1. Introduction

In a globalised and rapidly evolving economy, accurate forecasting of economic growth is essential for guiding monetary policy, anticipating market trends, and informing business strategies. Post-pandemic disruptions, policy shifts, and geopolitical tensions have amplified uncertainty, increasing the importance of real-time indicators. Industrial Production (INDPRO) serves as a key high-frequency proxy for real activity, offering timely insights into the state of the economy for central banks, traders, and corporate decision-makers.

Despite its importance, forecasting industrial production growth remains a difficult task. Macroeconomic data are complex, noisy, and affected by shocks and spillovers that make accurate prediction challenging. This paper evaluates a range of machine learning (ML) models using FRED-MD data to forecast industrial production growth and identify models best suited for forecasting in volatile, data-rich environments.

1.1. Literature Review

Forecasting real economic activity has long been central to macroeconomic analysis and policy design, with industrial production playing a vital role as a high-frequency proxy for output. The expansion of large macroeconomic panels, such as FRED-MD, has enabled researchers to move beyond simple models and exploit many predictors simultaneously.

Much of the literature focuses on managing high-dimensional predictor spaces. Penalized regression techniques have been proposed for variable selection and shrinkage in macro panels, with evidence showing these approaches perform robustly in high-dimensional forecasting problems, making them useful econometric benchmarks (Smeekes & Wijler, 2018). Factor models and other dimensionality-reduction techniques also remain popular for their regularization and interpretability benefits.

Recent studies have increasingly explored ML methods in macroeconomic forecasting. Comparative evidence suggests that ML models can yield meaningful gains when nonlinearities and complex interactions are present, with nonlinearity identified as a key source of improvement (Coulombe et al., 2020). Empirical work on monthly targets such as inflation finds that flexible, tree-based ensembles like random forests outperform classical univariate and linear multivariate benchmarks in certain environments, provided that cross-validation and horizon-specific tuning are carefully implemented (Medeiros et al., 2019). Building on these insights, this paper extends existing ML-based forecasting frameworks, previously applied to inflation and other macro variables, to the prediction of industrial production growth.

1.2. Executive Summary

This study evaluates a range of econometric and machine learning models for forecasting industrial production growth using the FRED-MD dataset. Among the models tested, Post-LASSO* achieved the strongest out-of-sample (OOS) performance with an R^2 = 0.153, MSE = 2.43, MAE = 0.775 outperforming both linear benchmarks and nonlinear ensemble methods. Gradient Boosting and Random Forests followed with R^2 = 0.040 and R^2 = 0.013, while traditional linear models such as AR(3) and ADL produced negative R^2 values (–0.304 and –0.325), indicating poor predictive power relative to a mean forecast.

2. Methodology

This study evaluates multiple econometric and machine learning approaches for forecasting monthly U.S. industrial production growth (Δlog(INDPRO)) using the FRED-MD dataset. Two traditional models, AR(3) and ADL, serve as linear benchmarks to capture autoregressive and lagged predictor dynamics, respectively.

We then estimate a series of regularized regression models to address high dimensionality and multicollinearity in the predictor set. These include LASSO, Adaptive LASSO, Elastic Net, and Post-LASSO. The LASSO and Adaptive LASSO impose $L1$-based penalties for variable selection, while Elastic Net combines $L1$and $L2$ penalties to balance sparsity and shrinkage. Post-LASSO refits an ordinary least squares (OLS) model on variables selected by LASSO to reduce shrinkage bias.

For nonlinear learners, we apply Random Forests and Gradient Boosting to capture interactions and nonlinear effects among predictors. Both are ensemble methods based on regression trees: Random Forests average over many decorrelated trees to reduce variance, while Gradient Boosting sequentially builds weak learners to minimize residual error.

All models are estimated in R using standard libraries: glmnet for penalized regressions, randomForest and ranger for Random Forests, and xgboost for Gradient Boosting. The dataset is processed with dplyr, tidyr, zoo, and lubridate; evaluation metrics (MSE, MAE, R^2) are computed using metrics. Robust standard errors are derived using the sandwich and lmtest packages to ensure valid inference under potential heteroskedasticity. Models are estimated recursively using expanding training windows to replicate real-time forecasting conditions. Hyperparameters are tuned via blocked cross-validation or the Bayesian Information Criterion (BIC), depending on model type. This unified framework allows direct comparison between linear, regularized, and nonlinear methods in terms of predictive accuracy and robustness to macroeconomic volatility.

2.1. Benchmark Models

Given time series data, 2 baseline models were employed: the Autoregressive (AR) model and the Autoregressive Distributed Lag (ADL) model.

The general formulation of an AR process with p lags is:

$$y_t = \alpha + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t$$

To determine the optimal lag length p* for the AR(p) model, specifications with up to six lags were estimated and evaluated using the Bayesian Information Criterion (BIC). As the main goal of this study is forecasting, BIC was selected for model selection due to its stronger penalisation of model complexity, which helps mitigate overfitting and enhances OOS predictive performance.

The general formulation of an ADL process with p lags of the target variable and q lags of auxiliary variables is:

$$y_t = \alpha + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=0}^{q} \beta_j x_{t-j} + \varepsilon_t$$

The ADL model extends the AR(p) framework by incorporating lagged values of additional explanatory variables. Four key predictors were selected: `BAA_AAA` (credit spread), `HOUST` (housing starts), `T10Y3M` (term spread), and `UMCSENTx` (consumer sentiment). We include 4 lags for each variable, reflecting quarterly effects and allowing for dynamic transmission of macroeconomic shocks over time.

