Media Statistika

p-ISSN 1979 – 3693 e-ISSN 2477 – 0647
MEDIA STATISTIKA 17(2) 2024: 128-139
http://ejournal.undip.ac.id/index.php/media_statistika

APPLICATION OF THE DYNAMIC FACTOR MODEL ON NOWCASTING SECTORAL ECONOMIC GROWTH WITH HIGH-FREQUENCY DATA

Putu Krishnanda Supriyatna[1,2], Dedy Dwi Prastyo[1], Muhammad Sjahid Akbar[1]
[1] Institut Teknologi Sepuluh Nopember, Kampus ITS - Sukolilo, Surabaya, Indonesia
[2] Badan Pusat Statistik Provinsi Papua Barat, Manokwari, Indonesia

e-mail: dedy-dp@statistika.its.ac.id

DOI: 10.14710/medstat.17.2.128-139

Article Info:
Received: 14 November 2023
Accepted: 21 December 2024
Available Online: 31 December 2024

Keywords:
ARMAX; *DFM*; *High-Frequency*; *Sectoral GDP*; *Nowcasting.*

Abstract: Economic growth is crucial for planning, yet delayed data releases challenge timely decision-making. Nowcasting offers near-real-time insights using high-frequency indicators (released monthly, weekly, or even daily) to predict low-frequency variables (quarterly or yearly). This study uses high-frequency indicators (monthly), such as stock price changes, air quality, transportation data, financial conditions, and Google Trends, to nowcast quarterly GDP through the Dynamic Factor Model (DFM). The data used span from January 2010 until March 2023, which is split into two: January 2010 until March 2022 for training data and the rest as testing data. Compared to the benchmark Autoregressive Moving Average with Exogenous Variables (ARMAX) model, DFM demonstrates superior accuracy with lower symmetric Mean Absolute Percentage Error (sMAPE). In addition, to evaluate the model performance in nowcasting the GDP across the sector using DFM, the additional metrics, i.e., Root Mean Square Error (RMSE), Mean Absolute Deviation (MAD), and Adjusted R-squared, concluded that in the industrial and transportation sectors results in sufficient nowcasting of GDP, Meanwhile, In the financial sector, the results of the nowcasting GDP give poor estimation results that need improvement.

1. INTRODUCTION

Economic growth, derived from changes in Gross Domestic Product (GDP), relies on numerous indicators, often causing delays in its release by over a month after the observation period. This delay is problematic, as timely indicators can provide accurate insights into economic conditions (Dauphin, 2022), while delays may hinder a country's ability to address economic crises effectively (Eurostat, 2016). Nowcasting offers a timely solution by estimating economic conditions using indicators that were available earlier. High-frequency data, such as monthly economic indicators, can bridge the gap by providing real-time insights compared to quarterly releases. Such data allows closer tracking of economic movements, enabling swift identification of directional changes in economic trends (Schorfeide, 2014; Hoven, 2013).

Nowcasting using high-frequency predictors to estimate low-frequency response variables presents a challenge in model development. The chosen model must effectively address the disparity in frequencies between predictors and responses. A suitable approach is the Dynamic Factor Model (DFM), which utilizes dynamic factors derived from time series data. These factors capture temporal dependencies and are governed by a multivariate time series framework, making DFM well-suited for handling diverse variables.

The DFM has been widely utilized for nowcasting economic conditions. Dauphin (2022) compared DFM with various Machine Learning (ML) techniques such as Support Vector Machines (SVMs), Random Forests (RF), and Neural Networks (NN), including regularized regression models, for nowcasting European economic conditions. The study found that DFM performs better under normal data trends, while ML methods are superior at detecting trend turning points. Fornaro (2013) applied DFM to estimate economic activity indicators in Finland, demonstrating its advantage over autoregressive and moving average models. Similarly, Chernis (2017) used DFM to nowcast regional economic growth in Canadian provinces. Additionally, Rahayu et al. (2023) employed time series regression and Support Vector Regression (SVR) for nowcasting daily Consumer Price Index (CPI) data.

This study will also examine nowcasting economic growth without utilizing high-frequency data. As a benchmark for DFM, the Autoregressive Moving Average with Exogenous Variables (ARMAX) method is employed due to its simplicity and ability to incorporate exogenous variables without converting high-frequency data to low-frequency formats. ARMAX and its variant, ARIMAX, have been widely used for forecasting GDP and economic growth. For instance, Ugoh et al. (2021) utilized ARIMAX (0,0,1) to estimate Nigeria's GDP using four predictor variables. At the regional level, Suhartono et al. (2015) applied ARIMAX with an Eid calendar effect to model pants sales in Boyolali, Central Java. The ARMAX model continues to be developed for predicting regional economic indicators in Indonesia, such as Bank Indonesia's (BI) incoming and outgoing transaction flows, as seen in studies by Suhartono et al. (2018a, 2018b) using the ARIMAX Quantile Regression (ARIMAX-QR) model. Prastyo et al. (2018) compared ARIMAX with Quantile Regression Neural Network (QRNN), while Maghfiroh (2021) applied the ARIMAX Hybrid Basis Function Network (ARIMAX-HBFN) to estimate transaction flows in Central Java Province.

