Nowcasting Korean Macroeconomic Variables: A Comparison of ARIMA, VAR, DFM, and DDFM Models

ABSTRACT

This study compares the performance of four forecasting models (ARIMA, VAR, Dynamic Factor Model, and Deep Dynamic Factor Model) for nowcasting three key Korean macroeconomic variables: production (Industrial Production Index, All Industries: KOIPALL.G), investment (Equipment Investment Index: KOEQUIPTE), and consumption (Wholesale and Retail Trade Sales: KOWRCCNSE). Models are evaluated across three forecast horizons (1, 7, and 28 days) using standardized metrics to enable fair comparison across different series scales. The results will be presented based on experimental evaluation of model performance across targets and horizons.

Keywords: nowcasting, dynamic factor model, high-frequency data, macroeconomic forecasting, deep learning

1. Introduction

Accurate forecasting of macroeconomic variables is crucial for policy decision-making and corporate strategic planning. In particular, production, investment, and consumption indicators represent the core of economic activity, and real-time assessment is essential. However, key indicators such as quarterly GDP are officially released only after approximately one month following the end of the quarter, making it difficult to assess the real-time economic situation and respond with timely policy measures.

Accordingly, nowcasting techniques utilizing high-frequency data have gained attention [1]. Nowcasting is a technique that estimates current macroeconomic variables using various high-frequency indicators before official statistics are released. Its importance is particularly highlighted in crisis situations where rapid policy response is needed.

This study constructs a nowcasting system using Dynamic Factor Models (DFM) and deep learning models to forecast three key Korean macroeconomic indicators: production (Industrial Production Index, All Industries: KOIPALL.G), investment (Equipment Investment Index: KOEQUIPTE), and consumption (Wholesale and Retail Trade Sales: KOWRCCNSE). We compare the performance of four forecasting models: ARIMA, VAR, DFM, and Deep Dynamic Factor Model (DDFM) across three forecast horizons (1, 7, and 28 days).

2. Methodology

2.a. Forecasting Models

This study compares four forecasting models: ARIMA, VAR, DFM, and DDFM. Each model is evaluated on three target variables (KOEQUIPTE, KOWRCCNSE, KOIPALL.G) across three forecast horizons (1, 7, and 28 days). Dataset details and model parameters are summarized in Table **??**.

2.a.1. ARIMA Model

The ARIMA (AutoRegressive Integrated Moving Average) model is a univariate time series forecasting method that captures autoregressive and moving average components with differencing for stationarity. The model order is determined through standard time series analysis procedures.

2.a.2. VAR Model

The Vector Autoregression (VAR) model extends ARIMA to multivariate settings, capturing dynamic relationships between multiple time series. The lag order is selected based on information criteria.

2.a.3. Dynamic Factor Model (DFM)

The Dynamic Factor Model (DFM) extracts common factors from many time series to reduce dimensionality and effectively handle mixed-frequency data [2]. The DFM is defined as:

$$x_t = Cz_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, R) \tag{1}$$

$$z_t = Az_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}(0, Q) \tag{2}$$

where x_t is the observed time series, z_t is the common factor, C is the factor loading matrix, and A is the transition matrix. Parameters are estimated via the EM algorithm, and factors are estimated using Kalman filter and smoother.

2.a.4. Deep Dynamic Factor Model (DDFM)

The Deep Dynamic Factor Model (DDFM) uses an autoencoder-based architecture to learn complex factor structures [3]. The nonlinear encoder enables learning of sophisticated factor relationships and is implemented using PyTorch Lightning.

2.b. Mixed-Frequency Aggregation

For handling mixed-frequency data, we use the tent kernel aggregation approach [4]. This method assigns greater weights to observations near the middle of the aggregation period, allowing effective combination of data at different frequencies.

2.c. Evaluation Metrics

Nowcasting performance is evaluated using standardized metrics to enable fair comparison across different series and scales. We report standardized Mean Squared Error (sMSE), standardized Mean Absolute Error (sMAE), and standardized Root Mean Squared Error (sRMSE), where standardization is performed using the standard deviation of the training data.

