ITALIAN INDUSTRIAL PRODUCTION

a time series forecasting project

Course: Time Series and Forecasting** (a.a. 2025/26)

Teacher: *Andrea Bastianin*

Students: *Paolo Minini (*53969A),

Francesco Sebastiano Memmola (50893A),

Academic year: 2025/2026

1. Introduction

This study forecasts the **Italian Industrial Production Index (IPI)** using monthly data from 2000 to 2022, limitation due to the limited data found in the exogenous variables. Industrial production is a volatile but timely indicator of real economic activity, and accurate short-term forecasts are crucial for an economy like Italy's, which is strongly manufacturing-based.

The sample includes two major structural breaks, the **2008 Financial Crisis** and the **2020 COVID-19 collapse**, which make short-term forecasting particularly challenging. These breaks make our target variable an interesting target to forecast, in order to analyze whether adding model complexity allows us to capture such unpredictable changes in the economy.

The central question of this project is whether incorporating richer macroeconomic information can improve one-step-ahead forecasts of Italian industrial production relative to very simple time-series benchmarks. We investigate whether variables widely recognized as informative predictors for both domestic and global industrial dynamics contain predictive content beyond what is already embedded in past IPI itself.

2. Data and dataset structure

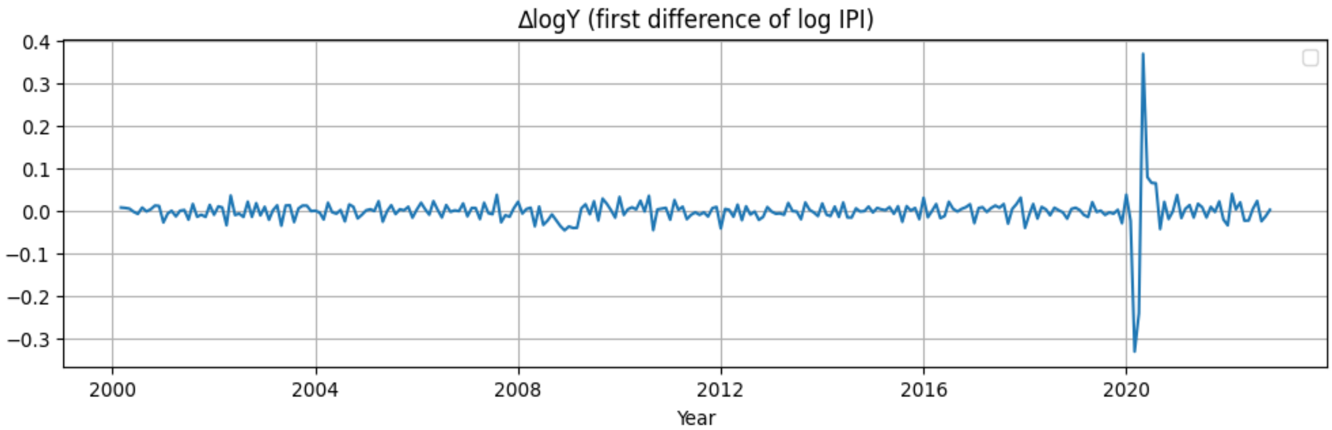
The data consists of **monthly observations from 2000 to 2022**. The target variable is the **Italian Industrial Production Index (IPI)**, which serves as a high-frequency indicator of real economic activity. The exogenous variables used throughout the study include: the **Producer Price Index (PPI)**, the **unemployment rate**, **Brent crude oil prices**, and the **OECD Composite Leading Indicator (CLI)**, all of which are widely recognized as informative predictors for both domestic and global industrial dynamics.

The IPI and the set of core explanatory variables were collected from *FRED*, ensuring consistency in frequency, coverage, and data quality.

The Italian Industrial Production Index exhibits no consistent long-term upward trend but is instead dominated by severe structural breaks corresponding to the 2008 Financial Crisis and the 2020 COVID-19 pandemic. Visual inspection confirms the series is seasonally adjusted but remains non-stationary in levels.

Transformations and stationarity

Before proceeding to the modelling stage, particular attention was devoted to **transformations and stationarity**. Augmented Dickey-Fuller test indicated that most series were non-stationary in levels, which is expected given the strong trends and the structural shifts observed over the sample. To address this, each variable was transformed according to its statistical properties and economic interpretation.



