Real-Time Economic Activity Monitoring in a Data-Rich Mixed-Frequency Environment:

Harnessing Machine Learning

Charalampos (Haris) Karagiannakis[1]

May 8, 2025

[1] King's Business School, King's College London

Dissertation for the degree of Doctor of Philosophy

PhD scholarship from the **Qatar Centre for Global Banking & Finance** at KCL is gratefully acknowledged.

Motivation

- GDP is a key macroeconomic indicator, but published infrequently, with delay and is subject to revisions.
- Real-time economic monitoring tools are essential for timely and informed policy decisions.
- Same and higher-frequency predictors (e.g., monthly, weekly, daily) available before GDP, enable real-time prediction updates throughout the quarter.
  $\Rightarrow$ The concept of *nowcasting* = early estimates of *current* economic activity.
- Recent systemic patterns of large forecast errors across central banks have highlighted the need to re-evaluate and improve current forecasting practices.

Motivation

Poor forecasts and substantial revisions hinder policymakers' ability to implement timely and effective policies, underscoring the need for improved real-time monitoring and forecasting tools.

- In the fourth quarter of 2021, the ECB posted its largest one-quarter-ahead forecast error (for 2022Q1) since staff projections began in 1998, with similar large underestimates by other central banks, international institutions, and private forecasters.
- By December 2021, the BLS had issued its largest-ever annual upward revision to monthly non-farm payroll estimates, adding nearly 1 million jobs for the entire year.

ECB issues mea culpa for poor inflation forecasts

Central bank says it was blindsided by 'exceptional' energy prices while German inflation hits fresh 40-year high

The ECB made its worst ever inflation forecast in December when it predicted eurozone consumer price growth would fall to 4.1 per cent in the first quarter of this year. Instead it shot up to 6.1 per cent, prompting the ECB to accelerate its plans for stopping net bond purchases and opening the door to a potential interest rate rise as early as July.

US struggles to measure jobs growth as pandemic distorts labour market data

Payroll growth estimates revised up by 976,000 jobs in 2021, their highest upward adjustment in a single year

The implications are enormous for the US Federal Reserve, which is closely watching the employment situation for the green light to tighten monetary policy next year.

Christopher Waller, a Fed governor, said the central bank had already supplemented its models with high-frequency data and other sources, including the weekly ADP employment report.

Literature

Recently, the availability of large datasets has led researchers to re-evaluate **machine learning models for macroeconomic forecasting**, with promising results, and the literature continues to grow.

- **ML Horse races**: Medeiros et al. (2021; inflation); Goulet Coulombe et al., 2021b; macro targets); Gu et al. (2020; stock returns); Groen and Kapetanios (2016; macro targets).
- CSR with regularization (Kotchoni et al., 2019); Boosting factor autoregressions (Bai and Ng, 2009); Bagging linear regressions (Inoue and Kilian, 2008; inflation).

Efforts to extend methodologies to incorporate **mixed-frequency datasets** for nowcasting and forecasting economic data (primarily GDP), have recently expanded to include **machine learning** models.

- Penalized MIDAS regression (sg-LASSO-MIDAS) (Babii et al., 2022)
- Echo State Network (MFESN) (Ballarin et al., 2024)
- Random Forest (MO-RFRN) (Clark et al., 2022)
- Three-pass regression filter (MF-3PRF) (Hepenstrick and Marcellino, 2019)
- MF-VAR with Bayesian additive regression trees (MF-BAVART) (Huber et al., 2020)

Contributions

- Systematic evaluation of a **wide range of ML models** and **standard econometric benchmarks** for nowcasting and forecasting U.S. GDP growth, ensuring an appropriate experimental setup through a:
  - Novel **large mixed-frequency** dataset of 258 series,
  - Set of pseudo **real-time** vintages reflecting the asynchronous releases of data, and
  - Dual evaluation framework assuming quarterly and monthly monitoring.
- Propose and compare **three mixed-frequency treatments**:
  1. Equally weighted temporal aggregation
  2. Unrestricted distributed lag inclusion
  3. Aggregation with Legendre polynomials and sets of orthogonal weights

Contributions

- Systematic evaluation of a **wide range of ML models** and **standard econometric benchmarks** for nowcasting and forecasting U.S. GDP growth, ensuring an appropriate experimental setup through a:
  - Novel **large mixed-frequency** dataset of 258 series,
  - Set of pseudo **real-time** vintages reflecting the asynchronous releases of data, and
  - Dual evaluation framework assuming quarterly and monthly monitoring.
- Propose and compare **three mixed-frequency treatments**:
  1. Equally weighted temporal aggregation
  2. Unrestricted distributed lag inclusion
  3. Aggregation with Legendre polynomials and sets of orthogonal weights
- Contribute to the **sparse vs. dense modeling debate**, by investigating the comparative performance of factor-only vs. composite sets (factors + individual series).

Literature (cont.)

Sparse vs *dense* modelling techniques debate:

- **Metanalyses** suggest training **ML models** on sets of **common factors** is often superior to any other strategy (e.g., Goulet Coulombe et al., 2022, 2021). Bai and Ng (2009) also conclude that applying least-squares boosting on principal components outperforms boosting the underlying set of observables.
- Medeiros et al. (2021) conclude that the best performing models for inflation forecasting **do not impose sparsity**, but the high-aggregation level of **factor models** is also **inadequate**.
- Giannone et al. (2021): **Bayesian framework** that allows for both, leaving the data to decide. They **reject** the assumption of **sparsity** in macroeconomic forecasting contexts, but also find that the posterior distribution of the parameters **does not imply** a single **fully dense model**.
- **Sparse-plus-dense** model class, nesting the two and departing from the assumption that coefficients must be either sparse or dense (e.g., Fan et al., 2023; Chernozhukov et al., 2017; Beyhum and Striaukas, 2023).

Contributions (cont.)

- Evaluation of a **wide range of ML models** and **standard econometric benchmarks** for nowcasting and forecasting U.S. GDP growth, ensuring an appropriate experimental setup through a:
  - Novel **large mixed-frequency** dataset of 258 series,
  - Set of pseudo **real-time** vintages reflecting the asynchronous releases of data, and
  - Dual evaluation framework assuming quarterly and monthly monitoring.
- Propose and compare **three mixed-frequency treatments**:
  1. Equally weighted temporal aggregation
  2. Unrestricted distributed lag inclusion
  3. Aggregation with Legendre polynomials and sets of orthogonal weights
- Contribute to the **sparse vs. dense modeling debate**, by investigating the comparative performance of factor-only vs. composite sets (factors + individual series).

Contributions

Some key findings:

- **ML algorithms outperform benchmarks** in nowcasting and short-term GDP forecasting, though the single best model differs across horizons and metrics. Selected promising models:
  - L_2 **factor-boosting** with linear base learner and **block-wise** lags.
  - Ensembles of **bagged linear regressions**.
- **Target-ARDI** with hard-thresholding preselection (Stock & Watson, 2002; Bai & Ng, 2009) remains a reliable and competitive benchmark.
- The horse race results, suggest **factors from quarterly aggregated predictors** as an effective companion information set for numerous algorithms. However, **mixed-frequency sets containing both factors and individual series** become increasingly relevant when GDP is monitored at higher observational frequencies.

Target Variable & Methodological Framework

Our goal is to forecast continuously compounded (c.c.) quarter-on-quarter (QoQ) growth rate of real US GDP over quarters $h = 1, \ldots, H$ using a **large set** of potential predictors **sampled at various frequencies**, $\mathbf{X}_t = (x_{1t}, \ldots, x_{Nt})'$.

The general **direct-forecasting** framework for predicting target y_{t+h}, is given by:

$$y_{t+h} = f_h(\mathbf{X}_t) + u_{t+h}, \quad h = 1, \ldots, H, \quad t = 1, \ldots, T,$$

where $f_h(\cdot)$ is an unknown function that maps the information spanned by the covariates to the future values of the target variable, and u_{t+h} the error term. Let Y_t denote the real US GDP, then, the **target variable** is defined as:

$$y_{t+h} = 100 \ln \left(Y_{t+h} / Y_{t+h-1} \right).$$

The objective is to identify the method that provides the best estimate $\widehat{f_h}$ that minimizes a given measure of prediction accuracy.

Table: List of time series forecasting models

Acronym	Model Description	Reference
AR(P)	Autoregressive iterated-specification	
RW	Random walk	
ARDI(K)	Autoregressive diffusion indices with K factors	Stock and Watson (2002)
T.ARDI(K)	ARDI with target-factors. Hard-threshold set to $\lvert t\text{-stat}\rvert > 1.96$	Bai and Ng (2008)
BVAR-Minn	Homoscedastic large Bayesian VAR	Bańbura et al. (2010)
BVAR-CSV	Large Bayesian VAR with heteroscedastic innovations	Carriero et al. (2016)
Ridge	Ridge regression with BIC for λ	Hoerl and Kennard (1970)
LASSO	Least absolute shrinkage and selection operator with BIC for λ	Tibshirani (1996)
AdaLASSO	Adaptive LASSO	Zou (2006)
EN	Elastic Net with $\alpha = 0.5$	Zou and Hastie (2005)
AdaEN	Adaptive EN	
CSR	Complete Subset Regressions (20C4) with hard-thresholding	Elliott et al. (2013)
Bag	Bagging linear regressions	Inoue and Kilian (2008)
BBoost	Boosting linear regressions, block-wise	Bai and Ng (2009)
CBoost	Boosting linear regressions, component-wise	Buehlmann (2006)
BTree	Boosting regression trees	Friedman (2001)
RF	Random forests	Breiman (2001)
SVR	Support vector machine regression with Gaussian Kernel function	Drucker et al. (1996)
LSTM	Long-short-term memory RNN with 3-hidden layers	Hochreiter and Schmidhuber (1997)
SgLASSO-MIDAS	Sparse-group LASSO-MIDAS with block-K-fold CV for λ & γ	Babii et al. (2022)

Univariate Benchmarks

- **Autoregressive model,** AR(P): The **iterated** AR(P) is obtained by estimating the parameters in the following one-period-ahead model with OLS:

$$y_{t+1} = \phi_0 + \sum_{p=1}^{P} \phi_p y_{t+1-p} + \varepsilon_t.$$

The h-step ahead forecast is given by: $\hat{y}_{t+h|t} = \hat{\phi}_0 + \sum_{p=1}^{P} \hat{\phi}_p \hat{y}_{t+h-p|t}$.

Hyperparameters

  - Lag-order P fixed at 1 & 4
  - P selected via BIC & standard 5-fold CV à la Bergmeir et al. (2018)

- **Random walk (RW)**: Setting $P = 0$ in the AR(P) we obtain the constant-growth model where the h-step-ahead forecast is the fixed-window rolling average of y_t.

Large Vector Autoregressions

Let $\mathbf{y}_t = (y_{1t}, \ldots, y_{Nt})'$ represent the vector of N observables. The general VAR(p) model is given by:

$$\mathbf{y}_t = \mathbf{A}_0 + \mathbf{A}(L)\mathbf{y}_{t-1} + \mathbf{u}_t$$

where $\mathbf{A}_0$ is a vector of intercepts, $\mathbf{A}(L) = \sum_{i=1}^{p} \mathbf{A}_i L^{i-1}$ the p-th order lag polynomial, and $\mathbf{A}_i$ an $N \times N$ coefficient matrix. Two variants are considered:

- **Homoscedastic VAR**, where errors are assumed to be iid: $\mathbf{u}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$.
- **Common stochastic volatility** (Carriero et al., 2016), where $\mathbf{u}_t \sim \mathcal{N}\left(\mathbf{0}, e^{h_t}\mathbf{\Sigma}\right)$, and $h_t = \rho h_{t-1} + \varepsilon_t^h$, with $|\rho| < 1$ with $\varepsilon_t^h \sim \mathcal{N}\left(0, \sigma_h^2\right)$.

Hyperparameters

- Systems involve 20 quarterly variables (à la Bańbura et al., 2010; Chan, 2020), and $p = 4$
- **Bayesian estimation**:
  - Standard BVAR: Minnesota prior.
  - BVAR-CSV: Normal-inverse-Wishart prior for $(\mathbf{A}, \mathbf{\Sigma})$; independent truncated normal and inverse-gamma priors for ρ and σ_h^2.

Autoregressive Diffusion Index Models (ARDI)

Another well established class of models in macroeconomic forecasting is the **factor-augmented regressions** (FAR), also known as **Autoregressive Diffusion Index** (ARDI) Models, where predictors are assumed to follow a factor structure:

$$X_t = \Lambda F_t + e_t$$

where F_t is an $r \times 1$ vector of common factors, $\Lambda = (\lambda_1, \dots \lambda_N)'$ the matrix of loadings, and e_t the idiosyncratic component. The forecasting equation is given by:

$$y_{t+h} = \gamma_h + \alpha_h(L) y_t + \beta_h(L) \tilde{f}_t + \varepsilon_{t+h}$$

where $\tilde{f}_t = \left(\tilde{F}_{t1}, \dots \tilde{F}_{tk}\right)'$ and $\tilde{F}_t$ the vector of principal component estimates of F_t. $\alpha_h(L)$ and $\beta_h(L)$ are finite order lag polynomials with dimensions p and V_f.

Hyperparameters

- Up to two static factors $(k = 1, 2)$ are extracted from the extended quarterly dataset.
- Lag orders for y_t and $\tilde{f}_t$ selected via BIC.

Target ARDI

Preselection is added **prior to factor extraction** in the standard FAR model to identify the subset of predictors most relevant to the target variable, based on their predictive power. Filtering via **hard-thresholding** involves the following steps:

1. For predictor $i = 1, \ldots, N$ in X_t, run an OLS of y_{t+h} on $x_{i,t}$ and controls W_t. Conduct a two-sided test for the null hypothesis that the parameter associated with $x_{i,t}$ is zero, and let t_i denote the resulting t-statistic.
2. The N^*_α-dimensional vector of **targeted predictors** is given by: $X^*_t = \{x_i \in X_t \mid |t_i| > c_\alpha\}$, where c_α is the critical value at significance level α.

PCA is then applied on the set of targeted predictors X^*_t and forecasts are generated using the standard FAR predictive regression.

Hyperparameters

- Preselection threshold set to $c_\alpha = 1.96$, corresponding to $\alpha = 5\%$.
- Screening regressions control for the first AR lag.

Shrinkage Estimators

The general shrinkage estimator is given by:

$$\widehat{\boldsymbol{\beta}}_h = \arg\min_{\boldsymbol{\beta}_h} \left[\sum_{t=1}^{T-h} \left(y_{t+h} - \alpha_h - \boldsymbol{\beta}_h' \mathbf{X}_t \right)^2 + \sum_{i=1}^{N} p\left(\beta_{h,i}\right) \right]$$

where $p\left(\beta_{h,i}\right)$ the penalty function that depends on tuning parameter $\lambda \geq 0$. We consider several popular alternatives for the penalty function:

- **Ridge regression (RR)**, where the loss function is penalized by the squared ℓ_2 norm of the coefficients vector:

$$\sum_{i=1}^{N} p\left(\beta_{h,i}\right) := \lambda \|\beta\|_2^2 = \lambda \sum_{i=1}^{N} \beta_{h,i}^2.$$

- **Least Absolute Shrinkage and Selection Operator (LASSO)**, which adds the ℓ_1 penalty to the loss function. The penalty is given by:

$$\sum_{i=1}^{N} p\left(\beta_{h,i}\right) := \lambda \|\beta\|_1 = \lambda \sum_{i=1}^{N} \left|\beta_{h,i}\right|.$$

Shrinkage Estimators

- **Adaptive LASSO (adaLASSO)**, which uses a weighted version of the ℓ_1 penalty:

$$\sum_{i=1}^{N} p\left(\beta_{h,i}\right) := \lambda \sum_{i=1}^{N} \omega_i \left|\beta_{h,i}\right|$$

where weights are given by $\omega_i = \left|\beta_{h,i}^{*}\right|^{-1}$, with $\beta_{h,i}^{*}$ the coefficient from a first-step regression.