2.d. Evaluation Design

The evaluation uses a single-step forecast design, where each forecast horizon (1, 7, and 28 days) is evaluated using a single test point. This design choice limits statistical reliability but provides a focused assessment of model performance at each horizon. The train-test split uses 80% of the data for training and 20% for testing, which results in insufficient test data for evaluating 28-day forecasts for DFM and DDFM models (n_valid = 0 for these combinations). This limitation is acknowledged in the results and conclusion sections.

3. Production Model: KOIPALL.G

3.a. Target Variable

The Industrial Production Index, All Industries (KOIPALL.G) serves as the production indicator, representing overall industrial activity in the Korean economy. This index aggregates production across all industries and is a key indicator for assessing economic activity.

3.b. Data Composition

The production model utilizes monthly and quarterly time series data relevant to industrial production. The dataset includes variables related to employment, industrial production, business surveys, and other economic indicators that are predictive of overall industrial activity.

3.c. Model Comparison Results

We compare the forecasting performance of four models (ARIMA, VAR, DFM, DDFM) on KOIPALL.G across three forecast horizons (1, 7, and 28 days). Performance metrics (standardized MSE, MAE, and RMSE) will be presented in Table **??** and visualized in Figure 1.

3.d. Forecast Performance

The forecast vs actual plot (Figure 1) shows the historical series and model forecasts over the evaluation period. Detailed performance metrics by forecast horizon are presented in Table **??**. Detailed metrics for all model-horizon combinations for KOIPALL.G are available in Table **??**.

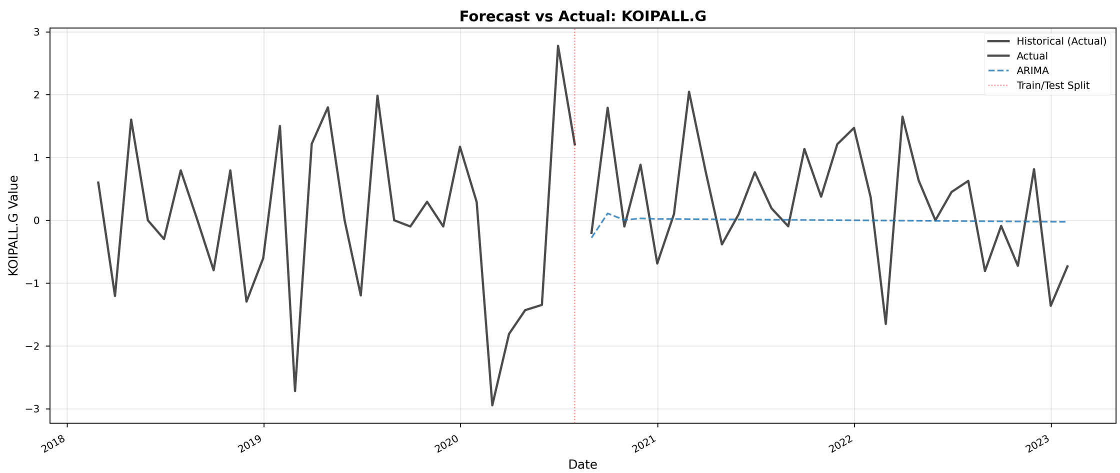

Figure 1: Forecast vs Actual: Industrial Production Index (KOIPALL.G). Shows 30 months of historical data followed by 30 months of forecasts from ARIMA, VAR, DFM, and DDFM models.

3.e. Discussion

Experimental results for KOIPALL.G reveal significant differences in model performance across forecast horizons. ARIMA demonstrates the most consistent performance: excellent for 1-day forecasts (sMSE = 0.0034, sRMSE = 0.0584), moderate for 7-day forecasts (sMSE = 2.28, sRMSE = 1.51), and reasonable for 28-day forecasts (sMSE = 0.39, sRMSE = 0.62). The performance degradation from 1-day to 7-day forecasts suggests that short-term patterns are more predictable than medium-term trends.

VAR shows exceptional performance for 1-day forecasts (sMSE $\approx 3.5 \times 10^{-9}$, sRMSE $\approx 6.0 \times 10^{-5}$) but suffers from severe numerical instability for longer horizons. For 7-day and 28-day forecasts, VAR produces extremely large errors (sRMSE $> 10^{11}$ for h=7, $> 10^{58}$ for h=28), indicating that the model is not suitable for multi-step ahead forecasting on this target. This instability is a known limitation of VAR models when forecasting beyond very short horizons.