2.2 LASSO

2.2.1. Standard LASSO

LASSO is well-suited for modeling the FRED-MD dataset due to its ability to handle high-dimensional macroeconomic data and automatically conduct variable selection. This allows us to systematically identify the predictors most relevant to industrial production growth. Macroeconomic time series often exhibit multicollinearity and noise, which can compromise the stability and generalisability of OLS, leading to biased and inconsistent OLS estimates. By imposing an L1 penalty, LASSO reduces model complexity by shrinking some estimates to 0, and facilitates an improved bias–variance tradeoff, enhancing OOS predictive performance. Furthermore, LASSO remains computationally efficient even with a large number of lagged variables, making it a practical and robust method for forecasting with the FRED-MD dataset.

Using $\alpha = 1$, we applied L1 regularisation to shrink coefficients toward zero and perform variable selection. The model was trained using predictors, lagged variables, and engineered features. The optimal shrinkage parameter λ was selected based on the BIC and cross validation (CV). Our choice of BIC over the Akiake Information Criterion is due to the stronger penalty applied in BIC, leading to more conservative models. We used rolling windows CV and the λ was chosen using the 1SE rule. Variables with nonzero coefficients were retained. Models with different lag lengths were compared, and the maximum lag was chosen where the OOS R^2 was highest.

2.2.2. Post-LASSO

Post-LASSO employs a two-step procedure. We first identify the relevant predictors with LASSO using both λ values from CV and BIC. Then, we fit the data with an OLS model using only these selected variables, without any additional penalisation. This approach allows for the elimination of potential bias reduction that may result from the shrinkage inherent in LASSO estimates.

2.2.3. Adaptive LASSO

Adaptive LASSO also employs a two-step procedure. We first use ridge (L2) regression to obtain the initial estimates on the predictor coefficients. Then, the L1 penalty weights are then set inversely proportional to the absolute values of the ridge (L2) coefficients, penalizing important predictors less.
Adaptive weighting helps to create sparser models and allows for variable selection. By modifying the regularisation behaviour for each variable, the adaptive penalty addresses what LASSO falls short of, which is the tendency to shrink large coefficients too much and small coefficients too little. Finally we use 10-fold CV to select the optimal λ to provide a more accurate estimate for OOS performance.

2.3 Elastic Net

Elastic net is used as it combines the strengths of both LASSO and ridge regression, making it more effective when dealing with datasets that have highly correlated predictors or when the number of predictors is large relative to the number of observations. It is also more robust than lasso in high dimensional settings and when multicollinearity is present. It can be shown that the estimate resulting from elastic net is a linear combination of the L1 and L2 regression estimates.

The model employs two primary hyperparameters for tuning: α (which controls the balance between ridge and lasso regularisation) and λ (which determines the overall regularisation strength). A grid search was conducted over [0,1] at 0.1 intervals for α. For each α, rolling CV with blocked folds was performed over the final 60 observations of the training set to respect temporal dependencies in the data. The MSE was computed for each α, and the optimal α was selected by the lowest MSE.

Using the selected α, a ridge-like elastic net model was fitted, with blocked CV applied to determine the optimal λ. The λ_min (corresponding to the minimum CV error) was chosen for OOS forecasting.

2.4 Random Forest

RF is an alternative to elastic net focusing on understanding the complex non-linear relationships and interactions among the predictors that may not be well-represented by linear models. It is less sensitive to outliners and noisy data as compared to linear models, and also allows for measures of predictor importance, to identify the important predictors, but without direction or magnitude like elastic net.

A hyperparameter grid was constructed for tuning the RF model, comprising the following parameters:

- mtry: the number of randomly selected features at each split (evaluated at √p and 2√p, where p is the total number of features)
- min.node.size: minimum number of observations in terminal nodes (set to 5 or 10)
- num.trees: fixed at 300 for cross-validation runs

A rolling CV procedure was implemented, where the model was iteratively fitted on expanding training windows, generating one-step-ahead forecasts at each fold. The MSE from these forecasts was used as the performance criterion for each hyperparameter combination. The configuration yielding the lowest rolling CV MSE was selected as optimal. Using the best hyperparameters, a final RF model was trained on the full training dataset with 500 trees and permutation-based variable importance enabled.

2.5 Gradient Boosting

Gradient Boosting (GB) is an ensemble learning method that builds models sequentially, where each new tree corrects the residual errors of the previous ensemble. Unlike Random Forests, which average over many independent trees to reduce variance, Gradient Boosting aims to minimize bias by iteratively improving model fit through gradient descent on a specified loss function. This makes it particularly suitable for capturing subtle nonlinear relationships and interaction effects in macroeconomic data.

A grid search was conducted to tune the key hyperparameters:

- learning rate (η): controls the contribution of each tree to the final prediction (tested at 0.05, 0.1, 0.2)
- max_depth: maximum depth of each regression tree (set between 2 and 4 to prevent overfitting)
- min_child_weight: minimum sum of instance weights in a leaf node (tested at 5, 10, 15)
- subsample and colsample_bytree: fractions of observations and features randomly sampled for each tree (varied between 0.5 and 0.8)

Rolling cross-validation was implemented using expanding training windows to respect temporal ordering. The number of boosting rounds was determined using early stopping based on out-of-sample mean squared error (MSE). The configuration minimizing validation MSE was selected as optimal.