The IPI series in **log-levels** displays strong persistence and clear evidence of a stochastic trend, consistent with the results of the Augmented Dickey–Fuller test, which indicates **non-stationarity in levels** but **stationarity in log-differences**.

Similar results emerge for most macroeconomic predictors: oil prices are non-stationary in log-levels but stationary in **log-differences**, PPI becomes stationary when expressed in **monthly growth**, unemployment requires **first differencing**, and the CLI becomes stationary after a **log transformation**.

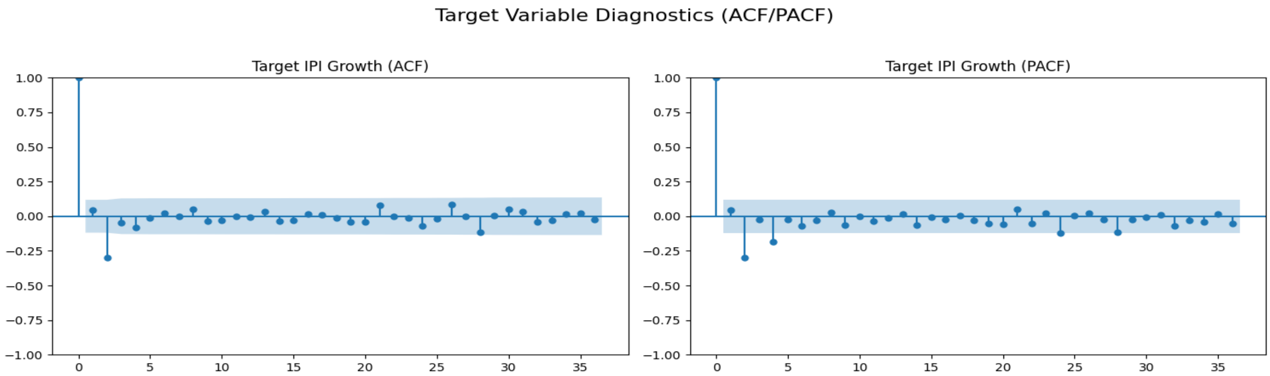
These transformations stabilize variance and eliminate trends, allowing ARIMA and VAR models to be estimated correctly.

To address major structural breaks in the sample, two deterministic dummy variables are included: an **impulse dummy for the COVID-19 shock**, capturing the temporary collapse and rebound in 2020, and a **shift dummy for the 2008 Financial Crisis**, reflecting its long-lasting impact on industrial activity. These controls help stabilize the coefficients of the multivariate models.

Following standard forecasting practice, the dataset is divided into a **training sample (2000–2018)** for model estimation and a **test sample (2018–2022)** for out-of-sample evaluation. All results are based on **1-step-ahead recursive forecasts** over the test period. All models are evaluated exclusively on **1-step-ahead forecasts** over the test period.

ACF/PACF Analysis of the target variable for models' choice

A closer inspection of the log-differenced IPI series, using its ACF and PACF, reveals no signs of residual monthly seasonality.



Crucially, the short-term dynamics exhibit a lack of immediate memory, as evidenced by an insignificant correlation at Lag 1, contrasting with a significant spike at Lag 2. This lack of first-order persistence suggests that the series shares key characteristics with a Random Walk process. Consequently, a simple Random Walk emerges as a plausible baseline candidate, to be formally evaluated against autoregressive specifications that attempt to capture the delayed correction observed at Lag 2.

Factor-model dataset

To incorporate broader macroeconomic conditions, we extend the dataset using a large panel of European indicators obtained from Barigozzi.eu. This data set includes monthly data of 31 variables. These include labor market statistics exchange rates and shares indices, interest rates and monetary aggregates, industrial turnover and producer prices broken down by main industrial groupings, consumer price indices, and various confidence indicators for business and consumers.

After aligning the panel to our IPI sample and removing series with missing observations, all variables are standardized and transformed to achieve stationarity. We then apply principal component analysis (PCA) to extract the common factors, with the first factor capturing the dominant co-movements across the macroeconomy and serving as the input to the Dynamic Factor Models, that will be used as alternative methods.

3. Methods and Models

All models are evaluated using a **recursive expanding-window design**. The initial estimation sample includes roughly the first 80% of the observations (2000–2018), and **one-step-ahead forecasts** are produced for the remaining portion of the sample (2018–2022). At each forecast origin, the model is re-estimated to incorporate all information available up to that date. This procedure mirrors real-time forecasting conditions and allows model coefficients to adjust gradually to the structural instabilities characterizing the sample.