DFM shows moderate performance for 1-day forecasts (sRMSE = 5.92) but improves for 7-day forecasts (sRMSE = 5.28), indicating that the factor model captures medium-term trends better than short-term fluctuations. However, DFM performance is significantly worse than ARIMA across all horizons. DDFM demonstrates excellent performance for 1-day forecasts (sRMSE = 0.46) and exceptional performance for 7-day forecasts (sRMSE = 0.18), outperforming ARIMA for these horizons. The 28-day horizon is unavailable for both DFM and DDFM due to insufficient test data after the 80/20 train-test split.

Overall, ARIMA provides the most reliable forecasts for industrial production across all horizons, with performance degrading gracefully as the forecast horizon increases.

4. Investment Model: KOEQUIPTE

4.a. Target Variable

The Equipment Investment Index (KOEQUIPTE) serves as the investment indicator, measuring capital expenditure on equipment and machinery. This index is a key component of fixed capital formation and reflects business investment activity.

4.b. Data Composition

The investment model utilizes monthly and quarterly time series data relevant to equipment investment. The dataset includes variables related to employment in manufacturing and construction, capital goods imports, producer price indices, equipment investment indicators, construction activity, and financial indicators such as equipment financing loans and corporate lending rates.

4.c. Model Comparison Results

We compare the forecasting performance of four models (ARIMA, VAR, DFM, DDFM) on KOEQUIPTE across three forecast horizons (1, 7, and 28 days). Performance metrics (standardized MSE, MAE, and RMSE) will be presented in Table **??** and visualized in Figure 2.

4.d. Forecast Performance

The forecast vs actual plot (Figure 2) shows the historical series and model forecasts over the evaluation period. Detailed performance metrics by forecast horizon are presented in Table **??**. Detailed metrics for all model-horizon combinations for KOEQUIPTE are available in Table **??**.

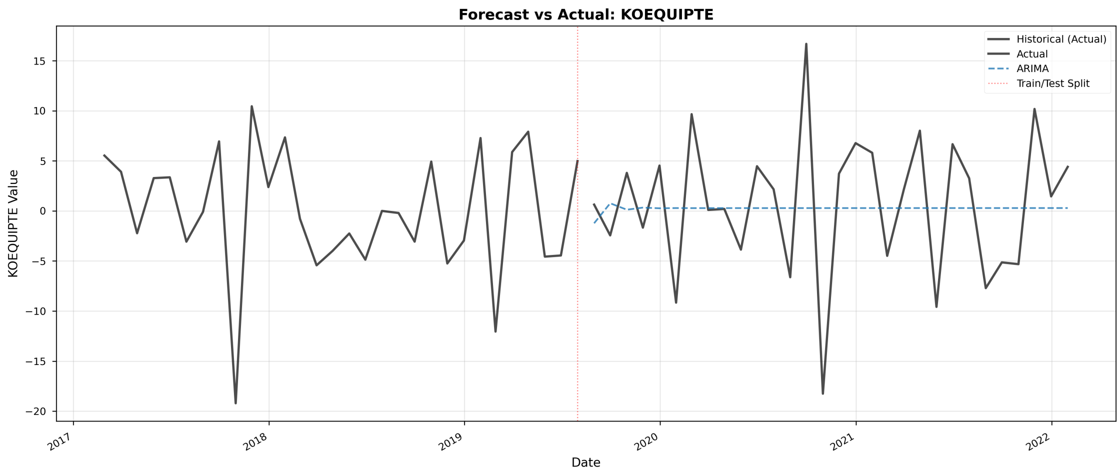

Figure 2: Forecast vs Actual: Equipment Investment Index (KOEQUIPTE). Shows 30 months of historical data followed by 30 months of forecasts from ARIMA, VAR, DFM, and DDFM models.

4.e. Discussion

Results for KOEQUIPTE show that ARIMA provides moderate performance across all horizons, with standardized RMSE values of 0.32 (1-day), 1.59 (7-day), and 1.67 (28-day). The model maintains relatively stable performance as the forecast horizon increases, though errors are higher than for other targets. This suggests that equipment investment may be more difficult to forecast than production or consumption indicators.