The final model was trained using the xgboost package in R with the selected parameters, producing a sequence of shallow trees that collectively formed a robust predictor. Regularization through shallow depth, shrinkage, and subsampling was used to avoid overfitting and enhance generalization performance.

3. Data

The primary data source used for this research is the monthly Federal Reserve Economic Data (FRED-MD). The period under observation is from 1987 to 2025 with 569 observations and 126 predictors. The training and test sets were split by time using a cutoff at 2016-01-01.

3.1. Data Cleaning

To clean the dataset, the Industrial Production: Total Index (INDPRO) was transformed into monthly growth rates using the formula:

$$INDPRO\_growth = 100 \times \frac{INDPROt - INDPROt-1}{INDPROt-1},$$

which serves as the dependent variable for this analysis. The first observation, for which the lag is unavailable, was removed. Missing values in each column were identified and imputed as appropriate. The sasdate column was converted to a date format, and the variable ACOGNO was dropped due to excessive missing values and discontinuation after 2001. The sample period was restricted to start from January 1978, and the most recent months, July and August 2025, were excluded due to incomplete data.

To construct key financial spread variables, which are critical predictors of economic activity, we defined:

- **Credit risk spread** (BAA_AAA): the difference between corporate bond yields and high-grade bond yields, calculated as $BAA - AAA$.
- **Yield curve spread** (T10Y3M): the difference between the 10-year Treasury yield and the 3-month Treasury bill rate, calculated as $GS10 - TB3MS$.

3.2. Data Transformation

To ensure stationarity of all time series, we followed the transformation codes provided in the FRED-MD dataset. Many variables exhibit stochastic trends or unit roots, which, if unaddressed, can lead to spurious regressions and invalid inference. The transformation codes were extracted and applied using the apply_tcode() function, implementing differencing, log-differencing, or growth-rate transformations as appropriate. These codes are based on the procedures established in McCracken and Ng (2016), which standardize FRED-MD variables to achieve approximate stationarity without requiring individual unit-root testing.

3.3 Data Summary Statistics

The FRED-MD dataset comprises 569 monthly observations from 1978 to 2025, containing 126 standardized macroeconomic and financial indicators. Descriptive statistics for the dependent variable (INDPRO_growth) and four key predictors—credit spread (BAA_AAA), housing starts (HOUST), term spread (T10Y3M), and consumer sentiment (UMCSENTx)—are summarized in Appendix A, Figure 1. Industrial production growth shows a small positive mean but wide dispersion, reflecting frequent expansions punctuated by sharp contractions.

Appendix A, Figure 2 illustrates the time series of INDPRO_growth, where two pronounced negative spikes correspond to the 2008 Global Financial Crisis and the 2020 COVID-19 shock. Both episodes represent structural breaks that introduce instability into the data-generating process and are likely to hinder the predictive performance of forecasting models trained on pre-crisis dynamics. Appendix A, Figure 3 shows that the key predictors exhibit similar volatility and abrupt shifts during these periods, indicating strong co-movement among macro-financial variables in times of stress.

The STL decomposition in Appendix A, Figure 4 isolates long-run trends, seasonal patterns, and irregular components of INDPRO_growth. The irregular component captures crisis-related shocks, while the seasonal pattern remains relatively stable across time. The boxplot in Appendix A, Figure 5 further highlights multiple outliers, with most observations tightly clustered near zero but several extreme downturns exceeding –10 %.

Overall, the presence of structural breaks, episodic volatility, and outliers underscores the inherent difficulty of forecasting monthly industrial production growth. Linear models may fail to adapt to these regime changes, while machine-learning methods risk overfitting transient anomalies—making comparative evaluation across model classes essential.

4. Results Analysis

Table 1 presents the summary statistics of the results from all models fitted to the FRED-MD dataset. It reports the OOS forecast performance, including the evaluation metrics MSE, MAE, and R^2.

Model	**MSE**	**MAE**	**R^2**
AR(3)	3.75	0.853	-0.304
ADL	3.82	0.864	-0.325
LASSO (CV)	3.46	0.824	-0.207
LASSO (BIC)	3.12	0.807	-0.088

Post-LASSO	3.46	0.824	-0.207
Post-LASSO*	2.43	0.775	0.153
Adaptive LASSO	3.00	0.739	-0.045
Elastic Net	5.44	0.988	-0.890
Random Forest	2.84	0.733	0.013
Gradient Boosting	2.77	0.740	0.040

*Post-LASSO is built on the LASSO model that was tuned by BIC

Table 1: OOS Model Evaluation Metrics

4.1. Benchmark Models

We select the optimal lag length p* based on the lowest BIC value to mitigate overfitting. As shown in Table 2, the selected lag length is p=3. The AR(3) model yields an OOS R^2 of -0.304, indicating limited predictive power beyond the mean forecast.

p	AIC	BIC
1	887.21	899.54
2	862.88	879.32
3	854.00	874.54
4	853.36	878.02
5	854.12	882.88
6	855.90	888.78

Table 2: Selecting optimal AR(p)

The ADL(3,4) model produces an even lower OOS R^2 = –0.325, suggesting that incorporating lagged exogenous predictors does not improve forecasting accuracy. From an economic perspective, we postulate that the additional predictors contribute limited new information about future industrial production growth or that the short run relationships with the lagged predictors are too unstable to provide good information. Additionally, the larger parameter set in the ADL model results in increased estimation variance and a higher risk of overfitting, which may have resulted in the weaker generalisation performance (OOS R^2).