GDP prediction has also been widely applied in Indonesia. For instance, Septiani (2019) analyzed Indonesia's economic growth alongside exchange rates and BI interest rates, using inflation as a predictor variable. In East Java, Imadudin and Prastyo (2022) estimated electricity consumption and Regional GDP with population growth as a key predictor. To further explore the application of the DFM and ARMAX methods, this study focuses on nowcasting economic growth at the sectoral level. Specifically, it examines Indonesia's Manufacturing Industry (C sector), Transportation (H sector), and Finance (K sector), with sector classifications based on the 2020 KBLI (Indonesian Business Field Standard Classification). These sectors were chosen due to their significant contribution to Indonesia's economy (BPS, 2020) and the availability of near real-time economic indicators.

2. LITERATURE REVIEW

2.1. Dynamic Factor Model Two-Stage Estimation (DFM-TS)

The DFM, introduced by Geweke (1977) and Sargent & Sims (1977), was initially designed to estimate models using frequency-domain methods. Giannone (2008) expanded this approach for formal forecasting, particularly when handling extensive datasets with backward-released series at varying lags. This methodology is now widely applied for GDP nowcasting through factor models. The process involves two primary steps: first, estimating model parameters based on principal components extracted from balanced-frequency panel data; second, deriving common factors by applying the Kalman smoother to the full dataset. The DFM utilizes stationary monthly indicators as common factors, which can be adjusted to align with quarterly observations, enabling effective transformations of monthly indicators into common factors (Giannone et al., 2008). The DFM model is formulated in Equation (1).

$$x_{i,t} = \mu + \lambda_{i1} F_{1,t} + \lambda_{i2} F_{2,t} + \cdots + \lambda_{ij} F_{j,t} + \ldots + \lambda_{ir} F_{r,t} + \varepsilon_{i,t} , \tag{1}$$

where the μ is an intercept, $x_{i,t}$ is predictor variable vector at time *t*, λ_{ij} is factor loading, and *j* = 1, 2, 3, …, *r* (number of common factors determined). The $F_{j,t}$ is a common factor vector. The $\boldsymbol{\varepsilon}_{i,t}$ is an idiosyncratic component. The common factor component $\lambda_{ij} F_{j,t}$ and the idiosyncratic component $\boldsymbol{\varepsilon}_{i,t}$ are two stationary unobserved processes. The common component process is assumed to be a linear function of several common factors with $r < n$. Common factors are considered to be able to capture almost all movements in the constituent variables. The common factor follows a Vector Autoregression process of order p or VAR(*p*) formulated in Equation (2).

$$\boldsymbol{F_t} = \sum_{i=1}^{p} \boldsymbol{\Lambda_i} \boldsymbol{F_{t-i}} + \boldsymbol{a_t} , \tag{2}$$

with $\boldsymbol{\Lambda}_i$ is coefficient matrix of the VAR(*p*) process, and vector $\boldsymbol{a_t}$ is assumed to be white noise with a covariance matrix $\boldsymbol{\Sigma}_a$. The reduced-form VAR(*p*) cannot accommodate contemporaneous effects, such that the structural version was developed as so-called Structural VAR(*p*) or SVAR(*p*), written as in Equation (3).

$$\boldsymbol{F_t} = \sum_{i=1}^{p} \boldsymbol{\widetilde{\Lambda}_i} \boldsymbol{F_{t-i}} + \boldsymbol{u_t} , \tag{3}$$

with $\boldsymbol{u}_t = \boldsymbol{B}^{-1} \boldsymbol{a}_t$, and matrix $\boldsymbol{B}$ captured the structural impact from $\boldsymbol{a}_t$ to $\boldsymbol{F}_t$, and the $\widetilde{\boldsymbol{\Lambda}}_i = \boldsymbol{B}^{-1}\boldsymbol{\Lambda}_i$.

One common factor estimation method that can be applied to the DFM method is the two-step estimation, which combines the Principal Component (PC) method and the Kalman Filter (KF) and smoothing technique. This method was developed by Doz (2012) for the identification of macroeconomic shocks. Unobserved common factors can be estimated consistently with the PC of the observed variables. However, if there is missing data, it is not enough to use the PC such that the KF is employed to estimate optimal parameters for unobserved data. The $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{u}_t$ are Normally distributed, allowing the use of the KF technique in common factor extraction. The common factor is influenced by past data up to the *p* lag and a number of common factors (*r*), which are relatively large to the number of common shocks (*q*) aiming to capture the relationship among the structural breaks or lead/lag relationship.

Once the common factors in Equation (3) were estimated, the next step to nowcast is employing a linear regression model to obtain the relationship between the low-frequency response variable and the estimate of the common factor. The model is formulated in the Equation (4).

$$\boldsymbol{y_t} = \boldsymbol{\beta}' \boldsymbol{\widehat{F}_t} + \boldsymbol{\varepsilon_t} , \tag{4}$$

with the Ordinary Least Square (OLS) estimator given in Equation (5).

$$\boldsymbol{\widehat{\beta}} = \left(\boldsymbol{\widehat{F}_t'} \boldsymbol{\widehat{F}_t}\right)^{-1} \boldsymbol{\widehat{F}_t'} \boldsymbol{y} . \tag{5}$$

The *h*-period-ahead forecast is formulated in Equation (6).

$$\hat{\boldsymbol{y}}_{t+h} = \hat{\boldsymbol{\beta}}' \hat{\boldsymbol{F}}_{t+h} . \quad (6)$$

The forecast of $\hat{\boldsymbol{F}}_{t+h}$ is obtained from the forecast of SVAR model formulated in Equation (2) and (3). If the *h*-period for the low-frequency variable is shorter than the time increment of the high-frequency variable, the forecast is considered a nowcast.