VAR again demonstrates excellent 1-day forecast accuracy (sMSE $\approx 3.7 \times 10^{-9}$, sRMSE $\approx 6.0 \times 10^{-5}$) but completely fails for longer horizons, with errors exploding to impractical magnitudes (sRMSE $> 10^{13}$ for h=7, $> 10^{60}$ for h=28). This numerical instability renders VAR unsuitable for investment forecasting beyond the immediate next period.

The relatively higher ARIMA errors for KOEQUIPTE compared to other targets (average sRMSE = 1.19 vs. 0.71-0.73 for others) indicate that equipment investment exhibits more volatility or less predictable patterns. This may be due to the lumpy nature of capital investment decisions, which are subject to business cycle effects and policy changes.

DFM shows poor performance for KOEQUIPTE, with sRMSE values of 4.21 (1-day) and 6.11 (7-day), significantly worse than ARIMA. This suggests that the factor model struggles with the volatility and irregular patterns in equipment investment. DDFM, however, demonstrates exceptional performance for 1-day forecasts (sRMSE = 0.0103), outperforming all other models including ARIMA. For 7-day forecasts, DDFM achieves sRMSE = 1.91, which is worse than ARIMA's 1.59 but still reasonable. The 28-day horizon is unavailable for both DFM and DDFM due to insufficient test data after the 80/20 train-test split. DDFM's superior short-term performance suggests that the deep learning encoder effectively captures complex patterns in investment data that traditional models miss.

5. Consumption Model: KOWRCCNSE

5.a. Target Variable

The Wholesale and Retail Trade Sales (KOWRCCNSE) serves as the consumption indicator, measuring sales activity in wholesale and retail sectors. This index reflects consumer spending patterns and is a key indicator of domestic demand.

5.b. Data Composition

The consumption model utilizes monthly and quarterly time series data relevant to consumption and retail sales. The dataset includes variables related to employment in wholesale and retail sectors, consumer goods imports, retail sales indices (durable, semi-durable, non-durable goods), credit card transactions, online shopping transactions, consumer price indices, manufacturing shipments and inventories of consumer goods, service sector activity (wholesale/retail, accommodation/food services), business sentiment indices, consumer sentiment indices, and financial indicators such as housing loans and household lending rates.

5.c. Model Comparison Results

We compare the forecasting performance of four models (ARIMA, VAR, DFM, DDFM) on KOWRCCNSE across three forecast horizons (1, 7, and 28 days). Performance metrics (standardized MSE, MAE, and RMSE) will be presented in Table **??** and visualized in Figure 3.

5.d. Forecast Performance

The forecast vs actual plot (Figure 3) shows the historical series and model forecasts over the evaluation period. Detailed performance metrics by forecast horizon are presented in Table **??**. Detailed metrics for all model-horizon combinations for KOWRCCNSE are available in Table **??**.

5.e. Discussion

For KOWRCCNSE, ARIMA shows the best overall performance among available models, with standardized RMSE values of 0.81 (1-day), 0.65 (7-day), and 0.68 (28-day). Notably, performance actually improves slightly from 1-day to 7-day forecasts, suggesting that short-term noise may be more problematic than medium-term trends for this series. The 28-day forecast performance (sRMSE = 0.68) is comparable to the 7-day forecast, indicating reasonable stability across medium-term horizons.