4.2. LASSO Models

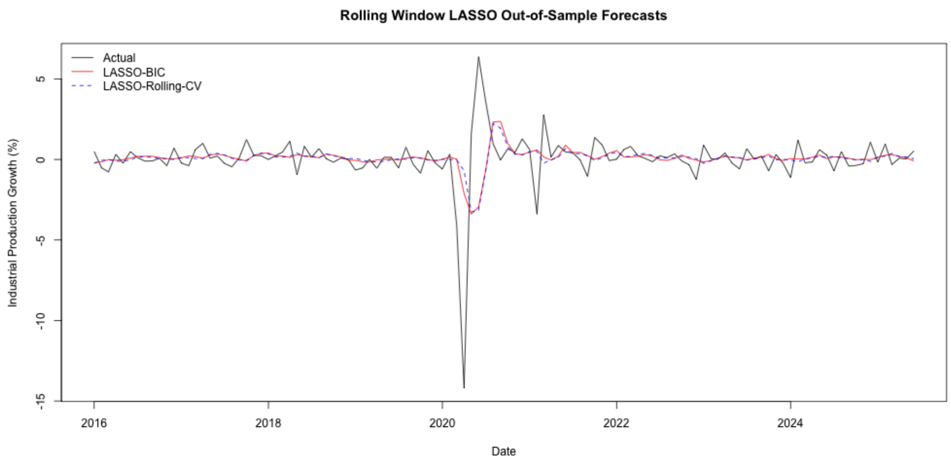

Fig. 1. OOS forecast vs actual with LASSO

By adjusting the maximum lag length for the predictor variables, we found that the best model performance was achieved at max_lag = 4, likely capturing the effect of quarterly seasonality common in economic and financial data, resulting in a total of 507 lagged features. Given the relatively small sample size, a larger maximum lag would have introduced an excessive number of predictors relative to observations, increasing the risk of overfitting. After lagging, we used a train-test split of 452-114.

The LASSO model's optimal penalty parameters were λ_BIC = 0.0984 and λ_CV = 0.1947. While the OOS R^2 for the BIC-selected λ was slightly worse than that from CV, both were negative, as shown in Table 1. Using BIC, 10 predictors were selected: CLAIMSx_lag1, ISRATIOx_lag2, T1YFFM_lag1, S&P 500_lag3, S&P div yield_lag2, BAA_AAA_lag1, BAA_AAA_lag2, INDPRO_growth_lag1, INDPRO_growth_lag2 and INDPRO_growth_lag3.

Unemployment claims or insurance claims CLAIMSx_lag1 reflect current labour market conditions. Rising claims indicate weakening employment conditions, which usually precede or coincide with slowed economic output and industrial production. The S&P 500 index captures business sentiment and investment trends that feed into industrial activity with a lag. Dividend yield and credit spreads BAA_AAA provide forward-looking market information about risk appetite and economic expectations, from investor expectations about corporate profitability, cash flow, to the broader economic environment, which in turn influence industrial production growth. The short term interest rate ratio ISRATIOx_lag2 and treasury yields T1YFFM_lag1 are relevant monetary factors that help explain industrial production.

Under the rolling CV setup, the LASSO model retained only the lagged values of the dependent variable, selecting INDPRO_growth with the 3 lags. This effectively reduces the model to an AR(3), suggesting that no other predictors provided additional explanatory power beyond the series' own past values.