- **Elastic Net (EN)**, which combines the ℓ_1 (LASSO) and ℓ_2 (Ridge) penalties:

$$\sum_{i=1}^{N} p\left(\beta_{h,i}\right) := \lambda\alpha \sum_{i=1}^{N} \beta_{h,i}^{2} + \lambda(1-\alpha) \sum_{i=1}^{N} \left|\beta_{h,i}\right|$$

where parameter $\alpha \in [0,1]$ determines the relative contributions of the two penalties.

Shrinkage Estimators

- **Adaptive EN model (adaEN)**, which uses a weighted version of the EN penalty, with the weights defined by a first-round estimation of the EN model.

Hyperparameters

- λ is chosen via BIC (Kock and Callot, 2015).
- The elastic-net mixing parameter is set to $\alpha = 0.5$.
- The penalty factors in the adaptive LASSO and EN models were set to: $\omega_i = \frac{1}{|\beta_i^*| + \frac{1}{\sqrt{T}}}$, where β_i^* is obtained from first-round regressions of the standard versions of the LASSO and EN models (as in Medeiros et al., 2021).

Complete Subset Regressions (with preselection)

CSR forecasts are obtained by aggregating the forecasts from the collection (known as *complete subset*) of the M predictive regressions formed by all possible **combinations** of k **regressors** from the total set of N **predictors**. The m-th model is given by:

$$y_{m,t+h} = \gamma_m + \alpha_m' W_t + \left(X_t' S_m\right) \beta_m + \epsilon_{m,t+h}$$

where S_m is the $N \times N$ diagonal selector matrix with k unity elements indicating the variables included in the model. The CSR forecast is the equal-weighted average:

$$\hat{y}_{t+h|t} = \frac{1}{M} \sum_{m \in \mathcal{M}} \hat{y}_{m,t+h|t}$$

where $\mathcal{M}$ denotes the model space of all $M = N!/k!(N-k)!$ k-variate combinations.

Hyperparameters

- Subset regressions are based on models with $k = 4$ predictors chosen from the $N^* = 20$ variables with the **highest** absolute t-statistics.
- 1 AR lag is included as control both in the screening regressions, and in the individual models.

$\Rightarrow$ Target CSR forecasts are obtained by aggregating $M = c_{N^*,k} = 4845$ forecasts.

Bagging Linear Regressions

Bagging or *bootstrap aggregation* for high-dimensional time-series proceeds as follows:

1. For each bootstrap sample $b = 1, \ldots, B$, repeat:
   a) Run a pre-selection step to identify the subset of predictors that are statistically significant at a specified significance level $\tilde{\alpha}$: $X_t^{*(b)} = \{x_i \in X_t \mid \left|t_i^{(b)}\right| > c_{\tilde{\alpha}}\}$.
   b) Run an OLS regression using the b-th bootstrap replica containing only the $N^{(b)}$ significant variables from the previous step, and calculate the h-step-ahead forecast, on the *original data*: $\hat{y}^{*(b)}_{t+h|t} = \hat{\gamma}^{*(b)'} + \hat{\alpha}^{*(b)'} W_t + \hat{\beta}^{*(b)'} X_t^{*(b)}$.
2. The bagged forecast is the average of all forecasts across the B bootstrap samples:

$$\hat{y}_{t+h|t} = \frac{1}{B} \sum_{b=1}^{B} \hat{y}^{*(b)}_{t+h|t}.$$

Hyperparameters

- Block bootstrap with $B = 100$ replicas, and block length $l = 4$.
- **Group-based** prescreen procedure, followed by **random subset-selection** step (of 16 variables)
  - Preselection threshold set to $c_\alpha = 2.576$, corresponding to $\alpha = 1\%$.
  - 1 AR lag included as control both in screening regressions, and in individual learners.

Boosting Linear Regressions

Boosting approximates an unknown nonlinear function $\Phi(.)$ by sequentially fitting multiple *weak learners*. Assuming that the *quadratic-error loss function* is used to penalize deviations of $\Phi(X_t)$ from y_t, the boosting algorithm solves the problem:

$$\hat{\Phi} = \arg\min_{\Phi} \frac{1}{2T} \sum_{t=1}^{T} (y_t - \Phi(X_t))^2.$$

Under quadratic loss, the algorithm approximates $\Phi(x) = \mathbb{E}(y_t | X_t = x)$, and boosting reduces to an iterative least-squares refitting of the residuals (Friedman, 2001).

Two alternative treatments for **including predictor dynamics** (Bai and Ng, 2009):

- **Component Boosting**: Each variable and its lags are treated as *distinct predictors*.
- **Block Boosting**: Lags of each covariate are treated as a *group*.

Hyperparameters

- Boosting factor-augmented regressions (FAR) with $P_{\text{max}} = 4$.
- Base learners' lag lengths in **block boosting** selected via BIC.
- $\nu = 0.2$; # of iterations $M = 10 \times N$, and a BIC-based *early stopping rule* to prevent overfitting.

Boosting Linear Regressions

Component-wise Boosting (CBoost) Considering the lags of each covariate **in a standalone fashion**, the least-squares boosting algorithm is given by:

1. Let $\hat{\Phi}_{t,0} = \bar{y}$ for each t, with $\bar{y} = \frac{1}{t}\sum_{s=1}^{t} y_s$
2. For iteration $m = 1, \ldots, M$:
   a) Calculate the *current residuals* $\hat{u}_t = y_t - \hat{\Phi}_{t,m-1}$
   b) For each variable $i = 1, \ldots, N$ regress the *current residuals* $\hat{u}$ on the i-th regressor to obtain $\hat{b}_i$. Compute $\hat{e}_{t,i} = \hat{u}_t - x_{t,i}\hat{b}_i$ and the corresponding $SSR_i = \hat{e}_i'\hat{e}_i$.
   c) Let i_m^* denote the index of the predictor selected at the m-th iteration, corresponding to that delivering the smallest SSR:

$$SSR_{i_m^*} = \min_{i \in [1,\ldots,N]} SSR_i = \min_{i=1,\ldots,N} \sum_{s=1}^{t} \left(\hat{u}_s - \hat{\phi}_m(x_{s,i})\right)^2,$$

   d) Let $\hat{\phi}_{t,m} = x_{t,i_m^*}\hat{b}_{i_m^*}$.
   e) Update $\hat{\Phi}_{t,m} = \hat{\Phi}_{t,m-1} + \nu\hat{\phi}_{t,m}$ where $0 < \nu \leq 1$ is the step length.

Boosting Linear Regressions

Block-wise Boosting (BBoost) Considering the lags of each predictor **as a group**, the least-squares boosting algorithm is given by:

1. Let $\hat{\Phi}_{t,0} = \bar{y}$ for each t.
2. For $m = 1, \ldots, M$:
   a) Calculate the *current residuals* $\hat{u}_t = y_t - \hat{\Phi}_{t,m-1}$
   b) For each variable $i = 1, \ldots, N$ estimate the model:
   $$\hat{u}_t = \sum_{p=1}^{p_i^*} \alpha_p y_{t-p} + \sum_{q=1}^{q_i^*} \beta_q x_{t-(q-1),i} + v_t$$
   where lag orders (p_i^*, q_i^*) for the i-th regressor are selected via BIC.
   c) Let $(p_i^*, q_i^*) = \arg\min_{p,\,q} \text{BIC}(p, q)$, and $\hat{b}_i$ the OLS estimator obtained by regressing $\hat{u}$ on $z_{t,i}$ where $z_{t,i} = (y_{t-1}, \ldots, y_{t-p_i^*}, x_{t,i}, \ldots, x_{t-(q_i^*-1),i})'$.
   Compute $\hat{e}_{t,i} = \hat{u}_t - z_{t,i}'\hat{b}_i$ and the corresponding $SSR_i = \hat{e}_i'\hat{e}_i$.
   d) Let i_m^* be such that $SSR_{i_m^*} = \min_{i \in [1,\ldots,N]} SSR_i$.
   e) Let $\hat{\phi}_{t,m} = z_{t,i_m^*}\hat{b}_{i_m^*}$.
   f) Update $\hat{\Phi}_{t,m} = \hat{\Phi}_{t,m-1} + \nu\hat{\phi}_{t,m}$.

Random Forests

Random forests forecasts are obtained by averaging the predictions from multiple regression trees trained on B bootstrap versions of the original dataset:

$$\widehat{y}_{t+h} = \frac{1}{B} \sum_{b=1}^{B} \left[\sum_{k=1}^{K_b} \widehat{\beta}_{k,b} \mathrm{I}_{k,b} \left(X_t ; \widehat{\theta}_{k,b} \right) \right]$$

where K is the number of terminal nodes (determined by the tree depth), and $\mathrm{I}_k \left(X_t ; \theta_k \right)$ denotes an indicator function determining membership in each region (each corresponding to one terminal node), such as:

$$\mathrm{I}_k \left(X_t ; \theta_k \right) = \begin{cases} 1 & \text{if } X_t \in R_k \left(\theta_k \right) \\ 0 & \text{otherwise.} \end{cases}$$

Hyperparameters

- $B = 1000$ independently drawn bootstrap replicas.
- Split-variable randomization: $1/3$ of variables considered at each split.
- Tree depth: trees grown until a leaf node contains < 5 observations (min leaf size $= 5$).

Boosting Regression Trees

Boosted tree ensembles employ the gradient boosting algorithm and use decision trees as base learners. With a quadratic loss function, each successive tree is trained to minimize the residuals from the aggregated predictions of previous iterations. The generic least-squares gradient boosting algorithm with a regression tree as the weak learner, proceeds as follows:

1. Initialize $\hat{\Phi}_{t,0} = \bar{y}$.
2. For $m = 1, ..., M$ repeat:
   a) Compute the *current residuals* $\hat{u}_t = y_t - \hat{\Phi}_{t,m-1}$
   b) Fit a **shallow** regression tree to the data $(\hat{u}_s, X_s,)_{s=1}^{t}$, and estimate $\phi_{t,m}$
   c) Update the fitted model: $\hat{\Phi}_{t,m} = \hat{\Phi}_{t,m-1} + \nu\hat{\phi}_{t,m}$.
3. The final fitted value is given by: $\hat{y}_{t+h} = \bar{y} + \nu \sum_{m=1}^{M} \hat{\phi}_{t,m}$.

Hyperparameters

- Tree depth: minimum leaf size $= 5$.
- $\nu = 0.05$; $M = 1000$ regression trees are aggregated.

Support Vector Regression

The nonlinear regression SVM finds the coefficients that minimize the Lagrangian:

$$\max -\frac{1}{2}\sum_{i,j=1}^{T}\left(\alpha_i-\alpha_i^*\right)\left(\alpha_j-\alpha_j^*\right)k\left(x_i,x_j\right)-\varepsilon\sum_{i=1}^{T}\left(\alpha_i+\alpha_i^*\right)+\sum_{i=1}^{T}y_i\left(\alpha_i-\alpha_i^*\right)$$

subject to $\sum_{i=1}^{T}\left(\alpha_i-\alpha_i^*\right)=0$ and $\alpha_i,\alpha_i^*\in[0,C]$, where α_i,α_i^* are Lagrange multipliers. $k\left(x,x'\right):=\langle\Phi(x),\Phi\left(x'\right)\rangle$, denotes a nonlinear kernel function with $\Phi(x)$ a transformation that maps the training data x_i to a (potentially) high-dimensional feature space. The optimal solution for the function is given by:

$$f(x)=\sum_{i=1}^{T}\left(\alpha_i-\alpha_i^*\right)k\left(x_i,x\right)+b.$$

Hyperparameters

Gram matrix computed using the Gaussian kernel function: $k(x_i,x_j)=\exp(-\left\|x_i-x_j\right\|^2)$.

Neural Networks

The components of a **Long Short-Term Memory (LSTM)** layer, are described by:

$$
\begin{aligned}
g_t &= \sigma_c\left(\boldsymbol{W}_g \mathbf{x}_t + \mathbf{R}_g \mathbf{h}_{t-1} + \boldsymbol{b}_g\right) && \text{cell candidate} \\
i_t &= \sigma_g\left(\boldsymbol{W}_i \mathbf{x}_t + \mathbf{R}_i \mathbf{h}_{t-1} + \boldsymbol{b}_i\right) && \text{input gate} \\
f_t &= \sigma_g\left(\boldsymbol{W}_f \mathbf{x}_t + \mathbf{R}_f \mathbf{h}_{t-1} + \boldsymbol{b}_f\right) && \text{forget gate} \\
\mathbf{c}_t &= f_t \odot \mathbf{c}_{t-1} + i_t \odot g_t && \text{cell state} \\
o_t &= \sigma_g\left(\boldsymbol{W}_o \mathbf{x}_t + \mathbf{R}_o \mathbf{h}_{t-1} + \boldsymbol{b}_o\right) && \text{output gate} \\
\mathbf{h}_t &= o_t \odot h\left(\mathbf{c}_t\right) && \text{hidden state}
\end{aligned}
$$

$$\text{and } \hat{y}_{t+h/t} = \boldsymbol{U}_y \mathbf{h}_t + \boldsymbol{b}_y$$

where σ_c, σ_g and h are point-wise non-linear activation functions for the cell input, the gates and the cell state. $\boldsymbol{W}$, $\mathbf{R}$, $\boldsymbol{b}$ are (input, recurrent, bias) weights to be estimated.

Hyperparameters

- 3 recurrent hidden layers, with 64, 32, and 16 units, and a dense layer with 8 units followed by a ReLU layer applying the activation function in that layer's outputs.
- Dropout layers after each LSTM layer, with a 0.3 dropout rate.

Sparse-group-LASSO-MIDAS

Sparse-group-LASSO-MIDAS is a penalized regression model tailored for high dimensional time series sampled at the same or different frequencies. The **sparse-group LASSO (sg-LASSO)** estimator solves:

$$\min_{b \in \mathbf{R}^p} \|\mathbf{y} - \mathbf{X}b\|_T^2 + \lambda \Omega(b)$$

with regularization parameter $\lambda \geq 0$, and the sg-LASSO penalty function:

$$\Omega(b) = \gamma |b|_1 + (1-\gamma)\|b\|_{2,1},$$

where weight parameter $\gamma \in [0,1]$ balances between the ℓ_1 **LASSO** penalty and the **group-LASSO** norm, $\|b\|_{2,1} = \sum_{G \in \mathcal{G}} |b_G|_2$ with $\mathcal{G}$ denoting a group structure. Groups are defined by assembling *different aggregated versions* of each predictor's lags.

Hyperparameters

- λ and γ are jointly selected via 5-fold block-CV.
- Lags are aggregated using (orthogonal) **Legendre polynomials** of degree 3.
- 4 AR lags are also assembled into a group, but no Legendre aggregation is applied.