2.2. Autoregressive Moving Average (ARMA)

The ARMA model consists of two processes, namely AR (Autoregressive) and MA (Moving Average). AR is a process that describes how data from the current period is affected by past conditions. The MA is a process that describes how data is affected by its past residual value. Mathematically, an AR model with order p or AR(p) can be defined in Equation (7), while the MA model is expressed in Equation (8).

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t \quad (7)$$

The systematic MA model of order q or MA(q) can be explained as follows.

$$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \quad (8)$$

Therefore, the ARMA (p, q) process, which is a combination of AR(p) and MA(q), can be seen in Equation (9).

$$Y_t = \mu + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \quad (9)$$

The identification phase of the ARMA model is carried out by examining the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. The sample ACF ($\hat{\rho}_j$) shows how the correlation between Y_t and Y_{t-j}, whereas the sample PACF ($\hat{\phi}_{t,t-j}$) shows the correlation between Y_t and Y_{t-j} by eliminating the existence of a shared linear relationship in the variables $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-j-1}$ (Wei, 2006).

2.3. ARMA with Exogenous Variables (ARMAX)

The ARMAX model extends the ARMA process by incorporating exogenous variables into the framework. This means that the ARMAX model is influenced not only by its historical data but also by the predictor or external variables included in the model. The general form of the ARMAX model is expressed in Equation (10) (Hyndman, 2010).

$$\begin{aligned} Y_t &= \varphi_1 X_{1t} + \varphi_1 X_{2t} + \cdots + \varphi_1 X_{ut} + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} \\ &\quad + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \end{aligned} \quad (10)$$

The assumptions of ARMA and ARMAX are the same, i.e., the errors are white noise and follow Normal distribution. Compared to the ARMA model, the ARMAX model has an additional variable X_{1t} to X_{ut} which are exogenous variables. The ARMAX consists of three specified orders. Order p for the AR model, order q for the MA model, and order u for the number of exogenous variables within the model so that it can be written ARMAX (p,q,u). Based on Equation (10), it can also be noted that the ARMAX model is a transfer function model, where the special case leads to an intervention model, with the order (b,r,s) being zero.

3. MATERIAL AND METHOD

This study uses high-frequency indicators (monthly), such as stock price changes, air quality, transportation data, financial conditions, and Google Trends, where the detail is written in Tabel 1, and the quarterly GDP from three sectors: Manufacturing Industry (C sector), Transportation (H sector), and Finance (K sector) that were chosen due to their significant contribution to Indonesia's economy (BPS, 2020). The data from January 2010 until March 2023 were split into two: (1) data training from January 2010 until March 2022, and (2) data testing for the rest from April 2022 until March 2023. The analysis using DFM was done using the "nowcasting" package in R software (de Valk et al., 2019).

3.1. DFM Model

The model specifications of the DFM method, from indicators to transformation into common factors, are as follows:

$$Y_{t,C} = \beta_{0,C} + \beta_{1,C} F_{t,1,C} + \beta_{2,C} F_{t,2,C} + \cdots + \beta_{r_C,C} F_{t,r,C} + \varepsilon_{t,C} \tag{11}$$

$$Y_{t,H} = \beta_{0,H} + \beta_{1,H} F_{t,1,H} + \beta_{2,H} F_{t,2,H} + \cdots + \beta_{r_H,H} F_{t,r,H} + \varepsilon_{t,H} \tag{12}$$

$$Y_{t,K} = \beta_{0,K} + \beta_{1,K} F_{t,1,K} + \beta_{2,K} F_{t,2,K} + \cdots + \beta_{r_k,K} F_{t,r,K} + \varepsilon_{t,K} \tag{13}$$

with

$$\begin{aligned} F_{t,j,C} &= \lambda_{1,C}\, x_{t,1} + \lambda_{2,C}\, x_{t,2} + \cdots + \lambda_{6,C}\, x_{t,6} + \lambda_{7,C}\, x_{t,1,C} \\ &\quad + \lambda_{8,C}\, x_{t,2,C} + \cdots + \lambda_{12,C}\, x_{t,6,C} \end{aligned} \tag{14}$$

$$\begin{aligned} F_{t,j,H} &= \lambda_{1,H}\, x_{t,1} + \lambda_{2,H}\, x_{t,2} + \cdots + \lambda_{6,H}\, x_{t,6} + \lambda_{7,H}\, x_{t,1,H} \\ &\quad + \lambda_{8,H}\, x_{t,2,H} + \cdots + \lambda_{14,H}\, x_{t,8,H} \end{aligned} \tag{15}$$