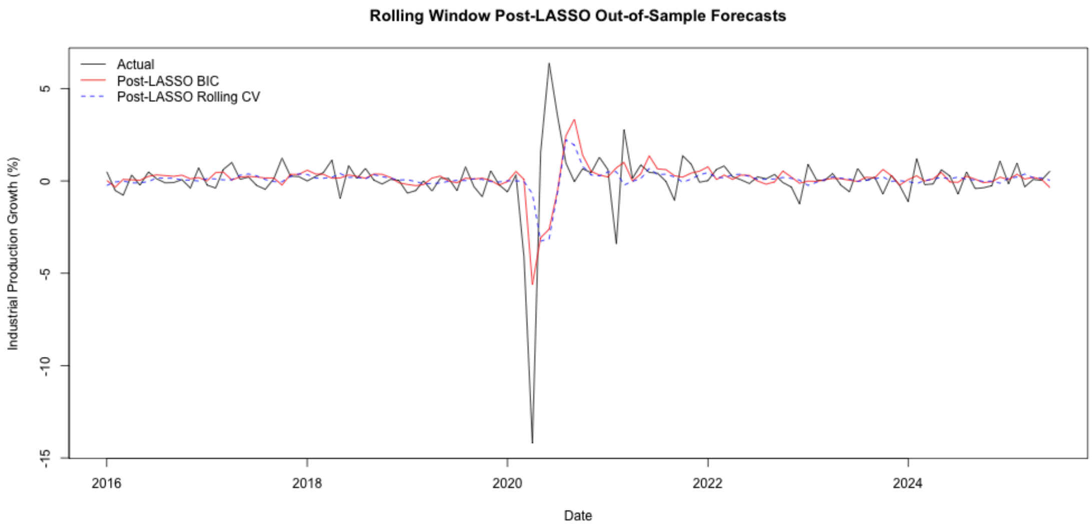

Fig. 2. OOS forecast vs actual with post-LASSO

The post-LASSO model, which refits OLS using the variables selected by LASSO, achieved the best performance among all models with λ_BIC, with OOS $R^2 = 0.1525$. As OLS does not shrink coefficients the bias introduced by LASSO is removed while keeping selected predictors. Thus LASSO selects useful predictors and post-LASSO improves the fit. We observed λ_CV performed better for LASSO itself but the BIC-based model outperformed it after OLS refit, likely as BIC tends to select fewer but stronger predictors, removing noise retained in CV. Once the shrinkage was removed in post-LASSO, the BIC-induced variable set provided a better bias–variance trade-off and improved generalisation.

With adaptive LASSO, it performed worse at $R^2 = -0.44$, likely due to over regularisation from the adaptive weights, causing increased effective penalisation for features with small ridge coefficients, leading to excessive shrinkage. This is further shown by the excessively large $\lambda = 114.629164$.

4.3. Elastic Net Model

The Elastic Net model achieved an out-of-sample $R^2 = -0.890$ and MSE of 5.44, indicating relatively weak predictive power compared to other regularized models. Although the Elastic Net combines both $L1$ (LASSO) and $L2$ (Ridge) penalties to handle correlated predictors, its performance suggests that the macroeconomic predictors in the FRED-MD dataset contribute limited additional explanatory power beyond lagged values of industrial production itself.

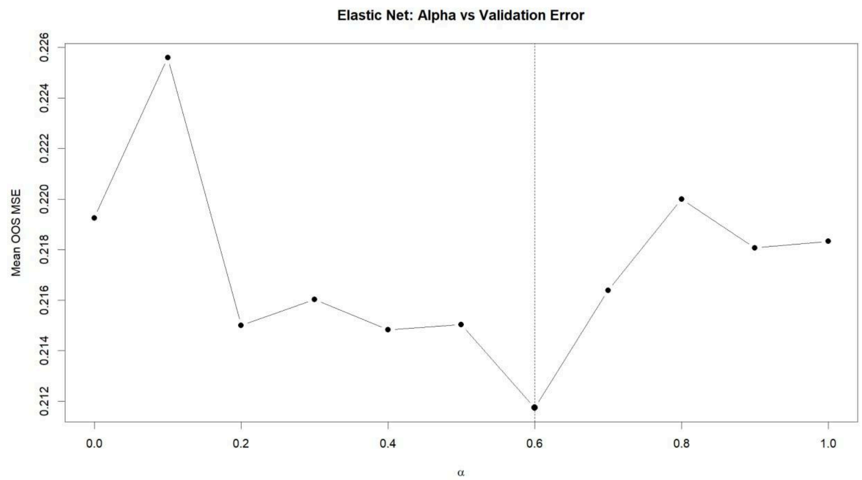

Fig. 3. Alpha selection for Elastic Net

The optimal mixing parameter α = 0.6 (Fig. 3) implies that the model leaned more towards LASSO-type sparsity than Ridge-type shrinkage, selecting only a few dominant predictors while moderately penalizing the rest. This balance helps mitigate multicollinearity but can still suffer when the underlying signal-to-noise ratio is low. The deterioration in R^2 likely reflects over-regularization, which is a tendency to overshrink coefficients when predictors are weakly informative, especially in the presence of structural breaks and non-stationary relationships.

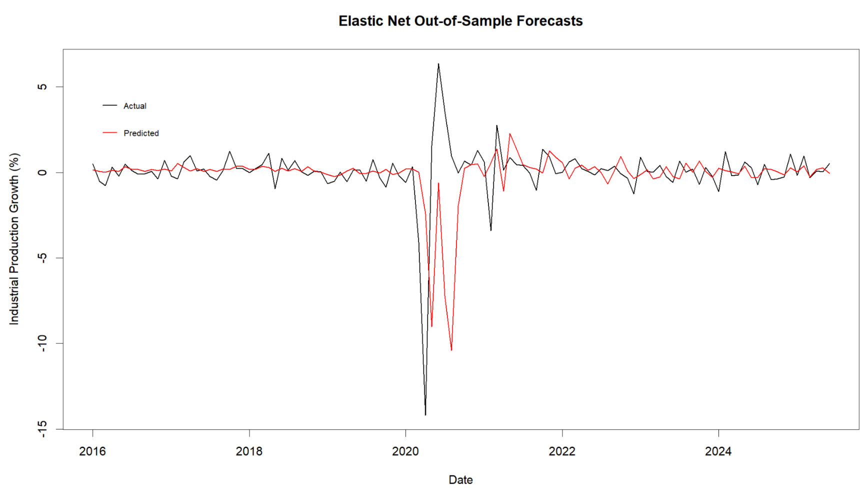

Fig. 4. OOS forecast vs actual with Elastic Net

The Elastic Net's performance indicates excessive sensitivity to short-term shocks rather than genuine adaptability. As shown in Figure 4, the model sharply overreacts to the COVID-19 collapse in 2020, producing extreme forecast deviations and a large deterioration in fit. The out-of-sample R^2 falls from approximately −0.110 (when 2020 is excluded) to −0.890 when the COVID period is included, underscoring how large exogenous shocks can severely distort model accuracy. This behaviour arises because Elastic Net retains many weakly correlated predictors, which collectively amplify transient fluctuations rather than filtering them out. The penalty structure shrinks coefficients continuously but does not eliminate uninformative variables, allowing noise to propagate through the model. Consequently, it responds too strongly to temporary volatility while failing to capture the true recovery dynamics.