Methods for Handling Mixed-Frequency Data

Handling Mixed-Frequency Data

The **MIDAS regression** (Ghysels et al., 2005; Andreou et al., 2010) for forecasting a low-frequency target h periods ahead using N covariates sampled at different frequencies, is given by:

$$y_{t+h} = \alpha + \sum_{p=0}^{P} \phi_p y_{t-p} + \sum_{i=1}^{N} \beta_i \psi\left(L^{1/m}; \omega_i\right) x_{t+w_i,i} + u_{t+h},$$

where $\psi(L^{1/m}; \omega_i) = \sum_{j=0}^{V_i-1} \omega_i(\theta, j) L^{j/m}$ is the *(high-frequency) lag polynomial*, and $L^{j/m}$ the lag operator, with $L^{j/m} x_{t+w_i,i} = x_{t+w_i-j/m,i}$. m denotes the frequency mismatch between the target and predictor i, V_i is the total number of (**leads** and) **lags** for that predictor, and $t + w_i$ its last available period.

Label	Description
D1	Equal-weighted Temporal Aggregation
D2	Unrestricted Lag Polynomials
D3	Legendre Polynomial Weights (3rd degree)

Single-frequency Information Set

To construct **information set D1**, the series in the high-frequency panel are temporally aggregated by calculating the equal-weighted average of the periods corresponding to each quarter. These are then combined with the indicators from the quarterly panel.

Formally, following the temporal aggregation literature (e.g., Chow and Lin, 1971), the conversion of high-frequency indicator data to aggregated low-frequency observations is achieved through a **deterministic aggregator function** $\psi(L^{1/m})$, applied in the lag operator $L^{1/m}$. In the context of a monthly-quarterly conversion, the expression for downsampling the monthly predictor x_t^M to x_t^Q takes the form:

$$x_t^Q = \psi(L^{1/3})x_t^M = \sum_{j=0}^{2} \omega_j L^{j/3} x_t^M,$$

with $\omega_j = 1/3$ providing the uniformly weighted average.

Mixed-frequency Information Sets

- **Unrestricted Lag Polynomials**:
  The unrestricted (U-MIDAS) model is obtained by removing functional restriction $\psi()$:
  $$y_{t+h} = \alpha + \sum_{p=0}^{P} \phi_p y_{t-p} + \sum_{i=1}^{N} \sum_{j=0}^{V_i - 1} \delta_{j,i} L^{j/m} x_{t+w_i,i} + u_{t+h}.$$
  Information set D2 is given by: $(Z_1, \ldots, Z_N)$, with $Z_{t,i} = \left(L^{j/m} x_{t+w_i,i}\right)_{j \in [0,1,\ldots,V_i - 1]}$.

- **Aggregation with Legendre Dictionaries**:
  The lag coefficients can be parameterized by expressing the MIDAS weight function as a linear combination of a *collection of approximating functions* $w_l(u)$, called **dictionary**:
  $$y_{t+h} = \alpha + \sum_{p=0}^{P} \phi_p y_{t-p} + \sum_{i=1}^{N} \sum_{l=1}^{L} \beta_{i,l} \frac{1}{V_i} \sum_{j=0}^{V_i - 1} w_l \left(\frac{j}{m}\right) L^{j/m} x_{t+w_i,i} + u_{t+h},$$
  with $\{w_l : l = 1, \ldots, L\}$, and L the dictionary size.

  Information set D3 is given by: $(Z_1 W_1, \ldots, Z_N W_N)$ where
  $W_i = (w_l(j/m)/V_i)_{j \in [0, V_i - 1], l \in [1, L]}$ is the $V_i \times L$ matrix of weights.

Aggregation with Legendre Dictionaries

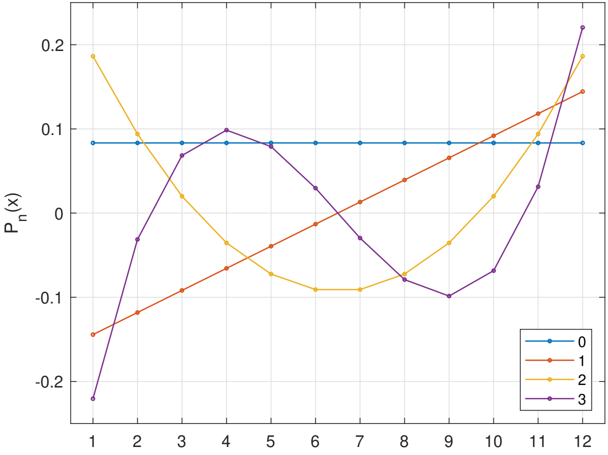


Figure: Legendre Polynomials for 12 Lags and Degrees 0 to 3

Data

A Comprehensive Mixed-Frequency Dataset

A large mixed-frequency dataset was built by *combining* **FRED-MD** and **FRED-QD** (McCracken & Ng, 2016, 2020), which replicate the coverage of the ‘Stock-Watson’ datasets widely used for applied forecasting and policy research in data-rich settings.

- Initial dataset: Apr-2021 FRED-MD/QD with 135 monthly and 248 quarterly series.
- Overlapping variables removed, and series downloaded at highest available frequency.

$\Rightarrow$ Resulting in two unbalanced panels: **171 monthly** and **87 quarterly** indicators.

This study is the **first attempt** to use the two large datasets together to inform the model selection process in a **data-rich mixed-frequency setting**.

Real-time Vintages

A set of pseudo real-time month-end vintages was constructed from the two unbalanced panels by applying publication delays inferred from the online FRED metadata for each predictor.

$\Rightarrow$ **219 monthly vintages** spanning the period **January 2003 - March 2021**.

	GDPC1	OUTBS	NWPIx	AHETPIx	MZMSL	UNRATE	CMRMTSPLx	UMCSENTx
Vintage 31/01/2021								
30/09/2020	18597	118	683	22	21250	8	1564146	80
31/10/2020	-	-	-	-	-	7	1572500	82
30/11/2020	-	-	-	-	-	7	1569672	77
31/12/2020	-	-	-	22	-	7	-	81
31/01/2021	-	-	-	-	-	-	-	79
28/02/2021	-	-	-	-	-	-	-	-
31/03/2021	-	-	-	-	-	-	-	-
Vintage 28/02/2021								
30/09/2020	18597	118	683	22	21250	8	1564146	80
31/10/2020	-	-	-	-	21369	7	1572500	82
30/11/2020	-	-	-	-	-	7	1569672	77
31/12/2020	-	120	-	22	-	7	-	81
31/01/2021	-	-	-	-	-	6	-	79
28/02/2021	-	-	-	-	-	-	-	77
31/03/2021	-	-	-	-	-	-	-	-
Vintage 31/03/2021								
30/09/2020	18597	118	683	22	21250	8	1564146	80
31/10/2020	-	-	-	-	21369	7	1572500	82
30/11/2020	-	-	-	-	21565	7	1569672	77
31/12/2020	18794	120	-	22	-	7	1566283	81
31/01/2021	-	-	-	-	-	6	-	79
28/02/2021	-	-	-	-	-	6	-	77
31/03/2021	-	-	-	-	-	-	-	85

Data

Preprocessing Steps

- Stationarity transformations:
  - Positive series not in rates or ratios are **log-transformed**.
  - **Differencing order** determined via iterative unit-root testing procedure SIC-based DF-GLS tests (Elliott et al., 1996) at $\alpha = 5\%$.
- **Outlier Adjustment**: Values exceeding median by $10 \times IQR$ are set to missing.
- **Missing values imputed** via EM-modified PCA algorithm of Stock and Watson (2002).
- Ragged-edge reintroduced to the 2 balanced (EM-filled) panels.

Implementation Specifics

Information sets D1, D2, D3, are created by augmenting the 171 monthly and 87 quarterly series in the 2 unbalanced panels with:

- K_p^* static PCA factors from the respective panel within each set,
- 4 AR terms; one year of **lags** + all available **leads** for each predictor and factor; and **Legendre-aggregation** of degree 3 for D3.

$\Rightarrow$ D1, D2, D3 contain 1226, 3081 and 1088 potential predictors (Jan-2003 vintage).

Table: List of the 20 variables for the Large Bayesian VARs

Series ID	Description	Freq.	Delay	Category
HOUST	Housing Starts: Total New Privately Owned	M	18	Housing
INDPRO	Industrial Production	M	14	Industrial Production
TCU	Capacity Utilization: Total Industry	M	14	Industrial Production
FEDFUNDS	Effective Federal Funds Rate	M	3	Interest Rates
GS10	10-Year Treasury Rate	M	3	Interest Rates
PAYEMS	All Employees: Total nonfarm	M	7	Labor
UNRATE	Civilian Unemployment Rate	M	7	Labor
M1SL	M1 Money Stock	M	25	Money and Credit
M2SL	M2 Money Stock	M	25	Money and Credit
NONBORRES	Reserves Of Depository Institutions	M	25	Money and Credit
TOTRESNS	Total Reserves of Depository Institutions	M	25	Money and Credit
RPI	Real Personal Income	M	27	Output, income, consumption
CPIAUCSL	CPI: All Items	M	12	Prices
PCEPI	Personal Cons. Expend.: Chain Index	M	27	Prices
PPIACO	Producer Price Index (PPI): All Commodities	M	13	Prices
WPSID61	PPI: Intermediate Materials	M	13	Prices
S&P 500	S&P's: Common Stock Price Index	M	0	Stock Markets
AHETPIx	Real Average Hourly Earnings (Core PCE deflated)	Q	7	Earnings & Productivity
GDPC1	Real Gross Domestic Product	Q	85	Output, income, consumption
PCECC96	Real Personal Consumption Expenditures	Q	85	Output, income, consumption

Out-Of-Sample Forecast Experiment Design

The vintages are used in 2 pseudo **out-of-sample evaluation experiments**, assuming:

- **Quarterly monitoring**: GDP tracked once at the end of each quarter (end-of-month 3 or 'EoM3'); 73 quarter-end vintages $\rightarrow v_0 = \text{2003:Q1},\ v_T = \text{2021:Q1} - n$
- **Monthly monitoring**: Nowcasts and forecasts are (re)calculated 3 times per quarter, at the end of each month. 219 month-end vintages $\rightarrow v_0 = \text{2003:M01},\ v_T = \text{2021:M03} - 3n$

Rolling-window estimation, with window size set to:

- $R_d = 132 - h$ observations for **direct** forecasting models,
- $R_{it} = 132 - P - 1$ observations for **iterated** forecasting models.

Forecast accuracy is assessed by **RMSE** and **MAE**, and model performance is compared for the **nowcast** and for **forecasts up to four quarters ahead**, $n = 0, 1, 2, 3, 4$ (with $n = 0$ the nowcast), obtained by setting $h = 1, 2, \ldots, 6$.

$$\text{RMSE}_{n,m} = \sqrt{\frac{1}{v_T - v_0 + 1} \sum_{v=v_0}^{v_T} \hat{e}^2_{v,n,m}} \qquad \text{MAE}_{n,m} = \frac{1}{v_T - v_0 + 1} \sum_{v=v_0}^{v_T} |\hat{e}_{v,n,m}|$$

where $\hat{e}_{v,n,m} = y_{v,n} - \hat{y}_{v,n,m}$ the OOS error.

Comparing Machine Learning Models and Benchmarks

ML Models on D1 & Benchmarks: **RMSE** wrt AR(1)

Models	n=0	n=1	n=2	n=3	n=4	avg	avgMCS
AR(1)	2.08	1.73	1.68	1.58	1.59	-	0.55
AR(4)	1.06	1.04	1.01	1.00	1.00	1.02	0.52
AR(BIC)	1.00	1.01	1.00	1.00	1.00	1.00	0.49
AR(CV)	1.03	1.01	1.00	1.00	1.00	1.01	0.45
RW	0.74	0.90	0.93	1.00	1.00	0.92	0.55
ARDI(1)	0.62	0.94	0.93	1.00	1.00	0.90	0.56
ARDI(2)	0.64	0.94	0.93	1.00	1.00	0.90	0.56
T.ARDI(1)	0.60	0.96	0.93	1.00	1.00	0.90	0.56
T.ARDI(2)	0.61	0.91	0.93	1.00	0.99	0.89	0.57
BVAR-Minn	0.70	0.95	0.93	1.00	1.02	0.92	0.52
BVAR-CSV	0.65	0.91	0.93	1.01	1.01	0.90	0.43
BBoost-D1F	**0.60**	**0.88**	0.93	1.00	1.00	**0.88**	**0.78**
CBoost-D1F	0.66	0.89	0.93	1.02	1.00	0.90	0.46
CSR-D1	0.66	0.92	0.93	1.01	1.01	0.91	0.43
Bag-D1	0.69	0.90	0.91	0.99	1.00	0.90	0.66
BTree-D1	0.72	0.90	0.94	1.01	1.03	0.92	0.31
RF-D1	0.69	0.90	0.93	1.00	1.00	0.91	0.48
SVR-D1	0.73	0.90	0.92	**0.99**	**0.99**	0.91	0.63
Ridge-D1	0.66	0.91	0.91	1.00	1.00	0.89	0.67
LASSO-D1	0.67	0.98	0.92	1.10	1.07	0.95	0.19
AdaLASSO-D1	0.77	0.92	0.98	1.06	1.01	0.95	0.26
EN-D1	0.68	0.90	**0.88**	1.05	1.04	0.91	0.43
AdaEN-D1	0.73	0.91	0.94	1.01	1.00	0.92	0.46
LSTM-D1	0.76	0.91	0.94	1.00	1.02	0.93	0.48
SgLASSO-D3	0.75	0.91	0.93	1.00	1.00	0.92	0.40

ML Models on D1 & Benchmarks: **MAE** wrt AR(1)

Models	n=0	n=1	n=2	n=3	n=4	avg	avgMCS
AR(1)	0.74	0.72	0.71	0.66	0.66	-	0.42
AR(4)	1.11	1.04	1.01	1.01	1.00	1.03	0.46
AR(BIC)	1.00	1.01	1.01	1.01	1.00	1.01	0.42
AR(CV)	1.06	1.01	1.00	1.01	1.00	1.02	0.45
RW	0.87	0.90	0.92	**0.99**	**1.00**	0.94	0.69
ARDI(1)	0.80	0.93	0.92	1.00	1.00	0.93	0.62
ARDI(2)	0.79	0.93	0.94	1.02	1.00	0.93	0.58
T.ARDI(1)	**0.75**	0.94	0.93	1.01	1.06	0.94	0.68
T.ARDI(2)	0.78	0.87	0.91	1.01	1.00	0.91	0.71
BVAR-Minn	0.80	1.01	0.94	1.05	1.10	0.98	0.65
BVAR-CSV	0.76	0.93	0.97	1.05	1.09	0.96	0.52
BBoost-D1F	0.78	**0.81**	0.93	1.01	1.01	**0.91**	**0.79**
CBoost-D1F	0.80	0.87	0.95	1.06	1.01	0.94	0.51
CSR-D1	0.75	0.93	0.98	1.05	1.08	0.96	0.52
Bag-D1	0.77	0.87	**0.90**	1.02	1.06	0.93	0.68
BTree-D1	0.85	0.95	1.00	1.04	1.13	1.00	0.21
RF-D1	0.78	0.90	0.93	1.02	1.02	0.93	0.60
SVR-D1	0.85	0.91	0.92	1.00	1.01	0.94	0.56
Ridge-D1	0.78	0.97	0.96	1.10	1.14	0.99	0.31
LASSO-D1	0.76	1.26	1.16	1.38	1.27	1.17	0.20
AdaLASSO-D1	0.90	1.02	1.13	1.18	1.09	1.06	0.11
EN-D1	0.77	0.87	0.98	1.19	1.13	0.99	0.27
AdaEN-D1	0.79	0.88	0.96	1.05	1.00	0.94	0.51
LSTM-D1	0.92	0.95	1.02	1.09	1.09	1.02	0.33
SgLASSO-D3	0.91	0.95	0.94	1.02	1.02	0.97	0.47

ML Models on D1 & Benchmarks: Key Takeaways

Alignment with established empirical findings:

- Block-wise boosting outperforms component-wise (Bai and Ng, 2009).
- CSV-BVAR outperforms the homoscedastic BVAR (Carriero et al., 2016; Chan, 2020).
- Pre-selection before factor extraction improves ARDI models (Bai and Ng, 2008).
- RW is a strong competitor among univariate models and overall (D'Agostino et al., 2007).
- Gains over the constant growth become harder at longer horizons (Bańbura et al., 2013).