$$\begin{aligned} F_{t,j,K} &= \lambda_{1,K}\, x_{t,1} + \lambda_{2,K}\, x_{t,2} + \cdots + \lambda_{6,K}\, x_{t,6} + \lambda_{7,K}\, x_{t,1,K} \\ &\quad + \lambda_{8,K}\, x_{t,2,K} + \cdots + \lambda_{13,K}\, x_{t,7,K} \end{aligned} \tag{16}$$

where $Y_{t,C}$ is industrial sector economic growth, $Y_{t,H}$ is transportation sector economic growth, $Y_{t,K}$ is finance sector economic growth, $F_{t,j,C}$ is a common factor in the DFM model of the industrial sector, $F_{t,j,H}$ is a common factor in the DFM model of the transportation sector; $F_{t,j,K}$ is a common factor in the finance sector's DFM model. The j is common factor index with j = 1, 2, …, r.

The assumption of the model in Equation (11)-(16) follows the assumption of DFM that has been explained in subsection 2.1. Some other assumptions are written within the following parameter estimation procedure. In the two-stage procedure, firstly, the matrices $\boldsymbol{\Lambda}$ and $\boldsymbol{F}_t$ are estimated using PC, where the data are standardized, balanced panel $\bar{\mathbf{x}}_t$, with no missing values and outliers. The second stage re-estimates the factors of an unbalanced panel $\mathbf{x}_t$ considering the parameters obtained in the previous step using KF smoothing. Another procedure that can be used to estimate the DFM parameters is the Expectation-Maximization (EM) algorithm. It is employed to estimate parameters where the missing observations exist.

3.2. ARMAX Model

The model specifications of the ARMAX method are written as follows.

$$\begin{aligned} Y_{t,C} &= \phi_{1,C} Y_{t-1,C} + \phi_{2,C} Y_{t-2,C} + \cdots + \phi_{p,C} Y_{t-p,C} + \varepsilon_{t,C} - \theta_{t-1,C}\varepsilon_{t-1,C} \\ &\quad - \theta_{t-2,C}\varepsilon_{t-2,C} + \cdots + \theta_{t-q,C}\varepsilon_{t-q,C} + \varphi_{1,C} x_{t,1,C} + \varphi_{2,C}\, x_{t,2} + \cdots \\ &\quad + \varphi_{6,C}\, x_{t,6} + \varphi_{7,C}\, x_{t,1,C} + \varphi_{8,C}\, x_{t,2,C} + \cdots + \varphi_{12,C}\, x_{t,6,C} \end{aligned} \tag{17}$$

$$\begin{aligned} Y_{t,H} &= \phi_{1,H} Y_{t-1,H} + \phi_{2,H} Y_{t-2,H} + \cdots + \phi_{p,H} Y_{t-p,H} + \varepsilon_{t,H} - \theta_{t-1,C}\varepsilon_{t-1,H} \\ &\quad - \theta_{t-2,H}\varepsilon_{t-2,H} + \cdots + \theta_{t-q,H}\varepsilon_{t-q,H} + \varphi_{1,H}\, x_{t,1} + \varphi_{2,H}\, x_{t,2} + \cdots \end{aligned} \tag{18}$$

$$+\varphi_{6,H}\, x_{t,6} + \varphi_{7,H}\, x_{t,1,H} + \varphi_{8,H}\, x_{t,2,H} + \cdots + \varphi_{14,H}\, x_{t,8,H}$$

$$\begin{aligned} Y_{t,K} = {} & \phi_{1,K} Y_{t-1,K} + \phi_{2,K} Y_{t-2,K} + \cdots + \phi_{p,K} Y_{t-p,K} + \varepsilon_{t,K} - \theta_{t-1,K}\varepsilon_{t-1,K} \\ & - \theta_{t-2,K}\varepsilon_{t-2,K} + \cdots + \theta_{t-q,K}\varepsilon_{t-q,K} + \varphi_{1,K}\, x_{t,1} + \varphi_{2,K}\, x_{t,2} + \cdots \\ & +\varphi_{6,K}\, x_{t,6} + \varphi_{7,K}\, x_{t,1,K} + \lambda_{8,K}\, x_{t,2,K} + \cdots + \lambda_{13,K}\, x_{t,7,K} \end{aligned} \quad (19)$$

The ARMAX models in Equation (17)-(19) were estimated separately for each sector. These benchmark models were employed with exogenous variables written in Table 1.

3.3. Data and Variables

The variables used in this study consisted of response variables from the three sectors (Y_C, Y_H, Y_K), predictor variables consist of six macroeconomic variables (X_1 until X_6), industrial sector-specific predictor variables ($x_{1,C}$ until $x_{6,C}$), transportation sector-specific predictor variables ($x_{1,H}$ until $x_{8,H}$), and financial sector-specific predictor variables ($x_{1,K}$ until $x_{7,K}$). The details of these variables are explained in Table 1. The chosen predictor variables consider their availability monthly as they are treated as high-frequency variables needed in the nowcasting. The variables with specification models that have been written in Equation (11)-(19) were not tested for their significance as these approaches, indeed, focus on accuracy.

