure, household consumption that is the main driver of domestic demand, the value of exports and imports that reflects the openness of the regional economy, and the quality of infrastructure that supports distribution and production efficiency (Aprilianita et al., 2024). In the context of Lampung Province, these variables have a significant influence given that the region's economic structure is still heavily dependent on the agricultural sector, manufacturing, and interregional trade.

III. RESEARCH METHODOLOGY

3.1 Research Time and Place

This research was conducted at the Mathematics Department Building, Faculty of Mathematics and Natural Sciences, University of Lampung from February to April of the 2025/2026 academic year.

3.2 Research Data

This study uses secondary time series data sourced from the official website of the Central Statistics Agency (BPS), covering the period 2010–2024, with a total of 60 quarterly data. The variables collected are Real GRDP and Nominal GRDP. Data shown in Appendix 1. And the monthly datasets, derived by interpolating the original quarterly data via the Denton Method, are presented in Appendix 2.

3.3 Research Methods

Computational analysis in this study was performed using Google Colaboratory (Google Colab) with the Python programming language. This platform was used due to its efficiency in processing data and its robust support for advanced econometric libraries required for factor extraction and state space modeling.

The steps taken in this study are as follows:

1. Descriptive analysis, performed by visualizing time series graphs to identify and observe data trends (such as growth or cyclical patterns).
2. Data preprocessing involves checking for missing values, duplicate data, and outliers to ensure data integrity and completeness.
3. Stationarity Testing using the Augmented Dickey-Fuller (ADF) Test, which is calculated based on the regression equation provided in Eq. (2.2). If the data is found to be stationary, the process continues to Static Factor Extraction. But if the data is found to be non-stationary, differentiation must be performed to achieve stationarity using the first-order differentiation formula, shown in Eq. (2.3).
4. Static factor extraction uses Principal Component Analysis (PCA) through data standardization, covariance or correlation matrix formation, calculation of eigenvalues and eigenvectors, determination of the number of principal components, and factors score formation.
5. Dynamic Factor Modeling (DFM), implemented through a State Space approach, which consists of Measurement Equations and Transition Equations. The Kalman Filter algorithm is then applied to predict, update, and smooth latent factors recursively, ensuring that the results achieve the Best Linear Unbiased Estimator (BLUE**)** status with minimal variance.
6. Evaluate the model quality through three assessment stages: first, evaluating model fit using Log-Likelihood as defined in Eq. (2.19), Akaike Information Criterion (AIC) in Eq. (2.20), and Bayesian Information Criterion (BIC) in Eq. (2.21); second, determining forecasting performance through error metrics such as Root Mean Squared Error (RMSE) in Eq. (2.22) and Mean Absolute Percentage Error (MAPE) in Eq. (2.23); and third, conducting a diagnostic evaluation using the Ljung-Box test in Eq. (2.24) to detect residual autocorrelation and the Jarque-Bera test in Eq. (2.25**)** to verify residual normality, ensuring the stability and statistical validity of the model.

The research flow is presented in Figure 1.