ML Models on D1 & Benchmarks: Key Takeaways (cont.)

Model comparison:

- **Best overall performance**:
  Block factor-boosting (BBoost-D1F) consistently outperforms other models:
  - For nowcasting especially: 40% RMSE reduction vs. AR(1) and 19% vs. RW.
  - Among the few models outperforming the RW at 1-quarter ahead.
  - Across horizons: **Highest average predictive accuracy** wrt. relative error and MCS p-values under both RMSE and MAE; 37% higher MCS p-value vs T.ARDI(2).
- **Targeted-factor ARDI** models remain competitive, nearly matching BBoost-D1F in nowcast accuracy. However, outperformed by RW at the 1-quarter-ahead.
- Most ML models outperform RW at nowcast horizon, but only a few at 1-quarter ahead.
- Across horizons, **ridge regression** has the **second strongest average performance** wrt. relative **RMSE** and MCS p-values (but not in MAE terms).
- Sparse models (e.g., LASSO, AdaLASSO, sg-LASSO-D3) underperform, suggesting that **sparse representations may not capture GDP dynamics** effectively (see Giannone et al., 2021).
- Results are robust to both RMSE and MAE criteria.

Comparing Information Sets

How to Handle Mixed-Frequencies?

Mixing Frequencies Approaches: RMSE **D2/D1**

Models	n=0	n=1	n=2	n=3	n=4	avg
BBoost-D2F	1.34	1.03	0.98	0.99	1.00	1.07
CBoost-D2F	1.13*	1.03	0.98	0.99	1.00	1.03
CSR-D2	0.93	1.01	1.02	0.99	1.00	0.99
CSR-D2F	0.93	1.01	1.00	1.00	1.00	0.99
Bag-D2	0.96	1.02	1.01	1.00	1.00	1.00
Bag-D2F	1.12	0.99	1.00	0.99	1.00	1.02
BTree-D2	1.07	1.03	1.03	1.00	0.99	1.03
BTree-D2F	1.11*	1.02	1.06*	1.01	0.98	1.04
RF-D2	1.04	1.00	1.01*	1.00	1.00	1.01
RF-D2F	1.04	1.01	1.01	1.00	1.00	1.01
SVR-D2	1.00	1.00	1.01	1.01	1.00	1.00
SVR-D2F	1.01	1.00	1.01	1.00	1.00	1.00
Ridge-D2	0.93	1.01	1.02	1.01	1.00	0.99
Ridge-D2F	1.04	0.97	0.94*	0.94*	0.97*	0.97
LASSO-D2	1.05	1.03	1.03	0.92*	0.94*	0.99
LASSO-D2F	1.13*	0.99	0.98	0.99	1.00	1.02
AdaLASSO-D2	0.89	1.10	0.96*	0.94	1.04*	0.99
AdaLASSO-D2F	1.17	0.99	0.98	0.99	1.01	1.03
EN-D2	1.02	1.11	1.07	0.95	0.96	1.02
EN-D2F	1.09	0.99	0.99	1.00	1.00	1.01
AdaEN-D2	0.87	1.10*	1.00	1.00	1.03*	1.00
AdaEN-D2F	1.15	0.99	0.98	0.99	1.01	1.02
LSTM-D2	1.02	1.03*	0.99	1.02	0.98	1.01
LSTM-D2F	1.00	1.01	1.00	0.99	1.01	1.00
avg	1.04	1.02	1.00	0.99	1.00	-

Mixing Frequencies Approaches: RMSE D3/D1

Models	n=0	n=1	n=2	n=3	n=4	avg
BBoost-D3F	1.41	1.02	0.98	1.01*	1.01	1.08
CBoost-D3F	1.22	1.06*	1.00	1.01	1.00	1.06
CSR-D3	1.19	1.01	1.00	1.00	1.00	1.04
CSR-D3F	1.06	1.02	1.00	1.00	1.00	1.02
Bag-D3	0.90	1.02	1.02	1.01*	1.00	0.99
Bag-D3F	1.24	1.00	0.99	1.00	1.00	1.05
BTree-D3	1.07	1.09*	1.01	1.01	0.99	1.04
BTree-D3F	1.22	1.11	1.05	1.00	1.03	1.08
RF-D3	1.02	1.02	1.00	0.99	0.99*	1.00
RF-D3F	1.08*	1.02	1.01	0.99	1.01	1.02
SVR-D3	1.00	1.00	1.00	1.01	1.00	1.00
SVR-D3F	1.01	1.00	1.00	1.00	0.99	1.00
Ridge-D3	1.04	1.00	1.01	1.01	1.00	1.01
Ridge-D3F	1.14	1.01	0.97	1.00	0.97	1.02
LASSO-D3	1.05	0.93*	1.05	0.93	0.95	0.98
LASSO-D3F	1.30*	1.02	0.98	1.00	1.00	1.06
AdaLASSO-D3	0.91	1.00	1.01	0.97	1.01	0.98
AdaLASSO-D3F	1.25*	1.05*	0.97*	1.00	1.00	1.05
EN-D3	1.03	1.01	1.06	0.95*	0.96*	1.00
EN-D3F	1.27*	1.02	0.99*	1.00	1.00	1.05
AdaEN-D3	0.99	1.01	1.00	0.99*	1.00	1.00
AdaEN-D3F	1.24	1.03	0.97*	1.00	1.00	1.05
LSTM-D3	0.99	1.02*	0.98	1.00	0.97	0.99
LSTM-D3F	1.02	1.01	1.00	0.99	1.01*	1.00
avg	1.11	1.02	1.00	0.99	1.00	-

Mixing Frequencies Approaches: RMSE D3/D2

Models	n=0	n=1	n=2	n=3	n=4	avg
BBoost-D3F	1.05	0.99	1.00	1.01*	1.01	1.01
CBoost-D3F	1.07	1.04	1.02	1.02	0.99	1.03
CSR-D3	1.28	1.01	0.98	1.01	1.00	1.06
CSR-D3F	1.14	1.02	1.00	1.00	1.00	1.03
Bag-D3	0.94	1.00	1.01	1.01	1.00	0.99
Bag-D3F	1.10	1.01	0.99	1.02*	1.00	1.02
BTree-D3	1.00	1.06*	0.98	1.01	1.01	1.01
BTree-D3F	1.10	1.09	0.99	0.99	1.05*	1.04
RF-D3	0.98	1.01	0.99	0.99	1.00	0.99
RF-D3F	1.03	1.01	1.00	0.99	1.01*	1.01
SVR-D3	1.00	1.00	1.00	1.00	0.99	1.00
SVR-D3F	1.00	1.00	0.99	1.00	0.99	1.00
Ridge-D3	1.11	0.99	0.99	1.00	1.00	1.02
Ridge-D3F	1.09*	1.04	1.03	1.06*	1.01	1.04
LASSO-D3	1.00	0.90	1.02	1.01	1.01	0.99
LASSO-D3F	1.15	1.03*	0.99	1.00	1.00	1.04
AdaLASSO-D3	1.03	0.91	1.05	1.04	0.96	1.00
AdaLASSO-D3F	1.07*	1.06	0.99	1.01	1.00	1.03
EN-D3	1.01	0.91	0.99	1.00	0.99	0.98
EN-D3F	1.16*	1.04*	1.00	1.00	1.00*	1.04
AdaEN-D3	1.14	0.92	0.99	1.00	0.97*	1.00
AdaEN-D3F	1.08*	1.04*	0.99	1.00	0.99*	1.02
LSTM-D3	0.98	0.99	0.99	0.97	1.00	0.99
LSTM-D3F	1.02	1.00	1.00	1.00	1.00	1.00
avg	1.06	1.00	1.00	1.01	1.00	-

Mixing Frequencies Approaches: Key Takeaways

- Models trained on D1(F) generally yield more precise nowcasts compared to those trained on D2(F) and D3(F), although the improvements vary by model.
- Only a limited number of pairs exhibit statistically significant differences (Giacomini-White test), though **all favor D1(F)** for **nowcasting**.
- Some models show substantial improvement when switching to D1(F), while others minimal benefits or even losses, highlighting the importance of a **model-specific assessment of information sets**.
- Differences across information sets become negligible for longer forecasting horizons, with RMSE and MAE ratios approaching unity.

Comparing Information Sets

Do Individual Series Add Value Beyond Factors?

Factor-Only and Composite Sets: RMSE D_xF/D_x

Models	n=0	n=1	n=2	n=3	n=4	avg
CSR-D1F	1.23	1.00	1.00	0.99	0.99	1.04
CSR-D2F	1.24	1.00	0.98	1.00	0.99	1.04
CSR-D3F	1.10	1.01	1.00	0.99	0.99	1.02
Bag-D1F	0.90	1.01	1.02	1.01	1.00	0.99
Bag-D2F	1.05*	0.97	1.00	1.00	0.99	1.00
Bag-D3F	1.23*	0.98	0.98	1.00	1.00	1.04
BTree-D1F	0.93	1.00	0.97	1.00	0.98	0.98
BTree-D2F	0.96	0.99	0.99	1.01	0.98	0.99
BTree-D3F	1.05	1.02	1.01	0.99	1.02	1.02
RF-D1F	1.01	0.99	0.99	1.00	0.99	1.00
RF-D2F	1.01	1.01	1.00	1.00	0.99	1.00
RF-D3F	1.07*	1.00	1.00	1.00	1.00	1.02
SVR-D1F	1.00	1.00	1.01	1.01	1.00	1.00
SVR-D2F	1.00	1.00	1.00	1.00	1.00	1.00
SVR-D3F	1.01	1.00	1.00	1.00	1.00	1.00
LASSO-D1F	0.98	1.00	1.03	1.01	1.00	1.00
LASSO-D2F	1.19*	0.98	0.99	0.99	1.00	1.03
LASSO-D3F	1.23*	1.02	1.00	0.99	1.00	1.05

Factor-Only and Composite Sets: RMSE D_xF/D_x (cont.)

AdaLASSO-D1F	0.96	0.93*	1.02	0.91*	0.93*	0.95
AdaLASSO-D2F	1.07	0.89	0.97*	0.98	0.99	0.98
AdaLASSO-D3F	1.14	1.05	0.94	0.98	0.98	1.02
Ridge-D1F	0.89	1.01	1.00	1.00	1.02	0.98
Ridge-D2F	1.04	0.89	0.98	1.00	0.95*	0.97
Ridge-D3F	1.10	1.01	0.96	1.02	0.99	1.02
EN-D1F	0.96	1.01	1.06	0.95*	0.96*	0.99
EN-D2F	1.03	0.90	0.98	1.00	1.00	0.98
EN-D3F	1.18*	1.02	0.99	1.00	1.00*	1.04
AdaEN-D1F	0.89	1.00	1.01	1.00	1.00	0.98
AdaEN-D2F	1.18	0.90	0.98	1.00	0.97*	1.01
AdaEN-D3F	1.11	1.02	0.98	1.00	1.00*	1.02
LSTM-D1F	0.99	0.99	0.98	1.01	0.97	0.99
LSTM-D2F	0.97	0.98	0.99	0.98*	1.00	0.98
LSTM-D3F	1.01	0.98	0.99	1.01	1.01	1.00
SgLASSO-D3F	1.03	0.98	0.99	1.01	1.00	1.00
avg	1.05	0.99	0.99	1.00	0.99	-

Factor-Only and Composite Sets: Key Takeaways

- **Individual series** contain predictive content **beyond the factors derived from them**, improving **nowcasts** on average. However:
  - Lower nowcast errors are primarily observed in mixed-frequency sets (D2 and D3), while for quarterly-aggregated data, factor-only (D1F) outperform composite sets (D1).
  - Only a limited number of pairs exhibit statistically significant differences (Giacomini-White test), though **all favor mixed-frequency sets (D2, D3)** at the **nowcast** horizon.
- All factor-only sets (D1F, D2F, D3F) appear to be preferable on average at the **one-quarter ahead** horizon, especially when considering the MAE criterion.
- The choice between factor-only and composite sets highly depends on the algorithm and horizon.

In summary:

- **Nowcast horizon** (for both RMSE and MAE):
  - Mixed-frequency sets: D2 $>$ D2F and D3 $>$ D3F,
  - Quarterly set: D1F $>$ D1.
- **1-quarter ahead**: $D_xF > D_x$ (primarily for MAE).
- **2-quarters ahead**: Some differences between composite and factor-only sets are observed, but ML (and other) models overall showed no gains over simple univariate benchmarks.

Factor-Only and Composite Sets: Key Takeaways

For the **nowcast** horizon where evidence is somewhat clearer, one could further experiment combining the components identified in the two pairwise comparison exercises:

$$\Downarrow$$

quarterly **factors** (D1F)

$$+$$

individual predictors from mixed-frequency panel
in an unrestricted fashion (D2)

Horse Race Results

Ranked Model Performance

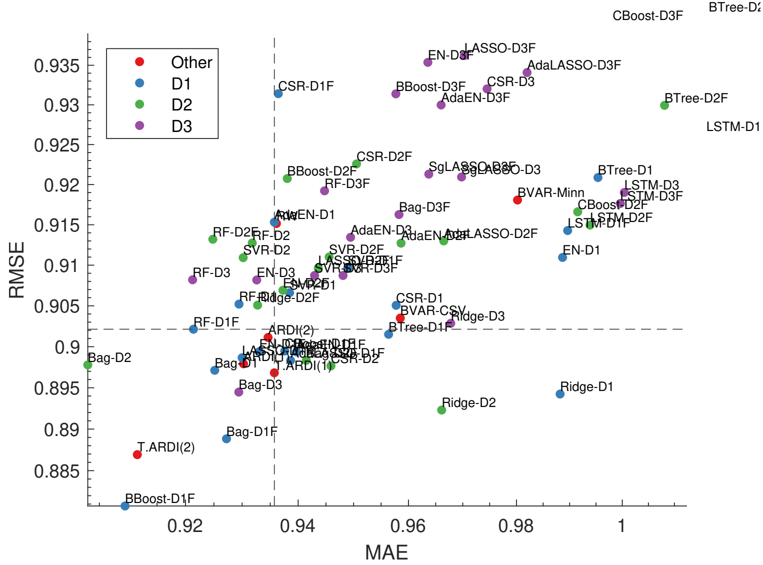
Quarterly Monitoring: 5-Horizon Avg. **RMSE** Ranking