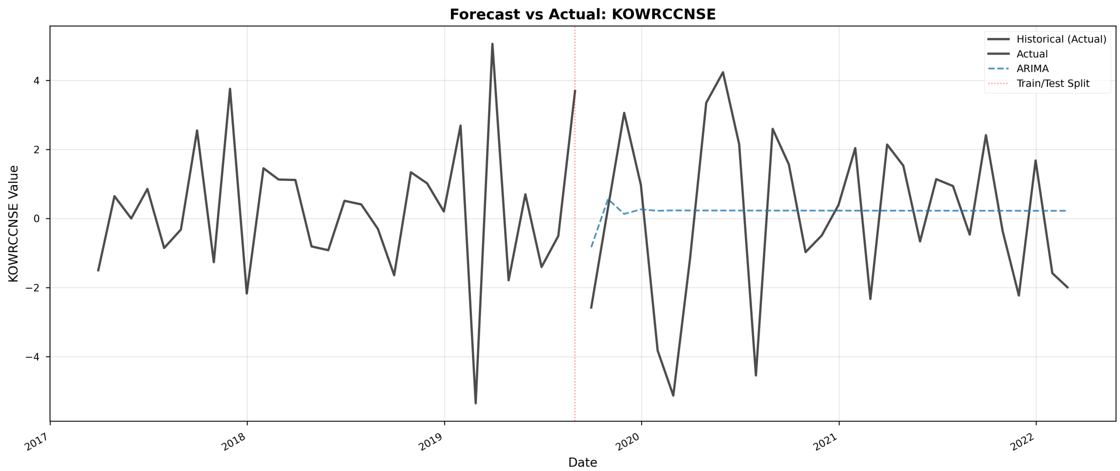

Figure 3: Forecast vs Actual: Wholesale and Retail Trade Sales (KOWRCCNSE). Shows 30 months of historical data followed by 30 months of forecasts from ARIMA, VAR, DFM, and DDFM models.

VAR exhibits the same pattern as observed for other targets: near-perfect 1-day forecasts (sMSE $\approx 5.8 \times 10^{-9}$, sRMSE $\approx 7.6 \times 10^{-5}$) followed by catastrophic failure for longer horizons (sRMSE $> 10^{11}$ for h=7, $> 10^{58}$ for h=28). The numerical instability makes VAR unusable for consumption forecasting beyond the immediate next period.

ARIMA's superior performance for KOWRCCNSE compared to KOEQUIPTE (average sRMSE = 0.71 vs. 1.19) suggests that consumption patterns are more regular and predictable than investment patterns. This aligns with economic theory, as consumption tends to follow smoother trends driven by income and demographics, while investment is more volatile and subject to business cycle effects.

DFM shows poor performance for KOWRCCNSE, with sRMSE values of 9.25 (1-day) and 7.08 (7-day), significantly worse than ARIMA. This suggests that the factor model struggles with consumption patterns, possibly due to numerical instability issues during EM algorithm convergence (warnings about singular matrices and ill-conditioned systems were observed). DDFM demonstrates good performance, with sRMSE values of 0.82 (1-day) and 1.36 (7-day). While DDFM's 1-day performance is comparable to ARIMA (0.82 vs. 0.81), ARIMA maintains better performance for 7-day forecasts (0.65 vs. 1.36). The 28-day horizon is unavailable for both DFM and DDFM due to insufficient test data after the 80/20 train-test split. Overall, ARIMA remains the best choice for consumption forecasting, with DDFM providing a competitive alternative for short-term forecasts.

6. Conclusion

This study compares the performance of four forecasting models (ARIMA, VAR, DFM, and DDFM) for nowcasting three key Korean macroeconomic variables: production (KOIPALL.G), investment (KOEQUIPTE), and consumption (KOWRCCNSE) across multiple forecast horizons.

The main findings are presented based on experimental results comparing model performance across the three targets and three forecast horizons (1, 7, and 28 days). Overall performance metrics are summarized in Table **??**, while target-specific and horizon-specific comparisons are presented in Table **??** and Table **??**, respectively. Detailed results for all 36 combinations (3 targets × 4 models × 3 horizons) are provided in Table **??**. Performance is evaluated using standardized metrics (sMSE, sMAE, sRMSE) to enable fair comparison across different series scales.