Table 1. Variables of Study

Variable	Details	Source
Y_C	GDP Growth of Manufacturing Industry Sector	Badan Pusat Statistik (BPS)
Y_H	GDP Growth of Transportation and Warehousing Sector	BPS
Y_K	GDP Growth of the Finance Sector	BPS
X_1	Growth of Consumer Confidence Index	BI
X_2	Growth of Consumer Expectation Index	BI
X_3	Growth of Income Expectation Index	BI
X_4	Growth of Job Availability Expectation Index	BI
X_5	Growth of Business Activity Expectation Index	BI
X_6	Growth of Price Expectation Index for the Next 3 Months	BI
$X_{1,C}$	Growth of Indonesia's Air Quality Index	www.iqair.com
$X_{2,C}$	Average Price Growth of Metal & Similar Industry's Stock	Yahoo Finance
$X_{3,C}$	Average Price Growth of the Plastic & Packaging Industry's Stock	Yahoo Finance
$X_{4,C}$	Average Price Growth of the Pharmaceutical Industry's Stock	Yahoo Finance
$X_{5,C}$	Average Price Growth of the Cosmetic Industry's Stock	Yahoo Finance
$X_{6,C}$	Average Price Growth for the Household Appliances Industry's Stock	Yahoo Finance
$X_{1,H}$	Growth in the Number of Train Passengers	BPS
$X_{2,H}$	Growth in the Number of Aircraft Passengers at Soekarno-Hatta Airport	BPS
$X_{3,H}$	Growth in the Number of Aircraft Passengers at Polonia Airport	BPS
$X_{4,H}$	Growth in the Number of Aircraft Passengers at Juanda Airport	BPS
$X_{5,H}$	Growth in Number of Aircraft Passengers at Ngurah Rai Airport	BPS

Table 1. Variables of Study

Variable	Details	Source
$X_{6,H}$	Growth in the Number of Aircraft Passengers at Sultan Hasanudin Airport	BPS
$X_{7,H}$	Average Price Growth of Toll Road Companies, Ports, Airports, and the Like's Stock	Yahoo Finance
$X_{8,H}$	Growth of Google Trend Index "*Macet*"	Google Trend
$X_{1,K}$	Average Price Growth of Bank's Stock	Yahoo Finance
$X_{2,K}$	Average Price Growth of Insurance Company's Stock	Yahoo Finance
$X_{3,K}$	Average Price Growth of Investment Company's Stock	Yahoo Finance
$X_{4,K}$	Growth of Consumer Saving Proportion	BI
$X_{5,K}$	Growth of Consumer Loan Repayments Proportion	BI
$X_{6,K}$	Growth of Google Trend Index "*Kredit*"	Google Trend
$X_{7,K}$	Growth of Google Trend Index "*Hutang*"	Google Trend

4. RESULTS AND DISCUSSION

Based on the results of the Phillips-Perron (PP) test, all predictor variables included in the DFM model are stationary at level. The six macroeconomic variables, the X_1 has a *p*-value of 0.0488, while the five remaining predictors result in *a p*-value less than 0.01. These results indicate that all six macroeconomic variables are stationary. Table 1 shows the *p*-value of the Phillip-Perron Test for the remaining predictors that concluded all predictors are stationary.

Table 2. Phillips-Perron Test on Predictors

Predictor	*p*-value	Predictor	*p*-value	Predictor	*p*-value
$X_{1,C}$	<0,01	$X_{1,H}$	<0,01	$X_{1,K}$	0,0488
$X_{2,C}$	<0,01	$X_{2,H}$	<0,01	$X_{2,K}$	0,0342
$X_{3,C}$	0,0202	$X_{3,H}$	<0,01	$X_{3,K}$	0,0427
$X_{4,C}$	<0,01	$X_{4,H}$	<0,01	$X_{4,K}$	<0,01
$X_{5,C}$	<0,01	$X_{5,H}$	<0,01	$X_{5,K}$	<0,01
$X_{6,C}$	0,0438	$X_{6,H}$	<0,01	$X_{6,K}$	<0,01
		$X_{7,H}$	0,0118	$X_{7,K}$	<0,01
		$X_{8,H}$	<0,01		

Table 3. DFM Model Specification

Sector	Full Data DFM			Training Data DFM		
	Optimum Lag	Factor Count	Variable Count	Optimum Lag	Factor Count	Variable Count
Industrial	1	4	12	1	4	12
Transportation	1	3	14	1	3	14
Finance	1	4	13	1	4	13

The DFM model formulated in Equations (1) and (3), with its specification as formulated in Equation (11) until (16), needs the determination of optimum lag, number of factor count, and variable count. The optimum lag is determined based on minimum *Akaike's information criterion* (AIC), *Hannan–Quinn information criterion* (HQ), *Schwarz criterion* (SC), and *Final Prediction Error criterion* (FPE). The factor counts (r) are determined by information criterion to minimize idiosyncratic components (ICr). Then, the specification of the DFM for each GDP sector is summarized in Table 3.