We begin with **univariate benchmark specifications**, including a random walk without drift, a random walk with drift, and selected ARIMA models for the log level of IPI. We then introduce two alternative multivariate structures: a **VAR-X model**, which incorporates a small set of macroeconomic predictors, and a **factor-augmented model (FAVAR-X)** constructed through principal component analysis applied to the larger macroeconomic dataset. Together, these models span a wide range of predictive structures, allowing us to evaluate whether additional economic information improves forecast accuracy relative to simple stochastic-trend benchmarks.

3.1 Benchmark models

Our main benchmark is the **Random Walk (RW)** for the log industrial production index $y_t = \log Y_t$:

$$y_t = y_{t-1} + \varepsilon_t,$$

which states that the best forecast of tomorrow's log level is today's level. We also consider a version **with drift**,

$$y_t = \mathrm{c} + y_{t-1} + \varepsilon_t,$$

but the estimated constant c is very small, so the two specifications are practically indistinguishable.

To allow for richer short-run dynamics while remaining in a univariate framework, we estimate several **ARIMA(p,1,q)** models for y_t. According to information criteria and according to ACF/PACF analysis the two most promising specifications are:

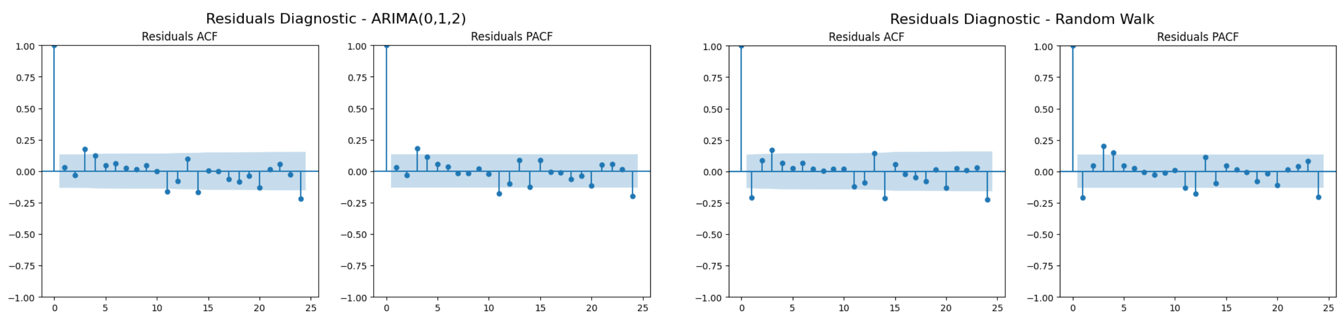
- **ARIMA(0,1,2)**:

$$\Delta y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$

- **ARIMA(2,1,0)**:

$$\Delta y_t = \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \varepsilon_t$$

For the ARIMA class we select the ARIMA(0,1,2) specification. Residual diagnostics are satisfactory: the t-test does not reject a zero mean, the Ljung–Box test fails to reject the null of no autocorrelation up to lag 12, the residual ACF/PACF stay within the 95% confidence bands, and all roots lie inside the unit circle. Taken together, these results indicate that ARIMA(0,1,2) provides a coherent in-sample representation of the differenced series.



However, when we move to forecasting performance, a different picture emerges. Although AIC and BIC both favour ARIMA(0,1,2) in-sample, out-of-sample the lowest RMSE and MAE on the test set are achieved by the Random Walk without drift. Consequently, we use ARIMA(0,1,2) as the reference ARIMA model for reporting residual diagnostics, but we treat the Random Walk as our benchmark forecasting model and express all accuracy comparisons relative to it.

3.2 VAR-X models

To examine whether macroeconomic variables help forecast industrial production, we estimate a **VAR-X(p)** model for a small system of monthly indicators. The endogenous vector is:

$$y_t = \begin{bmatrix} \text{IPI growth}_t \\ \text{PPI growth}_t \\ \text{Unemployment}_t \\ \text{Oil price growth}_t \\ \text{CLI}_t \end{bmatrix}$$

where $D_{\text{COVID},t}$ and $D_{2008,t}$ are dummy variables added like **exogenous variables**, which capturing the Covid-19 and Global Financial Crisis episodes.