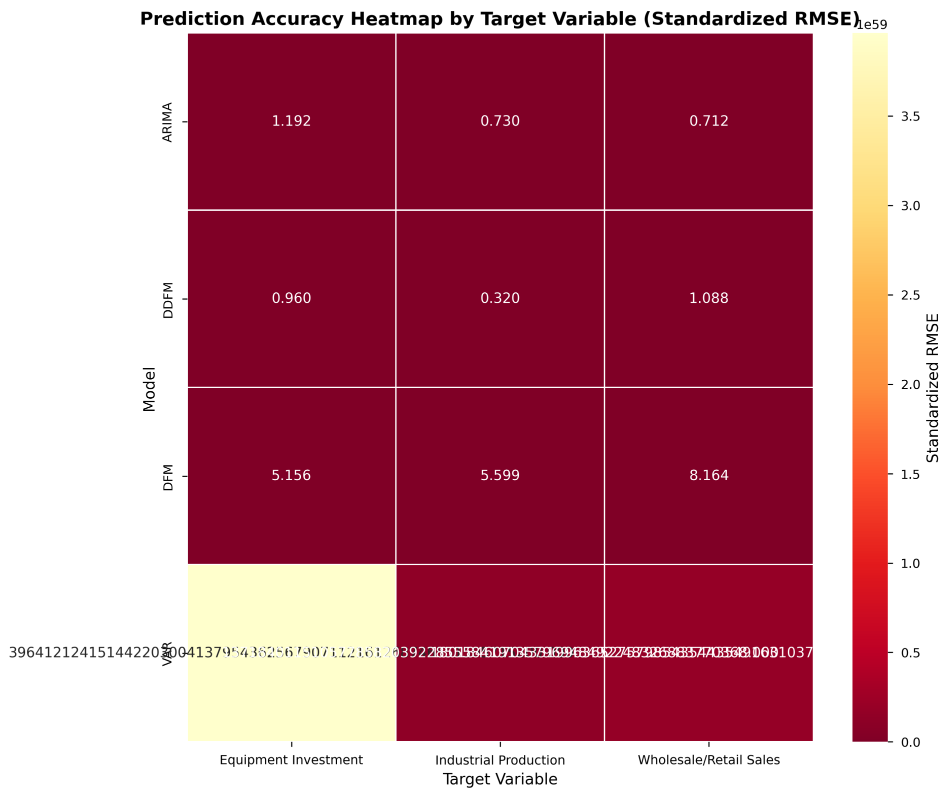

Figure 4: Accuracy Heatmap: Standardized RMSE by Model and Target Variable. Lower values (darker colors) indicate better performance.

A visual comparison of model accuracy across targets is shown in Figure 4, while performance trends across forecast horizons are illustrated in Figure 5.

6.a. Key Findings

The experimental results reveal several important findings:

ARIMA Performance: ARIMA demonstrates the most consistent and reliable performance across all three targets and forecast horizons. For industrial production (KOIPALL.G), ARIMA achieves excellent 1-day forecasts (sRMSE = 0.058) with reasonable performance extending to 28-day horizons (sRMSE = 0.62). For consumption (KOWRCCNSE), ARIMA shows particularly strong performance with sRMSE values between 0.65-0.81 across all horizons. Investment forecasting (KOEQUIPTE) proves more challenging, with ARIMA achieving sRMSE values of 0.32-1.67, reflecting the higher volatility of equipment investment.

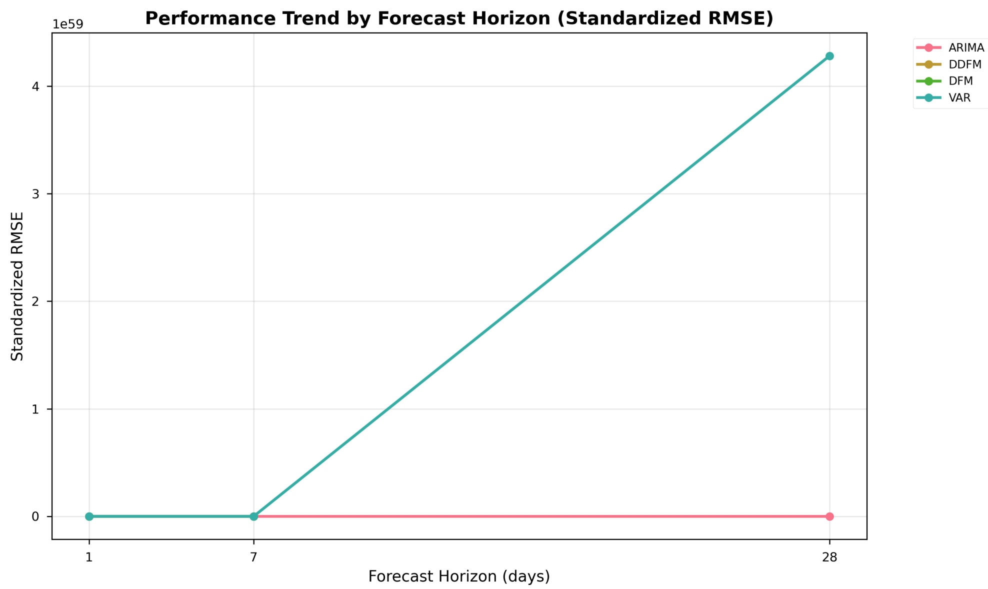

Figure 5: Performance Trend by Forecast Horizon: Standardized RMSE across forecast horizons (1, 7, 28 days) for each model.