Then, all the common factors that are formed are regressed using the OLS method, which includes predictor variables in each sector. There are two models for each sector: common factor estimation using the DFM-TS methods, based on the specification model in Equation (17)-(19), and OLS linear regression as defined in Equation (4). Those models are written in Equation (20)-(25).

$$Y_{t,C} = \beta_{0,C} + \beta_{1,C}F_{t,1,C} + \beta_{2,C}F_{t,2,C} + \beta_{3,C}F_{t,3,C} + \beta_{2,C}\, F_{t,4,C} + \varepsilon_{t,C} \tag{20}$$

$$Y_{t,H} = \beta_{0,H} + \beta_{1,H}F_{t,1,H} + \beta_{2,H}F_{t,2,H} + \beta_{3,H}F_{t,3,H} + \varepsilon_{t,H} \tag{21}$$

$$Y_{t,K} = \beta_{0,K} + \beta_{1,K}F_{t,1,K} + \beta_{2,K}F_{t,2,K} + \beta_{3,K}F_{t,3,K} + \beta_{4,K}F_{t,4,K} + \varepsilon_{t,K} \tag{22}$$

Using OLS regression, the models formed on the DFM-TS are as follows.

$$\hat{Y}_{t,C} = 3.91 + 0.9\,\hat{F}_{t,1,C} - 0.12\,\hat{F}_{t,2,C} + 0.63\,\hat{F}_{t,3,C} + 0.69\,\hat{F}_{t,4,C} \tag{20}$$

$$\hat{Y}_{t,H} = 5.31 + 2.27\,\hat{F}_{t,1,H} + 1.98\,\hat{F}_{t,2,H} + 1.18\,\hat{F}_{t,3,H} \tag{21}$$

$$\hat{Y}_{t,K} = 6.16 + 0.48\,\hat{F}_{t,1,K} - 0.03\,\hat{F}_{t,2,K} + 0.42\,\hat{F}_{t,3,K} + 1.05\,\hat{F}_{t,4,K} \tag{22}$$

As shown in Figure 1, the nowcasting results during the testing period exhibit notable estimation errors, particularly in the first two periods, where the values are significantly underestimated. In subsequent periods, the nowcasting estimates align more closely with the actual data. However, by the end of 2022, the nowcasting results substantially overestimate the industrial sector's growth rate, projecting it to exceed ten percent. At the start of 2023, the nowcasting results diverge, indicating a movement direction inconsistent with the actual data. The training period estimation captures most actual data trends, including the slowdown during COVID-19 and the high growth during economic recovery. However, some errors occurred, such as underestimations in 2019.

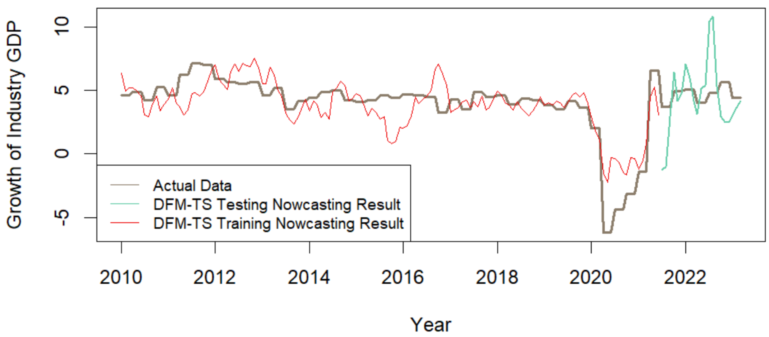

Figure 1. Comparison of DFM Model Nowcasting Results with Actual Data in the Industrial Sector (Percent)

Figures 2 and 3 highlight several challenges in nowcasting. During the testing period for the transportation sector, the nowcasting initially aligned with growth trends but included an overestimated point and diverged from actual data at the start of 2022, recovering accuracy mid-year. However, late 2022 through the testing period saw consistent underestimations, though the direction of movement was captured. For the training period, the nowcasting described the COVID-19 slowdown and post-recovery growth but underestimated at certain points. In the financial sector, nowcasting struggled during both training and testing periods, failing to capture actual trends due to the sector's unpredictable nature.

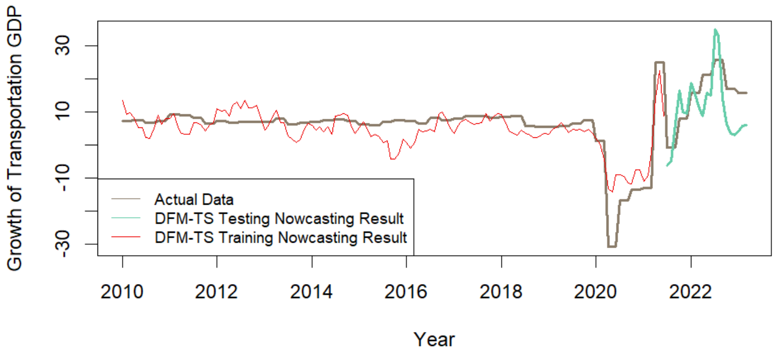

Figure 2. Comparison of DFM Model Nowcasting Results with Actual Data in the Transportation Sector (Percent)