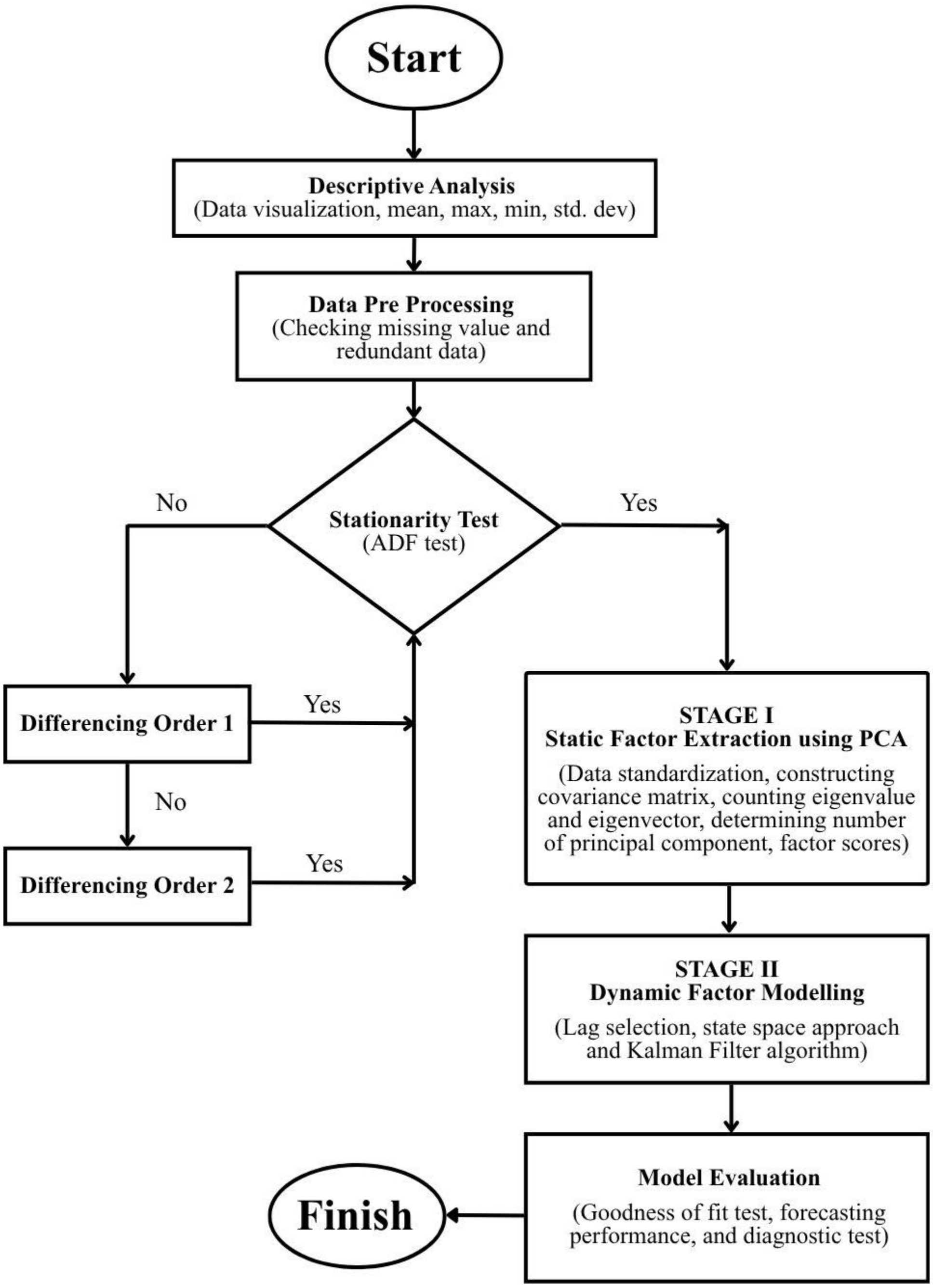

Fig. 1. Flowchart DFM Two Stage Estimation

V. CONCLUSION

Based on the results of the analysis and discussion that have been conducted regarding the application of the integration of Principal Component Analysis (PCA) and Dynamic Factor Model (DFM) on the GRDP data of Lampung Province, the following conclusions can be drawn:

1. The implementation of PCA has proven to be very effective in extracting crucial information from Real and Nominal GRDP variables to mitigate data redundancy due to high multicollinearity. The use of a single factor extraction (F_1) is able to provide a very strong representation by explaining more than 99% of the variance proportion from the original data. This capability has proven to be consistent in accurately capturing regional economic co-movement, both in quarterly and monthly frequency datasets, without causing unrealistic prediction spikes.
2. The integration of PCA and DFM is capable of producing stable and highly accurate forecast estimates on low-dimensional datasets, particularly for quarterly models that have a very high level of precision with MAPE values ranging from 1.43% to 1.52% and meet all classical econometric assumptions through success in the Ljung-Box test. Conversely, although the monthly model has an error level that is practically still accurate (MAPE $< 4\%$), it has a significant weakness in the form of failure in the residual autocorrelation test of Nominal GRDP, which is likely caused by the inability of the Denton Method to unravel the complexity of monthly inflation and the risk of noise due to the double differencing process.

For the Lampung Provincial Government, the use of the quarterly model is very useful as a reliable budget planning basis, while the monthly model still holds value as an early warning system instrument to monitor economic fluctuations more granularly. For researchers, this study enriches the literature regarding the robustness of integrating PCA and DFM on low-dimensional datasets, while also providing a methodological reference that single factor extraction is able to explain data variance despite technical challenges in high-frequency data.