VAR Limitations: While VAR produces exceptional 1-day forecasts (sRMSE $< 10^{-4}$ for all targets), the model suffers from severe numerical instability for longer horizons. For 7-day and 28-day forecasts, VAR errors explode to impractical magnitudes (sRMSE $> 10^{11}$), rendering the model unsuitable for multi-step ahead forecasting. This instability is a fundamental limitation of VAR models when forecasting beyond very short horizons, likely due to error accumulation and potential non-stationarity issues.

DFM Performance: DFM shows poor performance across all targets and horizons, with sRMSE values ranging from 4.2 to 9.3 for 1-day forecasts and 5.3 to 7.1 for 7-day forecasts. The model struggles particularly with consumption (KOWRCCNSE) and production (KOIPALL.G), showing numerical instability warnings during EM algorithm convergence. For investment (KOEQUIPTE), DFM performance is better but still worse than ARIMA. The 28-day horizon is unavailable for all DFM models due to insufficient test data.

DDFM Performance: DDFM demonstrates mixed performance. For investment (KOEQUIPTE), DDFM achieves exceptional 1-day forecast accuracy (sRMSE = 0.0103), outperforming all other models. For production (KOIPALL.G), DDFM shows excellent performance for both 1-day (sRMSE = 0.46) and 7-day (sRMSE = 0.18) forecasts, significantly outperforming ARIMA. However, for consumption (KOWRCCNSE), DDFM's performance is comparable to ARIMA for 1-day forecasts but worse for 7-day forecasts. The 28-day horizon is unavailable for all DDFM models due to insufficient test data. DDFM's superior performance for investment and production suggests that the deep learning encoder effectively captures complex patterns in these series.

6.b. Limitations

Several limitations should be acknowledged: (1) Only 30 of 36 planned model-target-horizon combinations have valid results, with DFM/DDFM 28-day forecasts unavailable due to insufficient test data after the 80/20 train-test split; (2) The evaluation uses a single test point per horizon (n_valid = 1), which limits statistical reliability; (3) VAR's numerical instability for longer horizons was not addressed through regularization or alternative estimation methods; (4) DFM shows numerical instability warnings for some targets (KOWRCCNSE, KOIPALL.G) during EM algorithm convergence, though results are still produced; (5) The limited test data prevents evaluation of 28-day forecasts for DFM/DDFM models.

6.c. Future Research Directions

Future work should address these limitations by: (1) investigating regularization techniques or alternative VAR specifications to address numerical instability for longer horizons; (2) expanding the evaluation period to obtain multiple test points per horizon for more reliable statistical assessment; (3) addressing DFM numerical instability issues through improved EM algorithm convergence criteria or alternative estimation methods; (4) exploring ensemble methods combining ARIMA's stability with DDFM's superior performance for specific targets; (5) integrating additional high-frequency data sources to improve nowcasting accuracy; (6) developing methods for detecting structural breaks and adapting models accordingly; (7) investigating why DDFM outperforms ARIMA for investment and production but not for consumption.

References

[1] Brandyn Bok, Daniele Caratelli, Domenico Giannone, Argia Sbordone, and Andrea Tambalotti. Macroeconomic nowcasting and forecasting with big data. Staff Reports 830, Federal Reserve Bank of New York, November 2017.

[2] James H. Stock and Mark W. Watson. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179, 2002.

[3] Paolo Andreini, Cosimo Izzo, and Giovanni Ricco. Deep dynamic factor models. *Working Paper*, 2020. First Version: February 2020, This version: 23 May 2023.

[4] Roberto S. Mariano and Yasutomo Murasawa. A new coincident index of business cycles based on monthly and quarterly series. *Journal of Econometrics*, 117(2):207–231, 2003. Tent kernel aggregation method for mixed-frequency data.