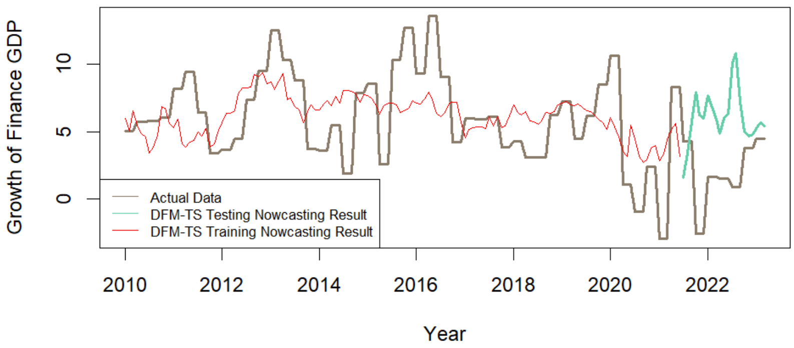

Figure 3. Comparison of DFM Model Nowcasting Results with Actual Data in the Finance Sector (Percent)

In addition, the selection of predictor variables limited to fast-released data only results in the absence of indicators capable of describing movements in the financial sector. This reason is supported by the very low Adjusted R-squared value in the financial sector DFM model shown in Table 3. Additional metrics, i.e., Root Mean Square Error (RMSE) and Mean Absolute Deviation (MAD), were also calculated and summarized in Table 4.

Table 4. The Goodness Specification of the DFM Model on Economic Growth Nowcasting

Sector	RMSE	sMAPE	MAD	*Adjusted R-Squared*
Industrial	2.7227	0.5022	2.0690	0.6371
Transportation	8.5845	0.7626	10.8949	0.6718
Finance	5.5307	0.9950	6.4078	0.1845

At the sectoral level, the DFM method performs best for nowcasting in the industrial sector, while the DFM-TS method is most suitable for the transportation sector. However, in the financial sector, the DFM method shows poor performance, failing to estimate economic growth trends.

In the ARMAX model, one of the conditions for data to be modeled is that it is already stationary. Based on the results of the PP-test, the three response variables were stationary at the level. Meanwhile, predictor variables do not need to be tested for stationarity. Data that has been tested for stationarity can be involved in the formation of the ARMA model.

Table 5. The Best ARMA and ARMAX Models of Each Model

Sector	ARMA (*p*,*q*) Model				ARMAX (*p*,*q*,*u*) Model			
	Model	LB Test	KS-Test	SMAPE	Model	LB Test	KS-Test	SMAPE
Industrial	ARMA (1,0)	*White noise*	Normal	0.17	ARMAX (1,0,12)	*White noise*	Normal	0.58
Transport	ARMA (1,0)	*White noise*	Not Normal	0.98	ARMAX (1,0,14)	*White noise*	Normal	0.76
Finance	ARMA (1,0)	*White noise*	Not Normal	1.04	ARMAX (1,0,13)	*White noise*	Not Normal	1.07

Using ACF and PACF plots, the best ARMA model candidates for each sector were identified. These models were evaluated based on parameter significance, sMAPE value, and residual assumptions of normality and white noise. The ARMA(1,0) model emerged as the best for all sectors and was extended with *u* predictor variables to form the ARMAX model. Table 5 summarizes the transition from ARMA to ARMAX and compares nowcasting results using DFM and ARMAX models across three sectors. The results, focused on sMAPE, show that the DFM-TS method outperforms ARMAX with lower sMAPE values, see Table 6.

Table 6. Comparison of sMAPE from DFM-TS and ARMAX models

Sektor	sMAPE	
	DFM-TS	ARMAX
Industrial	0.5022	0.5814
Transportation	0.7626	0.7631
Finance	0.9950	1.0727

This study identifies several key gaps that contribute to its novelty. First, it addresses the limited exploration of sector-specific GDP nowcasting by applying DFM to industrial, transportation, and financial sectors using near real-time indicators like stock prices, air quality, and Google Trends. Second, it highlights the varied performance of DFM across sectors, revealing sufficient accuracy in industrial and transportation sectors but suboptimal results in finance, underscoring the need for refinement. Third, it diversifies data sources by integrating less common high-frequency variables, enhancing nowcasting responsiveness. Fourth, it systematically benchmarks DFM against ARMAX, confirming DFM's superior accuracy across metrics like sMAPE and RMSE. Finally, the study identifies challenges in capturing the financial sector's volatility, suggesting opportunities for methodological improvements. This work advances sectoral nowcasting research and sets the stage for refining models and expanding high-frequency data use.

5. CONCLUSION

This study highlights the potential of nowcasting sectoral economic growth using near real-time indicator variables, demonstrating mixed outcomes across sectors. The DFM proved relatively effective for the industrial and transportation sectors, offering reasonably accurate predictions despite some limitations in estimating precise economic growth values. In contrast, the financial sector faced significant challenges, with the DFM yielding poor estimation results. However, even in this case, the DFM outperformed the benchmark ARMAX model across all three sectors, showcasing its superiority in handling high-frequency data for nowcasting purposes. These findings underscore the promise of the DFM approach while highlighting areas for further refinement, particularly in the financial sector.

ACKNOWLEDGEMENT

This research was funded by the Directorate of Research, Technology, and Community Service (DRTPM), Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia (*Kemdikbudristek*) through the Institut Teknologi Sepuluh Nopember (ITS) under contract numbers 038/E5/PG.02.00.PL/2024 and 1743/PKS/ITS/2024.