REFERENCES

Aprilianita, W. F., Azriani, Z., & Analia, D. (2024). Analisis faktor-faktor yang mempengaruhi produk domestik regional bruto sektor pertanian Provinsi Sumatera Barat. *Journal of Agribusiness and Community Empowerment (JACE)*, *7*(2), 92–101. https://doi.org/10.32530/jace.v7i2.777

Aslam, M., Sherwani, R. A. K., & Saleem, M. (2021). Vague data analysis using neutrosophic Jarque–Bera test. *PLOS ONE*, *16*(12), Article e0260689. https://doi.org/10.1371/journal.pone.0260689

Brewer, M. J., Butler, A., & Cooksley, S. L. (2016). The relative performance of AIC, AICC and BIC in the presence of unobserved heterogeneity. *Methods in Ecology and Evolution*, *7*(6), 679–692. https://doi.org/10.1111/2041-210X.12541

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), Article 1283. https://doi.org/10.3390/math10081283

Chen, E. Y., & Fan, J. (2023). Statistical inference for high-dimensional matrix-variate factor model. *Journal of the American Statistical Association*, *118*(542), 1038–1055. https://doi.org/10.1080/01621459.2021.1970569

Choi, J., & Yang, X. (2022). Asymptotic properties of correlation-based principal component analysis. *Journal of Econometrics*, *229*(1), 1–18. https://doi.org/10.1016/j.jeconom.2021.08.003

Cynthia, O., Ukamaka, & Onyinyechi, O. (2019). Principal component analysis and its derivation from singular value decomposition. *International Journal of Statistics and Probability*, *8*(2), 183–193. https://doi.org/10.5539/ijsp.v8n2p183

Dien, J., Beal, D. J., & Berg, P. (2005). Optimizing principal components analysis of event-related potentials: Matrix type, factor loading weighting, extraction, and rotations. *Clinical Neurophysiology*, *116*(8), 1808–1825. https://doi.org/10.1016/j.clinph.2004.11.025

Doz, C., Giannone, D., & Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics*, *164*(1), 188–205. https://doi.org/10.1016/j.jeconom.2011.02.012

Eze, C. M., Asogwa, O. C., Onwuamaeze, C. U., Eze, N. M., & Ifeanyi Okonkwo, C. (2020). On the Fourier residual modification of ARIMA models in modeling malaria incidence rates among pregnant women. *American Journal of Theoretical and Applied Statistics*, *9*(1), 1–9. https://doi.org/10.11648/j.ajtas.20200901.11

Fonzo, T. D., & Marini, M. (2012). *On the extrapolation with the Denton proportional benchmarking method* [Technical Report]. Eurostat.

Forkman, J., Josse, J., & Piepho, H.-P. (2019). Hypothesis tests for principal component analysis when variables are standardized. *Journal of Agricultural, Biological and Environmental Statistics*, *24*(2), 289–308. https://doi.org/10.1007/s13253-019-00355-5

Frost, H. R. (2022). Eigenvectors from eigenvalues sparse principal component analysis. *Journal of Computational and Graphical Statistics*, *31*(2), 486–501. https://doi.org/10.1080/10618600.2021.1987254

Hu, X., & Yao, F. (2022). *Dynamic principal component analysis in high dimensions* (arXiv:2104.03087). arXiv. https://doi.org/10.48550/arXiv.2104.03087

Iftikhar, H., Khan, F., Rodrigues, P. C., Alharbi, A. A., & Allohibi, J. (2025). Forecasting of inflation based on univariate and multivariate time series models: An empirical application. *Mathematics*, *13*(7), Article 1121. https://doi.org/10.3390/math13071121

Indah, L. S. M., Zakaria, W. A., & Sari, D. M. (2023). Pengaruh inflasi, PDRB, suku bunga dan tenaga kerja terhadap Nilai Tukar Petani Provinsi Lampung. *Jurnal Ekonomi Pertanian dan Agribisnis*, *7*(3), 1099–1111. https://doi.org/10.21776/ub.jepa.2023.007.03.16

Jeong, S. (2024). Forecast accuracy of demand for registered nurses and its determinants in South Korea. *Human Resources for Health*, *22*(1), Article 44. https://doi.org/10.1186/s12960-024-00910-3