Models	n=0	n=1	n=2	n=3	n=4	avg	avgMCS
BBoost-D1F	**0.60**	**0.88**	0.93	1.00	1.00	**0.88**	0.72
T.ARDI(2)	0.61	0.91	0.93	1.00	0.99	0.89	0.64
Bag-D1F	0.62	0.90	0.93	1.00	1.00	0.89	0.53
Ridge-D2	0.62	0.91	0.93	1.00	1.00	0.89	0.42
Ridge-D1	0.66	0.91	0.91	1.00	1.00	0.89	0.70
Bag-D3	0.62	0.92	0.93	1.01	1.00	0.89	0.50
T.ARDI(1)	0.60	0.96	0.93	1.00	1.00	0.90	0.59
Bag-D1	0.69	0.90	0.91	0.99	1.00	0.90	**0.75**
CSR-D2	0.62	0.92	0.95	1.00	1.01	0.90	0.53
Bag-D2	0.66	0.92	0.92	0.99	1.00	0.90	0.53
ARDI(1)	0.62	0.94	0.93	1.00	1.00	0.90	0.46
AdaLASSO-D1F	0.64	0.91	0.94	1.00	1.00	0.90	0.48
Bag-D2F	0.69	0.89	0.92	**0.99**	0.99	0.90	0.61
LASSO-D1F	0.65	0.91	0.94	1.00	1.00	0.90	0.43
AdaEN-D1F	0.65	0.91	0.94	1.00	1.00	0.90	0.43
EN-D1F	0.65	0.91	0.94	1.00	1.00	0.90	0.43
CBoost-D1F	0.66	0.89	0.93	1.02	1.00	0.90	0.44
ARDI(2)	0.64	0.94	0.93	1.00	1.00	0.90	0.50
BTree-D1F	0.67	0.90	0.91	1.01	1.01	0.90	0.56
RF-D1F	0.70	0.90	0.92	1.00	0.99	0.90	0.61

Quarterly Monitoring: 5-Horizon Avg. **MAE** Ranking

Models	n=0	n=1	n=2	n=3	n=4	avg	avgMCS
Bag-D2	0.72	0.88	**0.89**	1.00	1.01	**0.90**	**0.78**
BBoost-D1F	0.78	**0.81**	0.93	1.01	1.01	0.91	0.59
T.ARDI(2)	0.78	0.87	0.91	1.01	1.00	0.91	0.70
RF-D3	0.79	0.90	0.92	0.99	1.00	0.92	0.62
RF-D1F	0.80	0.89	0.92	1.00	0.99	0.92	0.59
RF-D2F	0.82	0.90	0.93	1.00	**0.98**	0.92	0.51
Bag-D1	0.77	0.87	0.90	1.02	1.06	0.93	0.44
Bag-D1F	0.77	0.88	0.92	1.02	1.04	0.93	0.48
Bag-D3	**0.72**	0.89	0.94	1.05	1.04	0.93	0.45
RF-D1	0.78	0.90	0.93	1.02	1.02	0.93	0.30
LASSO-D1F	0.79	0.89	0.96	1.01	1.00	0.93	0.36
SVR-D2	0.84	0.90	0.92	1.00	1.00	0.93	0.56
ARDI(1)	0.80	0.93	0.92	1.00	1.00	0.93	0.58
RF-D2	0.81	0.90	0.94	1.01	1.00	0.93	0.41
EN-D3	0.82	0.92	0.93	1.00	1.00	0.93	0.59
Ridge-D2F	0.80	0.86	0.91	1.03	1.07	0.93	0.37
EN-D1F	0.80	0.89	0.97	1.01	1.00	0.93	0.38
ARDI(2)	0.79	0.93	0.94	1.02	1.00	0.93	0.43
AdaEN-D1	0.79	0.88	0.96	1.05	1.00	0.94	0.52
T.ARDI(1)	0.75	0.94	0.93	1.01	1.06	0.94	0.57

Quarterly Monitoring: Avg. RMSE and MAE



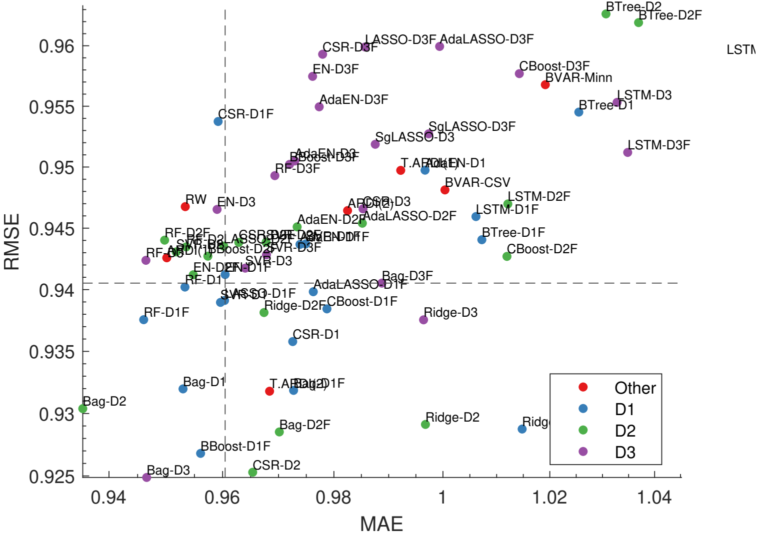
Monthly Monitoring: 5-Horizon Avg. **RMSE** Ranking

Models	n=0	n=1	n=2	n=3	n=4	avg
Bag-D3	0.73	0.93	0.97	1.00	0.99	**0.92**
CSR-D2	**0.71**	0.94	0.98	1.00	1.00	0.93
BBoost-D1F	0.72	0.92	0.98	1.01	1.00	0.93
Bag-D2F	0.77	0.91	0.97	0.99	1.00	0.93
Ridge-D1	0.76	0.92	0.96	1.00	1.00	0.93
Ridge-D2	0.74	0.93	0.97	1.00	1.00	0.93
Bag-D2	0.77	0.92	0.97	0.99	1.00	0.93
T.ARDI(2)	0.75	0.93	0.98	1.00	1.01	0.93
Bag-D1F	0.76	0.93	0.97	1.00	1.00	0.93
Bag-D1	0.78	0.92	0.96	1.00	1.00	0.93
CSR-D1	0.76	0.93	0.97	1.00	1.01	0.94
Ridge-D3	0.79	0.92	0.97	1.00	1.00	0.94
RF-D1F	0.82	0.92	0.97	1.00	**0.99**	0.94
Ridge-D2F	0.81	0.91	0.97	1.00	1.00	0.94
CBoost-D1F	0.77	0.93	0.98	1.01	1.00	0.94
SVR-D1	0.82	0.92	0.96	0.99	1.00	0.94
LASSO-D1F	0.79	0.93	0.98	1.00	1.00	0.94
AdaLASSO-D1F	0.78	0.94	0.98	1.00	1.00	0.94
RF-D1	0.80	0.92	0.97	1.00	1.00	0.94
Bag-D3F	0.82	0.92	0.97	1.00	0.99	0.94

Monthly Monitoring: 5-Horizon Avg. **MAE** Ranking

Models	n=0	n=1	n=2	n=3	n=4	avg
Bag-D2	0.77	0.89	0.96	1.01	1.04	**0.94**
RF-D1F	0.86	0.91	0.97	**0.99**	1.00	0.95
RF-D3	0.84	0.93	**0.96**	0.99	1.02	0.95
Bag-D3	**0.77**	0.91	0.99	1.03	1.03	0.95
RF-D2F	0.87	0.93	0.97	1.00	0.99	0.95
ARDI(1)	0.86	0.92	0.97	0.99	1.00	0.95
SVR-D2	0.87	0.91	0.97	1.00	1.01	0.95
Bag-D1	0.81	0.90	0.97	1.03	1.05	0.95
RF-D1	0.84	0.92	0.97	1.01	1.02	0.95
RW	0.89	0.92	0.97	0.99	1.00	0.95
RF-D2	0.85	0.92	0.97	1.01	1.01	0.95
EN-D2F	0.88	0.93	0.97	1.00	1.00	0.95
BBoost-D1F	0.81	0.92	0.99	1.04	1.02	0.96
BBoost-D2F	0.93	0.90	0.96	1.00	1.01	0.96
EN-D3	0.90	0.92	0.97	1.00	1.00	0.96
CSR-D1F	0.88	0.91	0.97	1.01	1.02	0.96
SVR-D1	0.88	0.92	0.97	1.01	1.02	0.96
LASSO-D2F	0.91	0.93	0.97	1.00	0.99	0.96
LASSO-D1F	0.85	0.93	1.00	1.02	1.00	0.96
EN-D1F	0.85	0.93	0.99	1.01	1.01	0.96

Monthly Monitoring: Avg. RMSE and MAE



The full picture: Key Takeaways

- Block-boosting with quarterly factors (D1F) performs consistently well across horizons, and error measures. Target-ARDI's remain a competitive alternative delivering strong performance.
- **Bagging linear regressions** proves to be a strong contender for nowcasting and short-term GDP forecasting, which suggests it merits more attention in future ML-based studies focusing on macroeconomic forecasting, beyond its limited application in the current literature.
- The significant presence of **linear models** in the upper quartiles of the fully fledged comparison, highlights their effectiveness: 16 out of the top 20 performers in the RMSE ranking, and 13 in the MAE ranking.
- Certain algorithms dominate specific rankings:
  - RF frequently appear in the upper quartile of the MAE ranking, but not the RMSE ranking.
  - Ridge regressions show strong results in terms of RMSE, but weaker MAE performance.
- Models trained on temporally aggregated high-frequency predictors (D1/D1F) are frequently found in the upper quartile, underscoring their effectiveness in handling mixed-frequency data.
- However, in the **monthly monitoring** setting, information sets containing high-frequency panels (D2 and D3) gain prominence, reflecting the added value of high-frequency predictors at capturing useful within-quarter signals, early in the quarter.

Implications and Concluding Remarks

- **Machine learning algorithms outperform benchmarks** in nowcasting and short-term GDP forecasting, though the single best model differs across horizons and metrics.
- **Temporally aggregated information sets** (obtained by equal-weight averaging) improve **nowcast** accuracy: average RMSE reductions range from 3.8% to 9.9%, depending on the MIDAS scheme benchmarked.
- **Factors + individual series $>$ factors only**
  Overall, composite sets yield 4.8% average **nowcast** error reduction, and up to 18.7% gain across models, especially linear ones. However, for the quarterly-aggregated information sets, factor-only (D1F) outperform composite sets (D1).

Implications and Concluding Remarks: Horse Race

- **Most promising algorithms**:
  - L_2 **factor-boosting** with linear base learner and **block-wise** lags (Bai & Ng, 2008).
  - Ensembles of **bagged regressions** with **linear** base learner.
- **Linear methods** dominate the top quartile in RMSE rankings (including bagged regressions, factor-models, and penalized regressions), while **nonlinear ML models** (particularly **Random Forests**) gain prominence in the MAE-based ranking, suggesting proneness to large forecast errors.
- **Target-ARDI** with hard-thresholding preselection (Stock & Watson, 2002; Bai & Ng, 2009) consistently found among top performers.
- Pooling all models together in a horse race, and assuming GDP is monitored at quarter-end, identifies the **quarterly-factor information set (D1F)** as a particularly effective companion set for numerous algorithms.
- Key insights from the quarterly OOS experiment remain valid under **monthly OOS monitoring**, with **mixed-frequency sets** gaining relevance, suggesting the **importance of high-frequency information early in the quarter**.

Thank you!

Questions or comments?

`ckarag.github.io`

References

Andreou, E., Ghysels, E., and Kourtellos, A. (2010). Regression models with mixed sampling frequencies. *Journal of Econometrics*, 158(2):246–261.

Babii, A., Ghysels, E., and Striaukas, J. (2022). Machine Learning Time Series Regressions with an Application to Nowcasting. *Journal of Business & Economic Statistics*, 40(3):1094–1106.

Bai, J. and Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317.

Bai, J. and Ng, S. (2009). Boosting diffusion indices. *Journal of Applied Econometrics*, 24(4):607–629.

Ballarin, G., Dellaportas, P., Grigoryeva, L., Hirt, M., van Huellen, S., and Ortega, J.-P. (2024). Reservoir computing for macroeconomic forecasting with mixed-frequency data. *International Journal of Forecasting*, 40(3):1206–1237.

Bańbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2013). Now-casting and the real-time data flow. In *Handbook of economic forecasting*, volume 2, pages 195–237. Elsevier.

Bańbura, M., Giannone, D., and Reichlin, L. (2010). Large bayesian vector auto regressions. *Journal of applied Econometrics*, 25(1):71–92.

References (cont.)

Bergmeir, C., Hyndman, R. J., and Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120:70–83.

Beyhum, J. and Striaukas, J. (2023). Sparse plus dense midas regressions and nowcasting during the covid pandemic. *arXiv e-prints*, pages arXiv–2306.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Buehlmann, P. (2006). Boosting for high-dimensional linear models. *The Annals of Statistics*, 34(2):559–583.

Carriero, A., Clark, T. E., and Marcellino, M. (2016). Common Drifting Volatility in Large Bayesian VARs. *Journal of Business & Economic Statistics*, 34(3):375–390.

Chan, J. C. (2020). Large Bayesian VARs: A Flexible Kronecker Error Covariance Structure. *Journal of Business & Economic Statistics*, 38(1):68–79.

Chernozhukov, V., Hansen, C., and Liao, Y. (2017). A lava attack on the recovery of sums of dense and sparse signals. *The Annals of Statistics*, 45(1):39 – 76.

Chow, G. C. and Lin, A.-l. (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The review of Economics and Statistics*, pages 372–375.

References (cont.)

Clark, T. E., Leonard, S., Marcellino, M., and Wegmüller, P. (2022). Weekly nowcasting us inflation with enhanced random forests. Technical report, Mimeo.

D'Agostino, A., Giannone, D., and Surico, P. (2007). (un) predictability and macroeconomic stability. *CEPR Discussion Papers*, 6594.

Drucker, H., Burges, C. J., Kaufman, L., Smola, A., and Vapnik, V. (1996). Support vector regression machines. *Advances in neural information processing systems*, 9.

Elliott, G., Gargano, A., and Timmermann, A. (2013). Complete subset regressions. *Journal of Econometrics*, 177(2):357–373. Dynamic Econometric Modeling and Forecasting.

Elliott, G., Rothenberg, T. J., and Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica: Journal of the Econometric Society*, pages 813–836.

Fan, J., Masini, R. P., and Medeiros, M. C. (2023). Bridging factor and sparse models. *The Annals of Statistics*, 51(4):1692–1717.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189 – 1232.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2005). There is a risk-return trade-off after all. *Journal of financial economics*, 76(3):509–548.

References (cont.)

Giannone, D., Lenza, M., and Primiceri, G. E. (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica*, 89(5):2409–2437.

Goulet Coulombe, P., Leroux, M., Stevanovic, D., and Surprenant, S. (2021). Macroeconomic data transformations matter. *International Journal of Forecasting*, 37(4):1338–1354.

Goulet Coulombe, P., Leroux, M., Stevanovic, D., and Surprenant, S. (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*, 37(5):920–964.

Hepenstrick, C. and Marcellino, M. (2019). Forecasting gross domestic product growth with large unbalanced data sets: the mixed frequency three-pass regression filter. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(1):69–99.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Huber, F., Koop, G., Onorante, L., Pfarrhofer, M., and Schreiner, J. (2020). Nowcasting in a pandemic using non-parametric mixed frequency VARs. *Journal of Econometrics*.

References (cont.)

Inoue, A. and Kilian, L. (2008). How useful is bagging in forecasting economic time series? a case study of u.s. consumer price inflation. *Journal of the American Statistical Association*, 103(482):511–522.

Kock, A. B. and Callot, L. (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics*, 186(2):325–344.

Kotchoni, R., Leroux, M., and Stevanovic, D. (2019). Macroeconomic forecast accuracy in a data-rich environment. *Journal of Applied Econometrics*, 34(7):1050–1072.

Medeiros, M. C., Vasconcelos, G. F. R., Álvaro Veiga, and Zilberman, E. (2021). Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics*, 39(1):98–119.

Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288.

Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, 101(476):1418–1429.

References (cont.)

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.