Before estimating the forecasting models, we run **Granger causality tests** between IPI growth and each predictor. The tests show that unemployment, oil prices and the CLI significantly Granger-cause industrial production, and in several cases industrial production also Granger-causes these variables. This bidirectional causality violates the strict exogeneity assumption of single-equation AR-X models and justifies the use of a **system-based VAR-X framework**.

The general VAR-X(p) specification is

$$y_t = \text{c} + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \text{B}x_t + \varepsilon_t$$

with c a vector of constants, Φ_k the $\text{lag} - \text{k}$ coefficient matrices, B the coefficients on the exogenous dummies and ε_t a white-noise vector.

Lag order is selected on the training sample using the **Akaike Information Criterion**. An initial unrestricted VAR-X suggested p=4 as the AIC-optimal lag length. Since simpler models tend to generalize better, we retain two specifications:

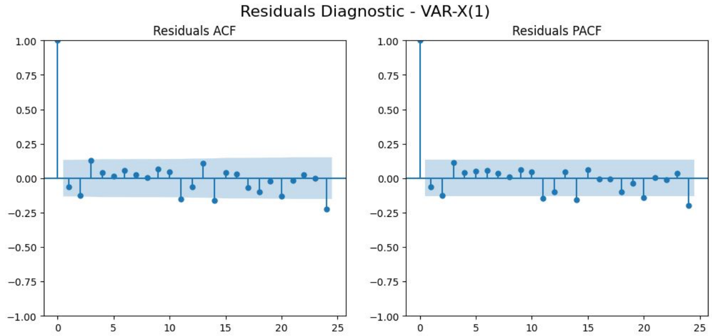
- **VAR-X(1)**:

$$y_t = \text{c} + \Phi_1 y_{t-1} + \text{B}x_t + \varepsilon_t$$

- **VAR-X(4)**:

$$y_t = \text{c} + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \Phi_3 y_{t-3} + \Phi_4 y_{t-4} + \text{B}x_t + \varepsilon_t$$

Diagnostics for both VAR-X specifications on the training sample show residuals with zero mean and no significant autocorrelation till lag 12; residual ACF/PACF lie within confidence bounds, and the stability test confirms that all eigenvalues are inside the unit circle. Because VAR-X(1) is much more parsimonious and performs better out of sample than VAR-X(4), we retain **VAR-X(1)** as our preferred multivariate alternative in the final evaluation.



3.3 Dynamic factor (FAVAR-X) models

To incorporate a richer set of macroeconomic indicators without over-parameterizing the model, we construct a **dynamic factor model**.

Forecast experiments with one, two and three factors indicate that a single factor performs best out of sample. We therefore retain **one common factor** $\widehat{F}_{1,t}$.

The factor-augmented models are:

- **Dynamic Factor VAR(1)** (FAVAR):

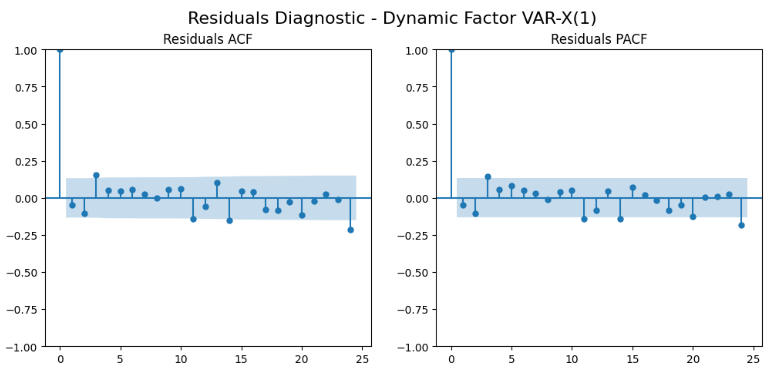
$$y_t = c + \rho_1 \; y_{t-1} + \lambda_1 \widehat{F}_{1,t-1} + \varepsilon_t$$

- **Dynamic Factor VAR-X(1)** (FAVAR-X with dummies):

$$y_t = c + \rho_1 \, y_{t-1} + \lambda_1 \widehat{F}_{1,t-1} + \beta_1 D_{\text{COVID},t} + \beta_2 D_{2008,t} + \varepsilon_t$$

As with the VAR-X, we check residual diagnostics: t-tests confirm zero-mean residuals, Ljung–Box statistics do not reject the null of no autocorrelation, residual ACF/PACF are consistent with white noise, and the VAR stability tests indicate that all roots lie inside the unit circle. These results support the FAVAR-X(1) as a coherent high-dimensional alternative.