Juanhua, M., & Donglin, X. (2025). China's economic growth from a comprehensive analysis perspective: A synergistic application of GRA, PCA, and GRU models. *Journal of Economics and Management Sciences*, *8*(1), p1–p12. https://doi.org/10.30560/jems.v8n1p1

Karunasingha, D. S. K. (2022). Root mean square error or mean absolute error? Use their ratio as well. *Information Sciences*, *585*, 609–629. https://doi.org/10.1016/j.ins.2021.11.036

Kim, J. H. (2019). Multicollinearity and misleading statistical results. *Korean Journal of Anesthesiology*, *72*(6), 558–569. https://doi.org/10.4097/kja.19087

Kuck, K., & Schweikert, K. (2021). Forecasting Baden-Württemberg's GDP growth: MIDAS regressions versus dynamic mixed-frequency factor models. *Journal of Forecasting*, *40*(5), 861–882. https://doi.org/10.1002/for.2743

Kulikova, M. V., Tsyganova, J. V., & Kulikov, G. Y. (2020). UD-based pairwise and MIMO Kalman-like filtering for estimation of econometric model structures. *IEEE Transactions on Automatic Control*, *65*(10), 4472–4479. https://doi.org/10.1109/TAC.2020.3004737

Kyo, K., Noda, H., & Fang, F. (2024). An integrated approach for decomposing time series data into trend, cycle and seasonal components. *Mathematical and Computer Modelling of Dynamical Systems*, *30*(1), 792–813. https://doi.org/10.1080/13873954.2024.2416631

Lever, J., Krzywinski, M., & Altman, N. (2017). Principal component analysis. *Nature Methods*, *14*(7), 641–642. https://doi.org/10.1038/nmeth.4346

O'Keeffe, H., & Petrova, K. (2025). *Component-based dynamic factor nowcast model* (Staff Report No. 1152). Federal Reserve Bank of New York. https://doi.org/10.59576/sr.1152

Qin, S. J., & Dunia, R. (2000). Determining the number of principal components for best reconstruction. *Journal of Process Control*, *10*(2-3), 245–250. https://doi.org/10.1016/S0959-1524(99)00045-3

Rahmi, S., Sunusi, N., & Ilyas, N. (2025). Spatial autoregressive quantile regression modeling of gross regional domestic product data in Java Island. *MethodsX*, *15*, Article 103621. https://doi.org/10.1016/j.mex.2025.103621

Ringo, J. N. B. S., & Monika, A. K. (2022). Nowcasting regional economic growth in Indonesia. *Buletin Ekonomi Moneter dan Perbankan*, *25*(3), 291–322. https://doi.org/10.21098/bemp.v25i3.1815

Ryan, O., Haslbeck, J. M. B., & Waldorp, L. J. (2025). Non-stationarity in time-series analysis: Modeling stochastic and deterministic trends. *Multivariate Behavioral Research*, *60*(3), 556–588. https://doi.org/10.1080/00273171.2024.2436413

Silva, R. P., Zarpelão, B. B., Cano, A., & Junior, S. B. (2021). Time series segmentation based on stationarity analysis to improve new samples prediction. *Sensors*, *21*(21), Article 7333. https://doi.org/10.3390/s21217333

Song, J., & Shin, S. J. (2018). Stability approach to selecting the number of principal components. *Computational Statistics*, *33*(4), 1923–1938. https://doi.org/10.1007/s00180-018-0826-7

Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *460*(97), 1167–1179. https://doi.org/10.1198/016214502388618960

Supriyatna, P. K., Prastyo, D. D., & Akbar, M. S. (2024). Application of the dynamic factor model on nowcasting sectoral economic growth with high-frequency data. *Media Statistika*, *17*(2), 128–139. https://doi.org/10.14710/medstat.17.2.128-139

Tarsidin, Idham, & Rakhman, R. N. (2018). Nowcasting household consumption and investment in Indonesia. *Buletin Ekonomi Moneter dan Perbankan*, *20*(3), 375–403. https://doi.org/10.21098/bemp.v20i3.858

Zhang, X., & Ding, F. (2020). Adaptive parameter estimation for a general dynamical system with unknown states. *International Journal of Robust and Nonlinear Control*, *30*(4), 1351–1372. https://doi.org/10.1002/rnc.4819