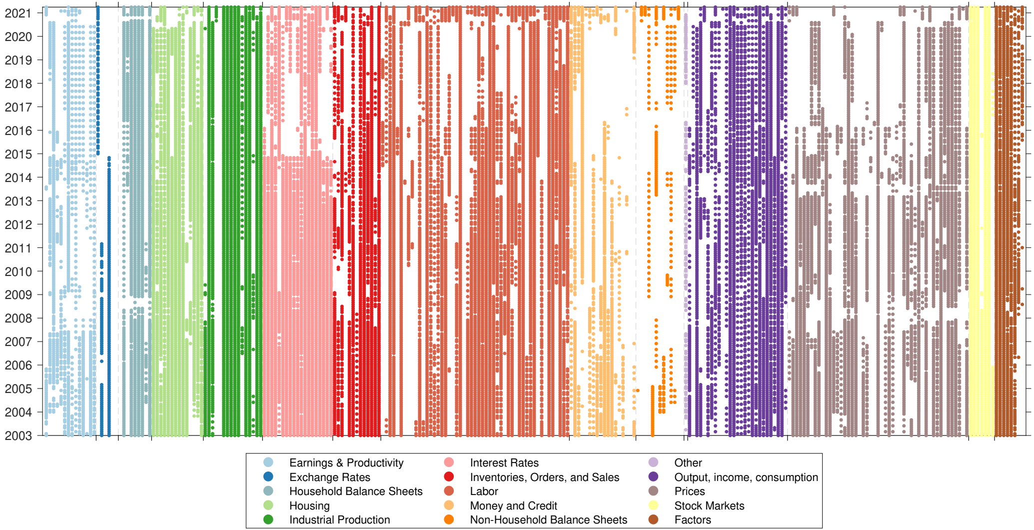
Series Distribution by Economic Category

Category	Count	Percent
Monthly		
Earnings & Productivity	3	1.8
Exchange Rates	6	3.5
Household Balance Sheets	1	0.6
Housing	13	7.6
Industrial Production	14	8.2
Interest Rates	18	10.5
Inventories, Orders & Sales	12	7.0
Labor Market	48	28.1
Money & Credit	13	7.6
Other	1	0.6
Output, Income & Consumption	6	3.5
Prices	29	17.0
Stock Markets	7	4.1
Quarterly		
Earnings & Productivity	11	12.6
Household Balance Sheets	8	9.2
Housing	1	1.1
Industrial Production	2	2.3
Interest Rates	1	1.1
Inventories, Orders & Sales	1	1.1
Labor Market	3	3.4
Money & Credit	5	5.7
Non-Household Balance Sheets	13	14.9
Output, Income & Consumption	22	25.3
Prices	20	23.0

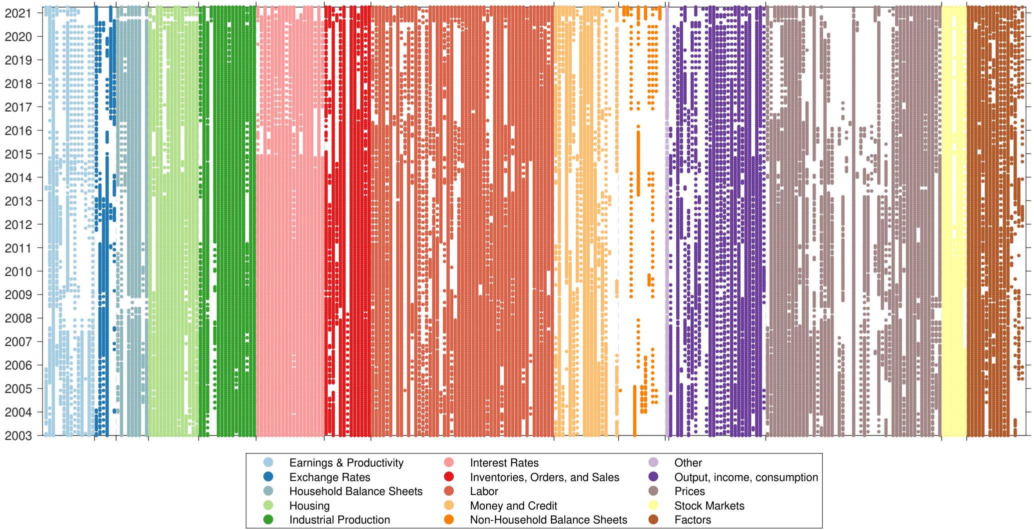
Distribution of Stationarity-Transformation Codes

Δ	Present Study		Original *tcode*	
	Count	Percent	Count	Percent
Monthly				
0	34	19.88	29	16.96
1	113	66.08	101	59.06
2	24	14.04	41	23.98
Quarterly				
0	3	3.45	7	8.05
1	70	80.46	60	68.97
2	14	16.09	20	22.99

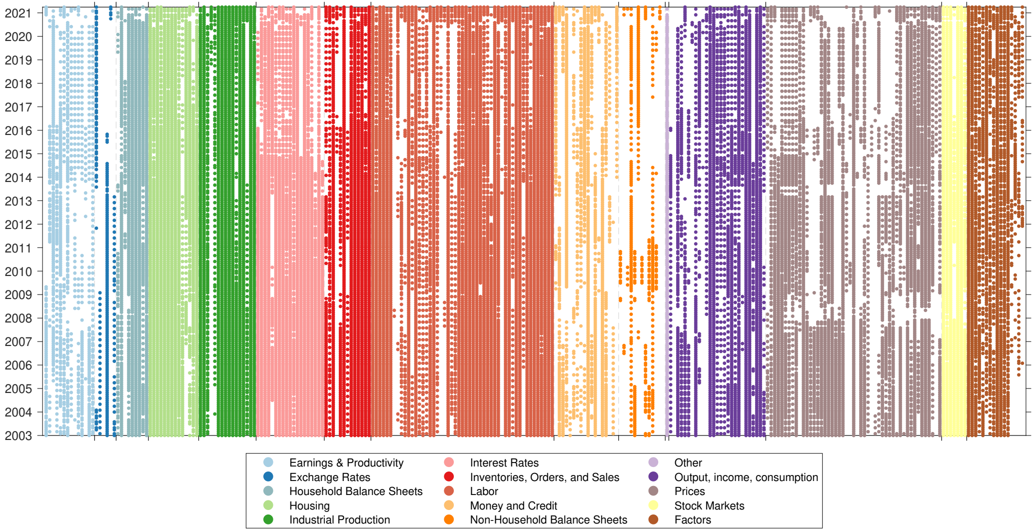
Sparsity Pattern for **Hard-thresholding** on **D1**



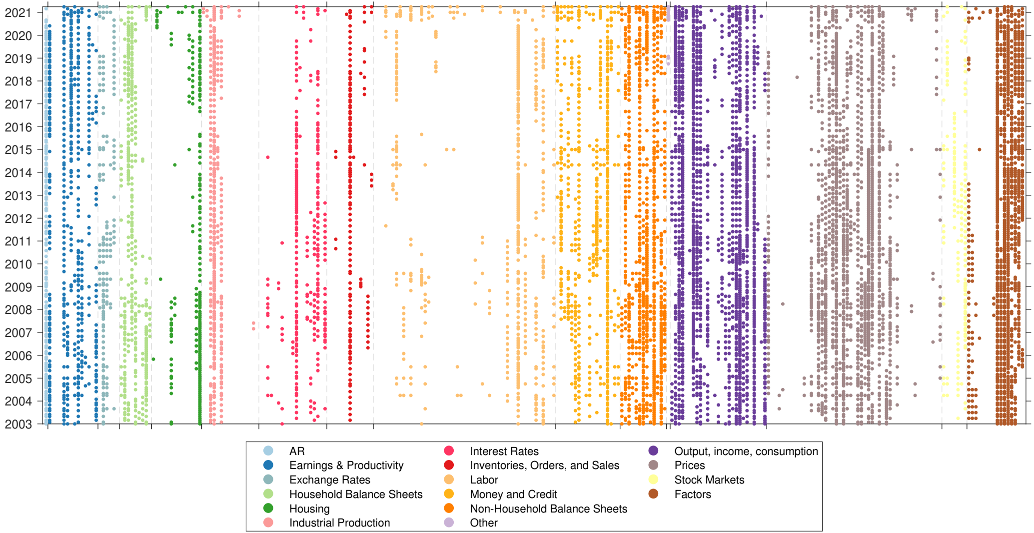
Sparsity Pattern for **Hard-thresholding** on **D2**



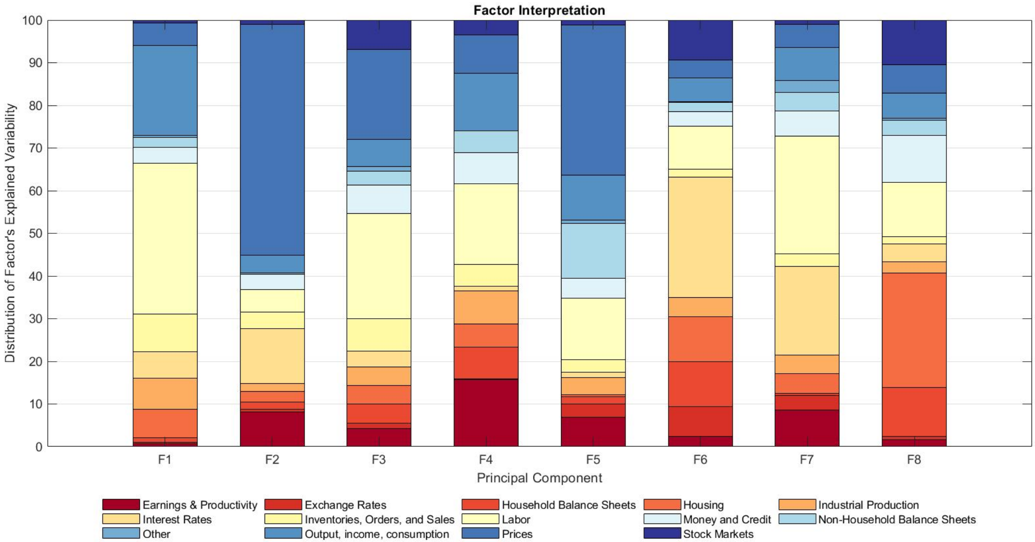
Sparsity Pattern for **Hard-thresholding** on **D3**



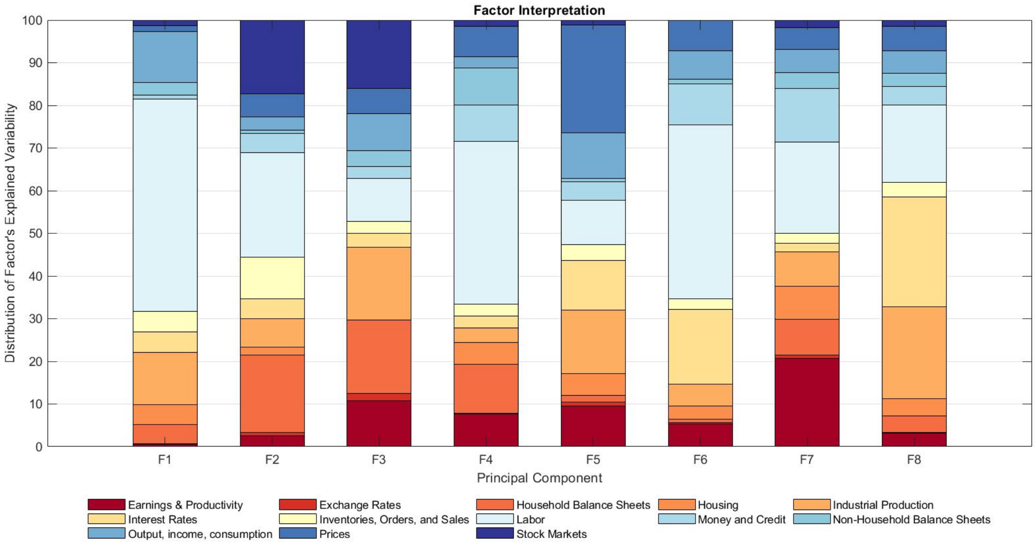
Sparsity Pattern for **Sg-LASSO-MIDAS** on **D3**



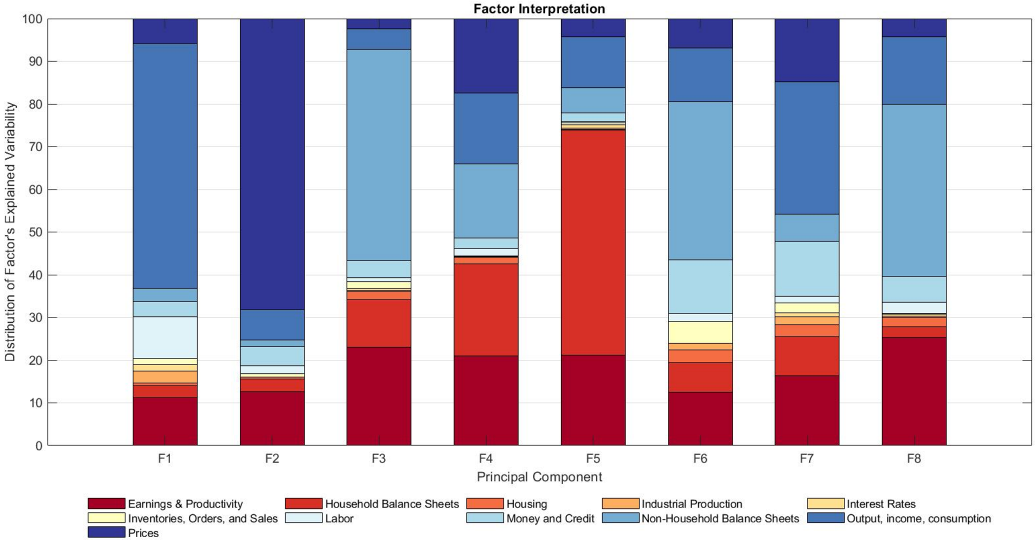
Factor Interpretation for Quarterly Set **D1** (258 series)



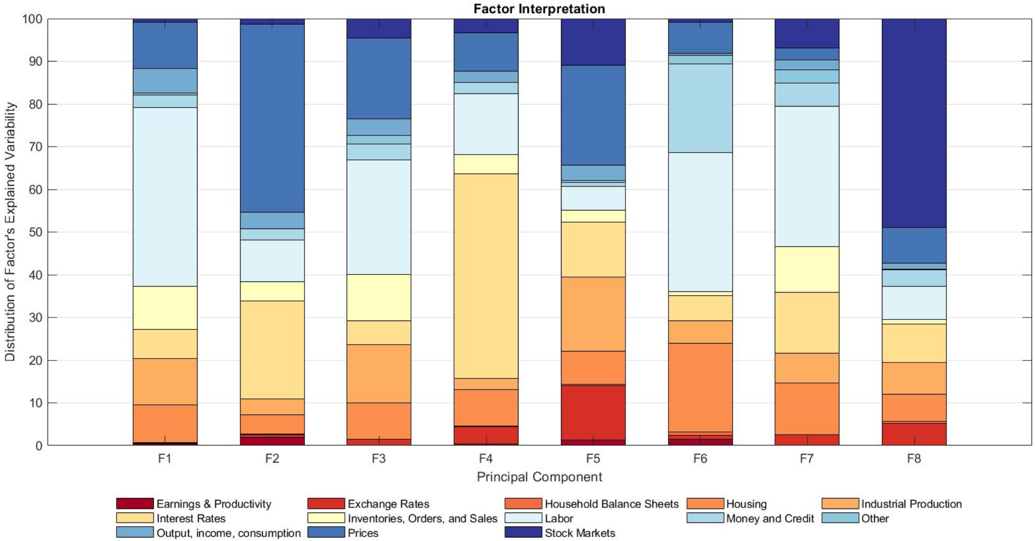
Factor Interpretation for Quarterly Set **D1**: Targeted-Predictors