Economically, this pattern suggests that the Elastic Net is overfitting short-lived macroeconomic disturbances rather than learning stable predictive relationships. Its combined $L1$–$L2$ regularization smooths across correlated features but does not confer robustness to regime changes. In contrast, the LASSO's stricter sparsity constraint yields more stable and interpretable behaviour during crisis periods, explaining its markedly superior out-of-sample R^2 and forecasting precision relative to Elastic Net.

4.4. Random Forest Model

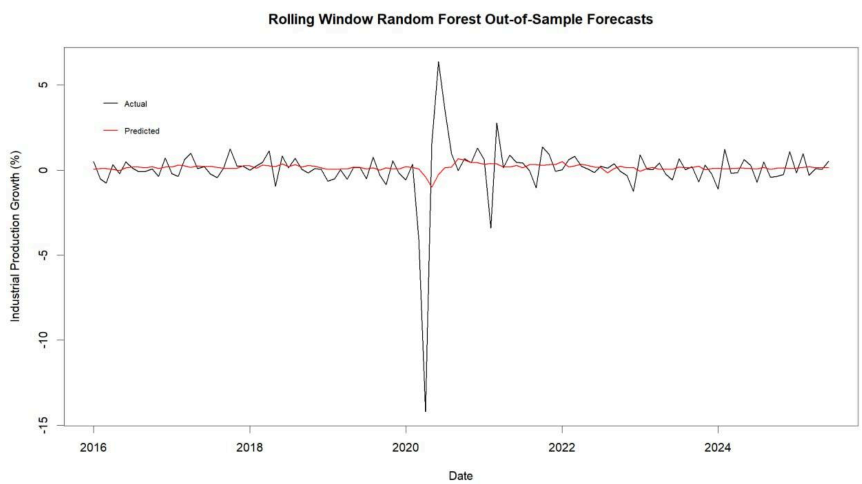

Fig. 5. OOS forecast vs actual with RF

The OOS R^2 for the RF model was 0.0223 (Table 1), indicating that the model explains approximately 2.23% of the variance in the test data. While this R^2 is low, it demonstrates that the model captures some predictive signal beyond the mean, though a substantial portion of the variance remains unexplained. Yet, with figure 9, we can observe the severe underfitting of the model as RF prioritises reducing variance by averaging out trees which only increases bias. CV results showed that tuning around mtry=45 and minimum node size = 10 yielded the lowest fold error, CV MSE ≈ 0.6285, suggesting that these hyperparameters provide an optimal balance between model complexity and generalisation.

The most important predictors identified by the model include lagged industrial production measures (IPDMAT_L1, IPDMAT_L2, IPDMAT_L3), labour market indicators (UNRATE_L1 – unemployment rate, MANEMP_L1 – manufacturing employment), and interest rate or yield variables (TB6MS_L1, T1YFFM_L2). The prominence of labour market variables aligns with economic theory, as employment conditions are closely linked to output and growth. The importance of interest rate and yield measures reflects the influence of financial conditions and monetary policy on economic activity. Additional variables, such as USTRADE_L1 (trade-related) and FEDFUNDS_L1 (federal funds rate), suggest that broader macroeconomic factors also play a role in shaping industrial production dynamics.

4.3. Gradient Boosting Model

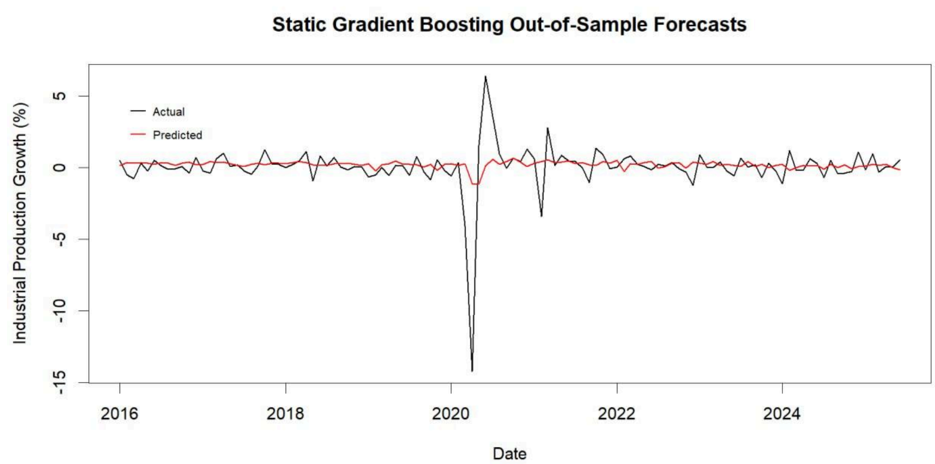

Fig. 6. OOS forecast vs actual with Gradient Boosting

The optimal hyperparameters identified via rolling CV included a modest learning rate (eta = 0.1), shallow trees (max_depth = 2), a relatively large minimum child weight (5), and subsampling of 80% of rows and 50% of columns. These settings indicate a preference for simpler trees and stronger regularisation to prevent overfitting. The optimal number of boosting rounds was approximately 43, selected using early stopping.