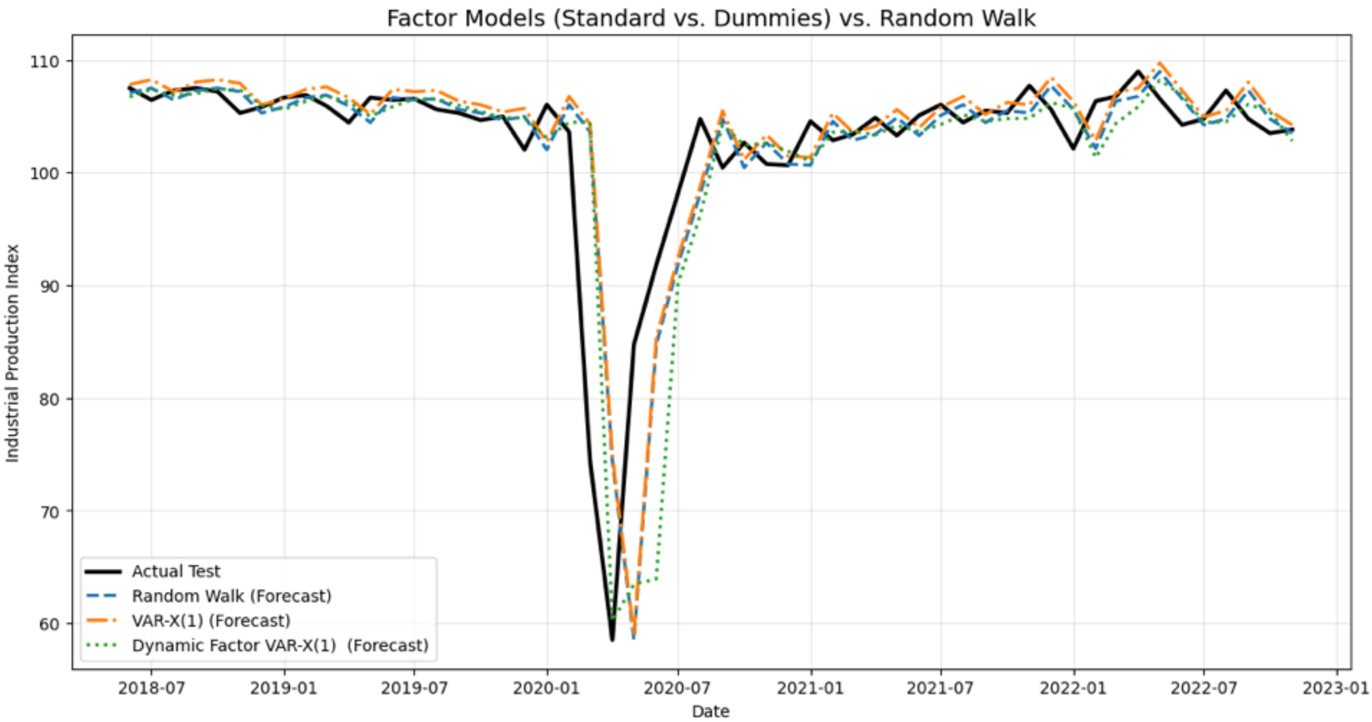
Given the limited length of the test sample and the already high dimensionality absorbed by PCA, we deliberately keep the factor model **parsimonious** (one factor, one lag). Additional lags are not supported by AIC/BIC and do not improve forecast accuracy.



4. Results

We now present the results of the forecasting exercise, focusing on one-step-ahead point forecasts produced by the benchmark and alternative models. All models are evaluated under the same recursive expanding-window scheme described in Section 3: the first 80% of the sample (2000–2018) is used for initial estimation, and forecasts are generated for the remaining 20% (2018–2022).

4.1 Forecasts and actual series



The figure plots the realised IPI (black line) together with one-step-ahead forecasts from:

- the **Random Walk** benchmark,
- the **VAR-X(1)** model with unemployment, inflation, oil prices and CLI, and
- the **Dynamic Factor** VAR-X(1) (FAVAR-X) model that augments the system with a PCA factor extracted from the broader macro dataset.