Factor Interpretation for Mixed-Frequency Set **D2**: Quarterly Panel (87)



Factor Interpretation for Mixed-Frequency Set **D2**: Monthly Panel (171)



Detailed Rankings: All 85 Candidate Models, **RMSE, EoM3**

Models	n=0	n=1	n=2	n=3	n=4	avg	avgMCS
BBoost-D1F	**0.60**	**0.88**	0.93	1.00	1.00	**0.88**	0.72
T.ARDI(2)	0.61	0.91	0.93	1.00	0.99	0.89	0.64
Bag-D1F	0.62	0.90	0.93	1.00	1.00	0.89	0.53
Ridge-D2	0.62	0.91	0.93	1.00	1.00	0.89	0.47
Ridge-D1	0.66	0.91	0.91	1.00	1.00	0.89	0.69
Bag-D3	0.62	0.92	0.93	1.01	1.00	0.89	0.50
T.ARDI(1)	0.60	0.96	0.93	1.00	1.00	0.90	0.59
Bag-D1	0.69	0.90	0.91	0.99	1.00	0.90	**0.73**
CSR-D2	0.62	0.92	0.95	1.00	1.01	0.90	0.52
Bag-D2	0.66	0.92	0.92	0.99	1.00	0.90	0.55
ARDI(1)	0.62	0.94	0.93	1.00	1.00	0.90	0.45
AdaLASSO-D1F	0.64	0.91	0.94	1.00	1.00	0.90	0.45
Bag-D2F	0.69	0.89	0.92	**0.99**	0.99	0.90	0.61
LASSO-D1F	0.65	0.91	0.94	1.00	1.00	0.90	0.42
AdaEN-D1F	0.65	0.91	0.94	1.00	1.00	0.90	0.41
EN-D1F	0.65	0.91	0.94	1.00	1.00	0.90	0.42
CBoost-D1F	0.66	0.89	0.93	1.02	1.00	0.90	0.44
ARDI(2)	0.64	0.94	0.93	1.00	1.00	0.90	0.48
BTree-D1F	0.67	0.90	0.91	1.01	1.01	0.90	0.56
RF-D1F	0.70	0.90	0.92	1.00	0.99	0.90	0.62
Ridge-D3	0.69	0.90	0.92	1.01	1.00	0.90	0.44
BVAR-CSV	0.65	0.91	0.93	1.01	1.01	0.90	0.31
CSR-D1	0.66	0.92	0.93	1.01	1.01	0.91	0.40
Ridge-D2F	0.71	0.89	0.92	1.00	1.00	0.91	0.51
RF-D1	0.69	0.90	0.93	1.00	1.00	0.91	0.46
SVR-D1	0.73	0.90	0.92	0.99	0.99	0.91	0.56
EN-D2F	0.71	0.90	0.92	1.00	1.00	0.91	0.47
EN-D3	0.70	0.91	0.93	1.00	1.00	0.91	0.50

Detailed Rankings: All 85 Candidate Models, **RMSE, EoM3** (cont.)

RF-D3	0.70	0.92	0.93	0.99	1.00	0.91	0.56
SVR-D3	0.73	0.90	0.92	1.00	0.99	0.91	0.46
SVR-D3F	0.73	0.90	0.92	1.00	**0.99**	0.91	0.45
LASSO-D2F	0.73	0.90	0.92	1.00	1.00	0.91	0.55
SVR-D1F	0.73	0.90	0.92	1.00	1.00	0.91	0.58
SVR-D2	0.73	0.91	0.93	1.00	1.00	0.91	0.54
EN-D1	0.68	0.90	**0.88**	1.05	1.04	0.91	0.45
SVR-D2F	0.73	0.90	0.93	1.00	0.99	0.91	0.46
AdaEN-D2F	0.74	0.90	0.92	1.00	1.01	0.91	0.50
RF-D2	0.72	0.90	0.94	1.00	1.00	0.91	0.43
AdaLASSO-D2F	0.75	0.90	0.92	0.99	1.00	0.91	0.53
RF-D2F	0.73	0.91	0.93	1.00	0.99	0.91	0.49
AdaEN-D3	0.72	0.92	0.93	1.00	1.00	0.91	0.46
LSTM-D1F	0.75	0.90	0.92	1.01	0.99	0.91	0.48
LSTM-D2F	0.75	0.91	0.92	1.00	1.00	0.91	0.50
RW	0.74	0.90	0.93	1.00	1.00	0.92	0.47
AdaEN-D1	0.73	0.91	0.94	1.01	1.00	0.92	0.43
Bag-D3F	0.76	0.90	0.92	1.01	0.99	0.92	0.41
CBoost-D2F	0.74	0.92	0.92	1.00	1.00	0.92	0.38
LSTM-D3F	0.76	0.91	0.92	1.00	1.00	0.92	0.48
BVAR-Minn	0.70	0.95	0.93	1.00	1.02	0.92	0.44
LSTM-D3	0.75	0.93	0.93	0.99	1.00	0.92	0.51
RF-D3F	0.75	0.92	0.93	0.99	1.00	0.92	0.48
AdaEN-D2	0.63	1.00	0.94	1.00	1.03	0.92	0.36
BBoost-D2F	0.80	0.90	0.91	0.99	1.00	0.92	0.66
BTree-D1	0.72	0.90	0.94	1.01	1.03	0.92	0.25
SgLASSO-D3	0.75	0.91	0.93	1.00	1.00	0.92	0.35
SgLASSO-D3F	0.77	0.90	0.92	1.01	1.00	0.92	0.33
CSR-D2F	0.76	0.92	0.93	1.00	1.00	0.92	0.43
LASSO-D3	0.71	0.91	0.97	1.02	1.02	0.92	0.38
LSTM-D1	0.76	0.91	0.94	1.00	1.02	0.93	0.50
EN-D2	0.69	1.00	0.94	1.00	1.00	0.93	0.39

Detailed Rankings: All 85 Candidate Models, **RMSE, EoM3** (cont.)

BTree-D2F	0.74	0.92	0.96	1.03	0.99	0.93	0.30
AdaEN-D3F	0.80	0.94	0.92	1.00	1.00	0.93	0.51
BBoost-D3F	0.84	0.89	0.91	1.01	1.00	0.93	0.50
CSR-D1F	0.82	0.92	0.93	1.00	1.00	0.93	0.56
CSR-D3	0.79	0.93	0.93	1.00	1.01	0.93	0.30
AdaLASSO-D3	0.70	0.92	0.99	1.03	1.02	0.93	0.35
LSTM-D2	0.77	0.93	0.94	1.02	1.00	0.93	0.38
AdaLASSO-D3F	0.81	0.95	0.91	1.00	1.00	0.93	0.50
EN-D3F	0.83	0.93	0.92	1.00	1.00	0.94	0.41
Ridge-D1F	0.68	0.92	0.98	1.06	1.04	0.94	0.20
LASSO-D3F	0.84	0.92	0.92	1.00	1.00	0.94	0.40
AdaLASSO-D2	0.68	1.00	0.94	0.99	1.06	0.94	0.36
LASSO-D2	0.71	1.00	0.95	1.01	1.01	0.94	0.34
CBoost-D3F	0.80	0.95	0.94	1.02	0.99	0.94	0.31
BTree-D2	0.77	0.93	0.97	1.02	1.02	0.94	0.18
Ridge-D3F	0.77	0.93	0.95	1.06	1.01	0.94	0.28
CSR-D3F	0.86	0.94	0.93	1.00	1.00	0.94	0.41
AdaLASSO-D1	0.77	0.92	0.98	1.06	1.01	0.95	0.19
LASSO-D1	0.67	0.98	0.92	1.10	1.07	0.95	0.24
BTree-D3	0.77	0.99	0.95	1.02	1.02	0.95	0.20
BTree-D3F	0.81	1.00	0.96	1.02	1.04	0.97	0.24
AR(BIC)	1.00	1.01	1.00	1.00	1.00	1.00	0.39
AR(CV)	1.03	1.01	1.00	1.00	1.00	1.01	0.39
AR(4)	1.06	1.04	1.01	1.00	1.00	1.02	0.38
AR(1)	2.08	1.73	1.68	1.58	1.59	-	0.42

Notes: The last column shows the average MCS p-value (squared loss) across horizons. Highlighted cells indicate models in the 60% MCS. Lower p-value suggests the model is less likely to in the MCS. Errors are relative to AR(1), for which absolute values are shown. Models are ranked by 5-horizon average relative error. Bold indicates lowest error or highest MCS p-value.

Detailed Rankings: All 85 Candidate Models, **MAE, EoM3**

Models	n=0	n=1	n=2	n=3	n=4	avg	avgMCS
Bag-D2	0.72	0.88	**0.89**	1.00	1.01	**0.90**	**0.78**
BBoost-D1F	0.78	**0.81**	0.93	1.01	1.01	0.91	0.58
T.ARDI(2)	0.78	0.87	0.91	1.01	1.00	0.91	0.71
RF-D3	0.79	0.90	0.92	0.99	1.00	0.92	0.61
RF-D1F	0.80	0.89	0.92	1.00	0.99	0.92	0.58
RF-D2F	0.82	0.90	0.93	1.00	**0.98**	0.92	0.50
Bag-D1	0.77	0.87	0.90	1.02	1.06	0.93	0.44
Bag-D1F	0.77	0.88	0.92	1.02	1.04	0.93	0.48
Bag-D3	**0.72**	0.89	0.94	1.05	1.04	0.93	0.45
RF-D1	0.78	0.90	0.93	1.02	1.02	0.93	0.30
LASSO-D1F	0.79	0.89	0.96	1.01	1.00	0.93	0.34
SVR-D2	0.84	0.90	0.92	1.00	1.00	0.93	0.55
ARDI(1)	0.80	0.93	0.92	1.00	1.00	0.93	0.59
RF-D2	0.81	0.90	0.94	1.01	1.00	0.93	0.41
EN-D3	0.82	0.92	0.93	1.00	1.00	0.93	0.57
Ridge-D2F	0.80	0.86	0.91	1.03	1.07	0.93	0.37
EN-D1F	0.80	0.89	0.97	1.01	1.00	0.93	0.36
ARDI(2)	0.79	0.93	0.94	1.02	1.00	0.93	0.41
AdaEN-D1	0.79	0.88	0.96	1.05	1.00	0.94	0.51
T.ARDI(1)	0.75	0.94	0.93	1.01	1.06	0.94	0.56
RW	0.87	0.90	0.92	0.99	1.00	0.94	0.57
CSR-D1F	0.84	0.89	0.92	1.00	1.03	0.94	0.43
EN-D2F	0.85	0.91	0.93	0.99	1.01	0.94	0.52
CBoost-D1F	0.80	0.87	0.95	1.06	1.01	0.94	0.36
BBoost-D2F	0.94	0.87	0.91	**0.98**	0.99	0.94	0.60
SVR-D1	0.85	0.91	0.92	1.00	1.01	0.94	0.41
AdaLASSO-D1F	0.81	0.87	0.98	1.04	1.00	0.94	0.28
AdaEN-D1F	0.80	0.88	0.99	1.02	1.00	0.94	0.30

Detailed Rankings: All 85 Candidate Models, **MAE, EoM3** (cont.)

Bag-D2F	0.80	0.89	0.93	1.02	1.07	0.94	0.32
SVR-D3	0.85	0.91	0.93	1.02	1.00	0.94	0.45
LASSO-D2F	0.88	0.91	0.94	0.99	1.00	0.94	0.49
RF-D3F	0.86	0.91	0.93	1.00	1.03	0.94	0.39
SVR-D2F	0.86	0.92	0.94	1.01	1.01	0.95	0.44
CSR-D2	0.74	0.94	1.00	1.01	1.04	0.95	0.42
SVR-D3F	0.88	0.91	0.93	1.02	1.01	0.95	0.45
SVR-D1F	0.86	0.91	0.95	1.01	1.02	0.95	0.32
AdaEN-D3	0.87	0.95	0.93	1.00	1.00	0.95	0.49
CSR-D2F	0.86	0.93	0.92	1.01	1.03	0.95	0.53
BTree-D1F	0.77	0.99	0.95	1.05	1.01	0.96	0.53
BBoost-D3F	0.96	0.86	0.94	1.03	1.00	0.96	0.40
CSR-D1	0.75	0.93	0.98	1.05	1.08	0.96	0.31
Bag-D3F	0.87	0.86	0.94	1.08	1.04	0.96	0.19
BVAR-CSV	0.76	0.93	0.97	1.05	1.09	0.96	0.40
AdaEN-D2F	0.89	0.93	0.94	1.01	1.03	0.96	0.32
EN-D3F	0.98	0.95	0.91	1.00	0.99	0.96	0.58
SgLASSO-D3F	0.94	0.87	0.93	1.05	1.02	0.96	0.27
AdaEN-D3F	0.98	0.96	0.90	1.01	0.98	0.97	0.57
Ridge-D2	0.72	0.97	0.98	1.07	1.08	0.97	0.31
AdaLASSO-D2F	0.90	0.94	0.98	1.01	1.01	0.97	0.40
CSR-D3F	0.92	0.95	0.93	1.02	1.02	0.97	0.24
Ridge-D3	0.80	0.91	0.92	1.09	1.12	0.97	0.19
SgLASSO-D3	0.91	0.95	0.94	1.02	1.02	0.97	0.24
LASSO-D3F	1.00	0.93	0.91	1.02	1.00	0.97	0.48
CSR-D3	0.90	0.93	0.95	1.04	1.06	0.97	0.24
BVAR-Minn	0.80	1.01	0.94	1.05	1.10	0.98	0.44
AdaLASSO-D3F	0.98	0.98	0.90	1.04	0.99	0.98	0.46
Ridge-D1	0.78	0.97	0.96	1.10	1.14	0.99	0.18
EN-D1	0.77	0.87	0.98	1.19	1.13	0.99	0.27
LSTM-D1F	0.91	0.98	0.99	1.05	1.02	0.99	0.31
CBoost-D2F	0.89	0.90	0.96	1.11	1.10	0.99	0.15

Detailed Rankings: All 85 Candidate Models, **MAE, EoM3** (cont.)

LSTM-D2F	0.91	0.98	0.96	1.07	1.04	0.99	0.16
BTree-D1	0.85	0.95	1.00	1.04	1.13	1.00	0.08
CBoost-D3F	0.92	1.00	0.98	1.08	1.02	1.00	0.07
LSTM-D3F	0.96	0.94	0.97	1.07	1.06	1.00	0.22
LSTM-D3	0.89	1.00	0.99	1.06	1.07	1.00	0.13
AR(BIC)	1.00	1.01	1.01	1.01	1.00	1.01	0.40
BTree-D2F	0.88	0.97	1.06	1.10	1.03	1.01	0.19
LSTM-D1	0.92	0.95	1.02	1.09	1.09	1.02	0.12
BTree-D2	0.89	1.00	1.02	1.07	1.12	1.02	0.08
AR(CV)	1.06	1.01	1.00	1.01	1.00	1.02	0.38
LSTM-D2	0.91	1.03	1.01	1.11	1.08	1.03	0.13
AR(4)	1.11	1.04	1.01	1.01	1.00	1.03	0.38
BTree-D3	0.91	1.08	1.03	1.13	1.10	1.05	0.04
EN-D2	0.82	1.10	1.08	1.18	1.11	1.06	0.03
AdaEN-D2	0.80	1.11	1.07	1.16	1.17	1.06	0.11
AdaLASSO-D1	0.90	1.02	1.13	1.18	1.09	1.06	0.04
BTree-D3F	0.94	1.08	1.01	1.14	1.17	1.07	0.03
LASSO-D2	0.84	1.10	1.10	1.21	1.13	1.08	0.02
AdaLASSO-D2	0.84	1.08	1.09	1.14	1.24	1.08	0.01
LASSO-D3	0.86	0.99	1.17	1.16	1.23	1.08	0.03
Ridge-D1F	0.87	0.96	1.19	1.26	1.24	1.10	0.01
AdaLASSO-D3	0.89	1.03	1.21	1.17	1.21	1.10	0.04
Ridge-D3F	0.90	0.99	1.12	1.30	1.22	1.11	0.01
LASSO-D1	0.76	1.26	1.16	1.38	1.27	1.17	0.16
AR(1)	0.74	0.72	0.71	0.66	0.66	-	0.47

Notes: The last column shows the average MCS p-value (absolute loss) across horizons. Highlighted cells indicate models in the 60% MCS. Lower p-value suggests the model is less likely to in the MCS. Errors are relative to AR(1), for which absolute values are shown. Models are ranked by 5-horizon average relative error. Bold indicates lowest error or highest MCS p-value.