The GB model achieved an OOS R^2 of roughly 0.040, explaining about 4% of the variance in the test data. While this performance surpasses that of previously evaluated models, the shallow tree depth and regularisation choices suggest the model favors smooth, simple response surfaces. This helps generalisation but does not allow for highly nonlinear patterns in industrial production growth.

5. Concluding Remarks

In conclusion, this study employed the FRED-MD dataset to forecast industrial production growth. The data were preprocessed and transformed to achieve stationarity, after which benchmark models, AR(3) and ADL(3,4) were established for comparison. Subsequently, several ML methods were implemented, including LASSO, post-LASSO, adaptive LASSO, elastic net, RF, and GB.

The results indicate that forecasting industrial production growth is challenging due to the noisy and complex nature of macroeconomic data, which limits model performance and the attainable R^2 values. Among all methods, only GB, RF and post-LASSO achieved positive OOS R^2 values, with post-LASSO performing best at $R^2 = 0.1525$. This suggests that while traditional and regularized linear models capture limited predictive structure, incorporating feature selection and nonlinear interactions can offer modest gains in forecasting accuracy.

Interestingly, the more complicated models did not necessarily improve the forecast of INDPRO_growth, and the simpler ones actually performed better for OOS prediction. As the predictors used were highly volatile, useful signals might not have been captured purely by the lagged predictors.

6. Limitations

6.1. Data

Several limitations in the dataset make forecasting industrial growth production challenging. For one, the noise and the complexity of macroeconomic dynamics cause many relevant factors like global shocks, policy changes and external influences to not be considered in the FRED-MD dataset. As a result of this, this limits the information that can be used to train the model. Subsequently, using monthly data might hide some higher-frequency dynamics or signals in real-time that could be critical for prediction of industrial production growth, leading to the loss of important information too. Finally, while efforts were made to account for structural breaks and ensure stationarity, it is hard to fully identify and model these changes with static models we used for training in this project, which may reduce predictive accuracy. Further statistical testing such as Chow or Bai-Perron tests may be necessary to identify structural breaks. This means that predicting while accounting for the structural break in 2020 due to COVID-19 was not possible when we trained data until 2015.

6.2. Model

The LASSO and elastic net models assume linear relationships between predictors and the target variable, which hinders in finding nonlinearities. In comparison, RF and GB can model such nonlinear patterns but are less straightforward to interpret. This is because their variable importance measures offer some insight into predictor relevance, but do not provide causal inference.

For RF, each tree in the forest is trained on a bootstrapped sample and predicts an average outcome for the region the feature space covers. Aggregating hundreds of trees, the ensemble prediction will be the average of many conditional means, which massively smooths volatility and suppresses extreme values, helping to stabilise and create low-variance predictions. However, this means it is insensitive to sharp turning points or big shocks like COVID. This also means that structural breaks are not captured by the model and will average values pre and post- structural break. Additionally, especially with volatile data like Industrial Production, all of the models used in this project struggle to extrapolate accurately beyond the range of historical data, limiting their predictive reliability.

7. Future Research

Future research could integrate policy-relevant economic interpretation with advanced data-driven forecasting methods. While traditional machine learning models provide strong baselines, emerging approaches in deep learning and hybrid ensembles present opportunities to enhance both accuracy and interpretability in predicting industrial production growth. Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures, are well suited for capturing complex temporal dependencies and nonlinear dynamics in macroeconomic time series without requiring a fixed lag structure. In addition, model stacking or blending could exploit complementary strengths across econometric and machine learning frameworks to improve robustness across economic regimes. More recently, transformer-based architectures leveraging attention mechanisms have shown promise in learning long-range dependencies and structural shifts, offering flexibility beyond traditional models. Future work should also assess the statistical significance of forecast gains relative to benchmarks using formal tests such as the Diebold-Mariano test. Together, these directions point toward integrating modern neural and ensemble techniques with econometric insight to advance macroeconomic forecasting precision and interpretability.

Bibliography

Mccracken, M, and S Ng. "FRED-MD: A Monthly Database for Macroeconomic Research, Federal Reserve Bank of St." *ECONOMIC RESEARCH FEDERAL RESERVE BANK of ST. LOUIS*, 2015, s3.amazonaws.com/real.stlouisfed.org/wp/2015/2015-012.pdf, https://doi.org/10.20955/wp.

Smeekes, Stephan, and Etienne Wijler. "Macroeconomic Forecasting Using Penalized Regression Methods." *International Journal of Forecasting*, vol. 34, no. 3, July 2018, pp. 408–430, https://doi.org/10.1016/j.ijforecast.2018.01.001.

Goulet Coulombe, P., Leroux, M., Stevanović, D., & Surprenant, S. (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics, 37*(5), 920–964. (See also arXiv preprint.)