In the pre-Covid period (2018–2019) all three models track the level of industrial production quite closely, with only small deviations around turning points. The crucial episode is the Covid-19 collapse in early 2020: none of the specifications is able to anticipate the sudden drop. Forecasts remain near the pre-crisis path and then adjust only after the first large negative observations are realised, so the initial fall and the start of the rebound are systematically missed.

Overall, the graph illustrates that differences between models are modest compared with their common inability to predict an unprecedented shock like Covid-19, which is exactly the limitation we discuss in the conclusions.

4.2 Evaluation metrics

Forecast accuracy is assessed using the **root mean squared forecast error (RMSE)** and the **mean absolute forecast error (MAE)** computed on the common out-of-sample window. In addition, we report RMSE and MAE ratios relative to the Random Walk benchmark (so that a value above one indicates worse performance than the benchmark).

Model	**RMSFE**	**MAFE**	**RMSFE ratio**	**MAFE ratio**
Random walk	6.2359	3.0395		
VAR-X(1)	6.3049	3.1607	1.0111	1.0399
Dynamic Factor VAR-X(1)	6.7165	3.0884	1.0771	1.0161

Two main messages emerge:

1. The **Random Walk benchmark remains hard to beat**. It achieves the lowest RMSE and MAE, confirming that short-horizon dynamics of industrial production are dominated by local persistence and peculiar shocks that are well captured by the naive “tomorrow ≈ today” rule.
2. The **VAR-X(1)** and **FAVAR-X(1)** models perform **very similarly**, but neither delivers systematic gains over the benchmark. VAR-X(1) yields a slightly lower RMSE than the factor model, while FAVAR-X(1) attains a marginally lower MAE than VAR-X(1), although both error ratios remain above one.

Overall, the additional macroeconomic information, either in the form of a small VAR-X system or condensed into a single factor, does not translate into lower one-step-ahead forecast errors relative to the Random Walk.

4.3 Interpretation of results

The ranking of models is consistent with the characteristics of the data and the shocks hitting the Italian economy over the evaluation period.

First, the **high persistence** of industrial production makes the Random Walk a very strong benchmark. Over short horizons, most of the predictive content is contained in the most recent observation, and simple persistence tends to outperform more parameter-rich specifications.

Second, while the **VAR-X(1)** model passes standard diagnostic checks and exploits economically meaningful predictors (unemployment, prices, oil, CLI and crisis dummies), these variables mainly help to explain medium-term co-movements rather than to improve very short-run forecasts month by month. The gains in fit are largely absorbed in-sample, with little benefit for out-of-sample one-step-ahead accuracy.

Third, the **Dynamic Factor VAR-X(1)** model incorporates a much richer set of macro indicators through the PCA factor. This factor captures broad business-cycle conditions and should produce forecasts that track the Covid-19 collapse and recovery reasonably well. However, the additional complexity does not lead to lower forecast errors: unexpected, extreme shocks such as the Covid-19 downturn are largely unpredictable from past macro information, even when summarized by a factor.

Conclusion

The project asks whether adding macro information improves one-step-ahead forecasts of Italian industrial production. Using monthly data from 2000–2022 and an expanding-window scheme, we compare a Random Walk with two multivariate models: VAR-X(1) with macro controls and dummies, and a Dynamic Factor VAR-X(1) built from PCA.

The Random Walk achieves the lowest RMSE and MAE, while both VAR-based models pass diagnostics but deliver slightly higher errors, indicating that at this horizon most predictive content is already in the latest IPI observation, that indicate no systematic gain over the naïve “tomorrow = today” rule. This suggests that, at the monthly one-step-ahead horizon, most predictive content is already contained in the latest observation of industrial production.

Nevertheless, the richer models remain useful from an economic perspective: macro variables, oil prices, business-cycle indicators and crisis dummies help clarify how industrial production co-moves with broader conditions and how major shocks (2008 and Covid-19) affect the dynamics, even if this extra structure does not translate into better short-run forecasts.

The failure of such models is mainly related to the fact that the test set at our disposal is dominated by large shocks (Covid-19) that standard linear models cannot anticipate. As a result, the potential benefits of richer multivariate or factor models are largely offset by the sudden structural break.

These considerations suggest that our conclusions mainly apply to “normal times” and to very short horizons. Different forecast horizons or a longer and calmer evaluation period could lead to a different model ranking. Future work could therefore consider richer dynamic factor or machine-learning approaches that better handle nonlinearities and structural breaks.