Detailed Rankings: All 85 Candidate Models, **RMSE, EoM**

Models	n=0	n=1	n=2	n=3	n=4	avg
Bag-D3	0.73	0.93	0.97	1.00	0.99	**0.92**
CSR-D2	**0.71**	0.94	0.98	1.00	1.00	0.93
BBoost-D1F	0.72	0.92	0.98	1.01	1.00	0.93
Bag-D2F	0.77	0.91	0.97	0.99	1.00	0.93
Ridge-D1	0.76	0.92	0.96	1.00	1.00	0.93
Ridge-D2	0.74	0.93	0.97	1.00	1.00	0.93
Bag-D2	0.77	0.92	0.97	0.99	1.00	0.93
T.ARDI(2)	0.75	0.93	0.98	1.00	1.01	0.93
Bag-D1F	0.76	0.93	0.97	1.00	1.00	0.93
Bag-D1	0.78	0.92	0.96	1.00	1.00	0.93
CSR-D1	0.76	0.93	0.97	1.00	1.01	0.94
Ridge-D3	0.79	0.92	0.97	1.00	1.00	0.94
RF-D1F	0.82	0.92	0.97	1.00	**0.99**	0.94
Ridge-D2F	0.81	0.91	0.97	1.00	1.00	0.94
CBoost-D1F	0.77	0.93	0.98	1.01	1.00	0.94
SVR-D1	0.82	0.92	0.96	0.99	1.00	0.94
LASSO-D1F	0.79	0.93	0.98	1.00	1.00	0.94
AdaLASSO-D1F	0.78	0.94	0.98	1.00	1.00	0.94
RF-D1	0.80	0.92	0.97	1.00	1.00	0.94
Bag-D3F	0.82	0.92	0.97	1.00	0.99	0.94
EN-D2F	0.82	0.92	0.97	1.00	1.00	0.94
EN-D1F	0.80	0.93	0.98	1.00	1.00	0.94
SVR-D3	0.83	0.92	0.97	0.99	1.00	0.94
RF-D3	0.81	0.94	0.97	0.99	1.00	0.94
ARDI(1)	0.80	0.94	0.97	1.00	1.00	0.94
CBoost-D2F	0.82	0.92	0.97	1.00	1.01	0.94
BBoost-D2F	0.84	0.91	0.97	1.00	1.00	0.94
SVR-D3F	0.83	0.91	0.97	1.00	1.00	0.94

Detailed Rankings: All 85 Candidate Models, **RMSE, EoM** (cont.)

SVR-D2	0.83	0.92	0.97	1.00	1.00	0.94
AdaEN-D2	0.77	0.95	0.98	1.00	1.02	0.94
EN-D2	0.78	0.97	0.98	1.00	0.99	0.94
RF-D2	0.82	0.93	0.98	1.00	1.00	0.94
LASSO-D2F	0.83	0.92	0.97	1.00	1.00	0.94
AdaEN-D1F	0.80	0.93	0.98	1.00	1.00	0.94
SVR-D1F	0.83	0.92	0.97	1.00	1.00	0.94
CSR-D2F	0.82	0.93	0.97	1.00	1.00	0.94
SVR-D2F	0.83	0.92	0.97	1.00	1.00	0.94
RF-D2F	0.83	0.93	0.97	1.00	0.99	0.94
BTree-D1F	0.81	0.92	0.98	1.00	1.01	0.94
AdaEN-D2F	0.83	0.93	0.96	1.00	1.00	0.95
AdaLASSO-D2F	0.84	0.92	0.97	1.00	1.00	0.95
AdaLASSO-D2	0.76	0.95	0.98	1.00	1.03	0.95
LSTM-D1F	0.85	0.92	0.97	1.00	0.99	0.95
ARDI(2)	0.80	0.95	0.98	1.00	1.01	0.95
EN-D3	0.83	0.92	0.98	1.00	1.00	0.95
CSR-D3	0.81	0.93	0.98	1.00	1.00	0.95
RW	0.84	0.92	0.98	1.00	1.00	0.95
LSTM-D2F	0.84	0.93	0.97	1.00	1.00	0.95
LASSO-D2	0.78	0.98	0.98	1.00	1.00	0.95
BVAR-CSV	0.81	0.93	0.98	1.01	1.01	0.95
RF-D3F	0.84	0.93	0.98	**0.99**	1.00	0.95
T.ARDI(1)	0.81	0.95	0.98	1.00	1.01	0.95
AdaEN-D1	0.82	0.93	0.99	1.01	1.00	0.95
BBoost-D3F	0.87	**0.91**	0.97	1.01	1.00	0.95
AdaEN-D3	0.85	0.93	0.98	1.00	1.00	0.95
LSTM-D3F	0.85	0.93	0.97	0.99	1.01	0.95
SgLASSO-D3	0.85	0.92	0.98	1.00	1.00	0.95
SgLASSO-D3F	0.86	0.91	0.98	1.01	1.00	0.95
CSR-D1F	0.88	0.92	0.97	1.00	1.00	0.95
EN-D1	0.81	0.94	**0.94**	1.04	1.04	0.95

Detailed Rankings: All 85 Candidate Models, **RMSE, EoM** (cont.)

BTree-D1	0.81	0.93	0.99	1.01	1.02	0.95
AdaEN-D3F	0.88	0.92	0.97	1.00	1.00	0.95
LSTM-D3	0.86	0.93	0.98	1.01	1.00	0.96
BVAR-Minn	0.85	0.93	0.97	1.01	1.01	0.96
EN-D3F	0.89	0.93	0.97	1.00	1.00	0.96
CBoost-D3F	0.85	0.94	0.98	1.01	1.00	0.96
LSTM-D1	0.87	0.94	0.98	1.00	1.01	0.96
CSR-D3F	0.89	0.94	0.97	1.00	1.00	0.96
LASSO-D3F	0.90	0.92	0.97	1.00	1.00	0.96
AdaLASSO-D3F	0.89	0.93	0.97	1.00	1.00	0.96
BTree-D2F	0.85	0.95	0.99	1.02	1.00	0.96
LSTM-D2	0.87	0.94	0.99	1.02	1.00	0.96
BTree-D2	0.84	0.95	0.99	1.01	1.02	0.96
Ridge-D3F	0.80	0.94	1.02	1.05	1.03	0.97
AdaLASSO-D1	0.83	0.95	1.02	1.04	1.02	0.97
BTree-D3	0.85	0.99	0.99	1.02	1.02	0.97
Ridge-D1F	0.78	0.96	1.03	1.06	1.05	0.98
LASSO-D3	0.83	0.93	1.02	1.06	1.06	0.98
AdaLASSO-D3	0.83	0.95	1.03	1.06	1.06	0.99
LASSO-D1	0.77	1.01	0.99	1.10	1.07	0.99
BTree-D3F	0.88	1.03	1.00	1.01	1.03	0.99
AR(BIC)	1.01	1.00	1.00	1.00	1.00	1.00
AR(CV)	1.02	1.00	1.00	1.00	1.00	1.00
AR(4)	1.05	1.02	1.00	1.00	1.00	1.01
AR(1)	1.85	1.69	1.61	1.58	1.59	-

Notes: Evaluation exercise assumes updating of predictions at the end of every month. Errors are relative to AR(1), for which absolute values are shown. Models are ranked by 5-horizon average relative error. Bold indicates lowest error.

Detailed Rankings: All 85 Candidate Models, **MAE, EoM**

Models	n=0	n=1	n=2	n=3	n=4	avg
Bag-D2	0.77	0.89	0.96	1.01	1.04	**0.94**
RF-D1F	0.86	0.91	0.97	**0.99**	1.00	0.95
RF-D3	0.84	0.93	**0.96**	0.99	1.02	0.95
Bag-D3	**0.77**	0.91	0.99	1.03	1.03	0.95
RF-D2F	0.87	0.93	0.97	1.00	0.99	0.95
ARDI(1)	0.86	0.92	0.97	0.99	1.00	0.95
SVR-D2	0.87	0.91	0.97	1.00	1.01	0.95
Bag-D1	0.81	0.90	0.97	1.03	1.05	0.95
RF-D1	0.84	0.92	0.97	1.01	1.02	0.95
RW	0.89	0.92	0.97	0.99	1.00	0.95
RF-D2	0.85	0.92	0.97	1.01	1.01	0.95
EN-D2F	0.88	0.93	0.97	1.00	1.00	0.95
BBoost-D1F	0.81	0.92	0.99	1.04	1.02	0.96
BBoost-D2F	0.93	0.90	0.96	1.00	1.01	0.96
EN-D3	0.90	0.92	0.97	1.00	1.00	0.96
CSR-D1F	0.88	0.91	0.97	1.01	1.02	0.96
SVR-D1	0.88	0.92	0.97	1.01	1.02	0.96
LASSO-D2F	0.91	0.93	0.97	1.00	0.99	0.96
LASSO-D1F	0.85	0.93	1.00	1.02	1.00	0.96
EN-D1F	0.85	0.93	0.99	1.01	1.01	0.96
CSR-D2F	0.88	0.92	0.97	1.01	1.04	0.96
SVR-D3	0.89	0.93	0.99	1.01	1.01	0.96
CSR-D2	0.78	0.96	1.00	1.03	1.05	0.97
Ridge-D2F	0.83	**0.89**	0.98	1.05	1.08	0.97
SVR-D2F	0.89	0.93	0.99	1.01	1.02	0.97
SVR-D3F	0.90	0.92	1.00	1.03	1.01	0.97
T.ARDI(2)	0.83	0.92	0.99	1.04	1.05	0.97
RF-D3F	0.91	0.94	0.98	1.00	1.01	0.97

Detailed Rankings: All 85 Candidate Models, **MAE, EoM** (cont.)

Bag-D2F	0.83	0.92	0.98	1.05	1.08	0.97
BBoost-D3F	0.93	0.91	1.00	1.02	1.00	0.97
CSR-D1	0.80	0.93	1.01	1.05	1.07	0.97
Bag-D1F	0.85	0.94	0.98	1.05	1.05	0.97
AdaEN-D3	0.94	0.94	0.97	1.00	1.01	0.97
AdaEN-D2F	0.92	0.97	0.97	1.01	1.00	0.97
AdaEN-D1F	0.87	0.95	1.01	1.02	1.01	0.97
SVR-D1F	0.89	0.93	1.00	1.02	1.04	0.97
EN-D3F	0.99	0.93	0.97	1.00	1.00	0.98
AdaLASSO-D1F	0.87	0.95	1.02	1.03	1.00	0.98
AdaEN-D3F	0.99	0.93	0.97	1.01	**0.99**	0.98
CSR-D3F	0.94	0.94	0.98	1.02	1.01	0.98
CBoost-D1F	0.86	0.94	1.02	1.05	1.02	0.98
ARDI(2)	0.87	0.97	1.00	1.03	1.04	0.98
AdaLASSO-D2F	0.93	0.97	1.00	1.03	1.00	0.99
CSR-D3	0.88	0.93	1.02	1.05	1.04	0.99
LASSO-D3F	0.99	0.92	0.98	1.01	1.02	0.99
SgLASSO-D3	0.93	0.95	1.00	1.02	1.04	0.99
Bag-D3F	0.89	0.90	1.03	1.08	1.04	0.99
T.ARDI(1)	0.88	0.94	1.00	1.08	1.06	0.99
Ridge-D3	0.84	0.93	1.02	1.10	1.09	1.00
AdaEN-D1	0.89	0.94	1.06	1.06	1.03	1.00
Ridge-D2	0.81	0.98	1.03	1.07	1.09	1.00
SgLASSO-D3F	0.96	0.92	1.01	1.07	1.02	1.00
AdaLASSO-D3F	0.99	0.95	1.00	1.04	1.02	1.00
BVAR-CSV	0.87	0.95	1.02	1.07	1.09	1.00
AR(BIC)	1.01	1.01	1.01	1.00	1.00	1.01
LSTM-D1F	0.94	0.95	1.04	1.05	1.04	1.01
BTree-D1F	0.93	1.00	1.04	1.03	1.03	1.01
AR(CV)	1.03	1.01	1.00	1.01	1.00	1.01
CBoost-D2F	0.91	0.93	1.00	1.11	1.11	1.01
LSTM-D2F	0.93	0.99	1.03	1.07	1.05	1.01

Detailed Rankings: All 85 Candidate Models, **MAE, EoM** (cont.)

CBoost-D3F	0.94	0.96	1.05	1.07	1.05	1.01
Ridge-D1	0.85	0.98	1.03	1.11	1.11	1.01
BVAR-Minn	0.93	0.96	1.01	1.08	1.10	1.02
AR(4)	1.06	1.02	1.01	1.00	1.00	1.02
BTree-D1	0.93	0.97	1.06	1.08	1.09	1.03
BTree-D2	0.92	1.02	1.03	1.06	1.12	1.03
LSTM-D3	0.95	1.00	1.04	1.08	1.08	1.03
LSTM-D3F	0.97	0.99	1.02	1.10	1.09	1.03
BTree-D2F	0.95	1.01	1.08	1.08	1.06	1.04
LSTM-D1	0.98	1.01	1.07	1.09	1.12	1.05
EN-D1	0.87	1.00	1.06	1.18	1.19	1.06
BTree-D3	0.94	1.07	1.08	1.11	1.12	1.06
LSTM-D2	0.97	1.05	1.07	1.13	1.10	1.06
AdaEN-D2	0.87	1.06	1.14	1.14	1.16	1.08
EN-D2	0.87	1.10	1.15	1.16	1.12	1.08
BTree-D3F	1.01	1.09	1.07	1.15	1.16	1.10
AdaLASSO-D2	0.89	1.09	1.16	1.15	1.20	1.10
LASSO-D2	0.88	1.11	1.16	1.19	1.14	1.10
AdaLASSO-D1	0.99	1.13	1.19	1.15	1.15	1.12
Ridge-D3F	0.90	1.03	1.25	1.31	1.29	1.15
LASSO-D3	0.96	1.03	1.21	1.27	1.32	1.16
Ridge-D1F	0.93	1.09	1.24	1.27	1.26	1.16
AdaLASSO-D3	0.99	1.08	1.24	1.27	1.34	1.18
LASSO-D1	0.98	1.34	1.26	1.39	1.34	1.26
AR(1)	0.72	0.71	0.68	0.65	0.65	-

Notes: Evaluation exercise assumes updating of predictions at the end of every month. Errors are relative to AR(1), for which absolute values are shown. Models are ranked by 5-horizon average relative error. Bold indicates lowest error.