Medeiros, M. C., et al. (2019). Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics* / working paper (see full article). Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3155480.

Appendix A

INDPRO_growth	BAA_AAA	HOUST	T10Y3M	UMCSENTx
Min. :-13.2418	Min. :-0.630000	Min. :-597.0000	Min. :-2.190000	Min. :-17.30000
1st Qu.: -0.2260	1st Qu.:-0.050000	1st Qu.: -64.0000	1st Qu.:-0.160000	1st Qu.: -2.40000
Median : 0.2087	Median :-0.010000	Median : -3.0000	Median :-0.010000	Median : -0.10000
Mean : 0.1419	Mean :-0.000123	Mean : -0.5905	Mean :-0.002408	Mean : -0.04042
3rd Qu.: 0.5628	3rd Qu.: 0.040000	3rd Qu.: 63.0000	3rd Qu.: 0.150000	3rd Qu.: 2.50000
Max. : 6.5849	Max. : 0.940000	Max. : 363.0000	Max. : 3.330000	Max. : 17.30000

Fig. 1. Descriptive Summary of key predictors

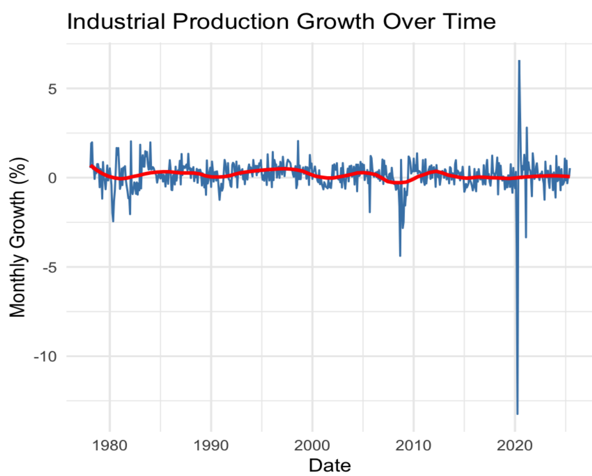

Fig. 2. Time trend of INDPRO_growth

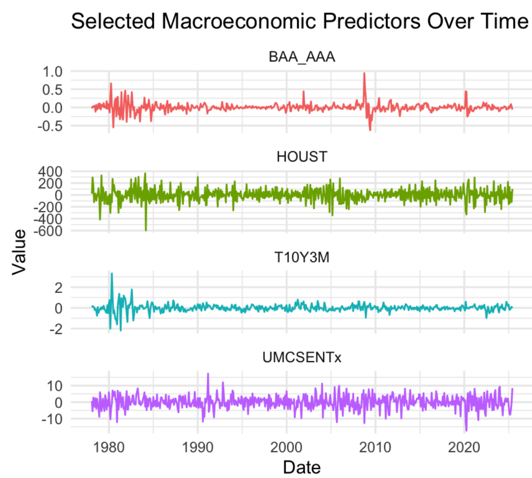

Fig. 3. Time trend of key predictors

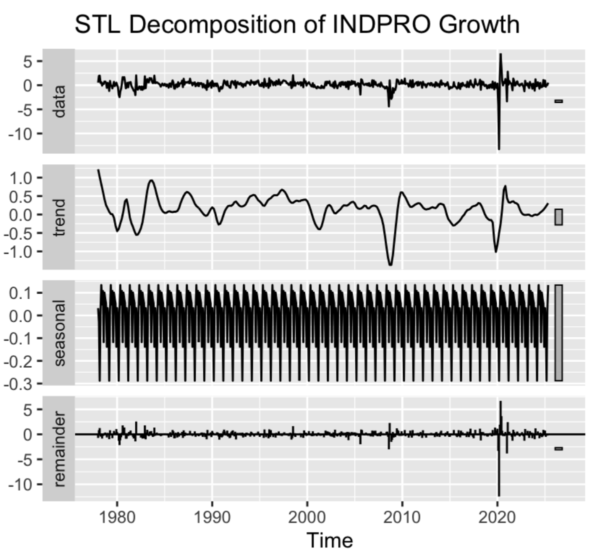

Fig. 4. STL Decomposition of INDPRO_growth

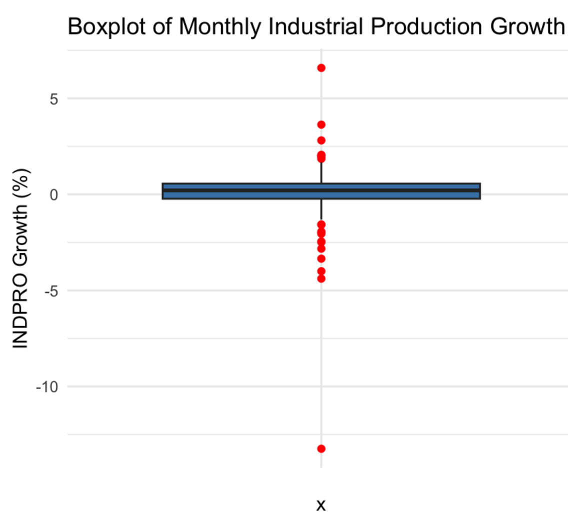

Fig. 5. Boxplot of INDPRO_growth