REFERENCES

Badan Pusat Statistik (BPS). (2020). Klasifikasi Baku Lapangan Usaha Indonesia (KBLI) 2020, Jakarta: Badan Pusat Statistik.

Chernis, T. & Sekkel, R. (2017). A Dynamic Factor Model for Nowcasting Canadian GDP Growth. *Empirical Economics*, *53*(1), 217-234, August. DOI: 10.1007/s00181-017-1254-1.

Dauphin, J., Dybczak, K., Maneely, M., Sanjani, M.T., Suphaphipat, N., Wang, Y., & Zhang, H. (2022). Nowcasting GDP A Scalable Approach Using DFM, Machine Learning and Novel Data. *Applied to European Economies*. IMF Working Paper.

de Valk, S., de Mattos, D., & Ferreira, P. (2019). Nowcasting: An R Package for Predicting Economic Variables Using Dynamic Factor Models. *The R Journal*, *11*(1), 230-244. https://doi.org/10.32614/RJ-2019-020

Doz, C., Giannone, D., & Reichlin, L. (2006). A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models. European Central Bank Working Paper Series No 674 / September 2006.

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.

Eurostat. (2016). Overview of GDP flash estimation methods. Luxembourg: Publications Office of the European Union.

Farih, I. & Prastyo, D. D. (2022). Forecasting Electricity Consumption Based On Economics and Social Indicators Using VAR Model with Exogenous Variable: Evidence From East Java Province. 2022 6th ICITISEE. DOI: 10.1109/ICITISEE57756.2022.10057638.

Fornaro, P. (2013). Predicting Finnish Economic Activity using Firm-level Data. Working Paper University of Helsini Faculty of Social Sciences.

Geweke, J. (1977). The Dynamic Factor Analysis of Economic Time Series in Latent Variables in Socio-Economic Models, ed. by D.J. Aigner and A.S. Goldberger, Amsterdam: North-Holland.

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time Informational Content of Macroeconomic Data. *Journal of Monetary Economics*, *55*, 665–676.

Hoven, L., & Schreurs, G. (2013). Towards a Monthly Indicator of Economic Growth. e Joint EU/OECD Workshop on Recent Developments in Business and Consumer Surveys, Brussels, 14-15 November 2013.

Hyndman, R. J. (2010). The ARIMAX Model Muddle. Accessible in: http://robjhyndman.com/hyndsight/arimax/. Accessed on 17 March 2023.

Maghfiroh, Z.F., Suhartono, S., Prabowo, H., Salehah, N. A., Prastyo, D. D., & Setiawan, S. (2021). Forecasting Inflow and Outflow of Currency in Central Java using ARIMAX, RBFN and Hybrid ARIMAX-RBFN. *J. Phys.: Conf. Ser*. 1863 012066.

Prastyo, D.D., Suhartono, S., Puka, A.O. B., & Lee, M.H. (2018). Comparison between Hybrid Quantile Regression Neural Network and Autoregressive Integrated Moving Average with Exogenous Variable for Forecasting of Currency Inflow and Outflow in Bank Indonesia. *Jurnal Teknologi*, *80*(6), 61-68.

Rahayu, S.D., Prastyo, D.D., & Setiawan, S. (2023). Nowcasting of Daily Consumer Price Index using Time Series Regression and Support Vector Regression. *AIP Conference Proceedings*, 2540, 080033.

Sargent, T. J. & Sims, C. A. (1977). *Business Cycle Modeling without Pretending to Have Too Much a Priori Economic Theory*. Minneapolis: Federal Reserve Bank of Minneapolis.

Schorfheide, F. & Song, D. (2014). Real-time Forecasting with a Mixed Frequency VAR. *Journal of Business & Economic Statistics*, *3*, 366380.

Septiani, A., Sumertajaya, I. M., & Aidi, M.N. (2019). Vector Autoregressive X (VARX) Modeling for Indonesian Macroeconomic Indicators and Handling Different Time Variations with Cubic Spline. Department of Statistics, IPB University, Bogor, Indonesia.

Suhartono, S., Lee M. H., & Prastyo, D.D. (2015). Two Levels ARIMAX and Regression Models for Forecasting Time Series Data with Calendar Variation Effects. *AIP Conference Proceedings* 1691, 050026 (2015). Doi: 10.1063/1.4937108.

Suhartono, S., Salehah, N. A., Prastyo, D. D., & Rahayu, S.P. (2018a). Hybrid ARIMAX Quantile Regression Model for Forecasting Inflow and Outflow of East Java Province. *J. Phys.: Conf. Ser*. 1028, 012228

Suhartono, S., Saputri, P.D., Prastyo, D.D., & Rahayu, S.P. (2018b). Hybrid Quantile Regression Neural Network Model for Forecasting Currency Inflow and Outflow in Indonesia. *J. Phys.: Conf. Ser*. 1028, 012213

Ugoh, C. G., Uzuke, C. A., Ugoh, D. O. (2021). Application of ARIMAX Model on Forecasting Nigeria's GDP. *American Journal of Theoretical and Applied Statistics, 10*(5): 216-225.

Wei, W. W. (2006). *Time Series Analysis Univariate and Multivariate Methods*. New York: Pearson Education.