EUROPEAN CENTRAL BANK

EUROSYSTEM

Nowcasting world trade with machine learning: a three-step approach

NBU open research seminar

15/11/2023

B. Meunier (ECB)

Joint with M. Chinn (U. Wisconsin) and S. Stumpner (BdF)

Motivations

- Data on global trade published with **significant lag**
- In the meantime, numerous **indicators** available for trade and macroeconomic environment

*What about **nowcasting?***

- World trade **highly volatile** (Bussiere et al., 2013)

*What about non-linear methods? And **machine learning?***

Data

- Target: monthly **world trade** (*volumes*) available since Jan. 2000
- **600** potential regressors identified based on the **literature** on nowcasting trade (e.g., Guichard and Rusticelli, 2011; Jakaitiene and Dees, 2012; Bahroumi et al., 2016; Martinez-Martin and Rusticelli, 2021)

200 **Trade indicators**

- PMI (surveys)
- Port traffic
- Customs trade (*values*)

300 **Macro. outlook**

- Industrial production
- Retail sales
- Business confidence

100 **Financial**

- Stock markets
- Commodity prices
- Exchange rates

The three-step approach

Pre-selection

- Ranking regressors by predictive power

Fan and Lv (2008)
Runstler (2016)

Factor extraction

- Summarizing information
- Orthogonalizing inputs

Stock and Watson (2002)
Goulet-Coulombe et al. (2022)

Machine learning

Tree-based	*Regression-based*
Random Forest (RF)	*Macro.* RF
Gradient boosting (GB)	Gradient *linear* boosting

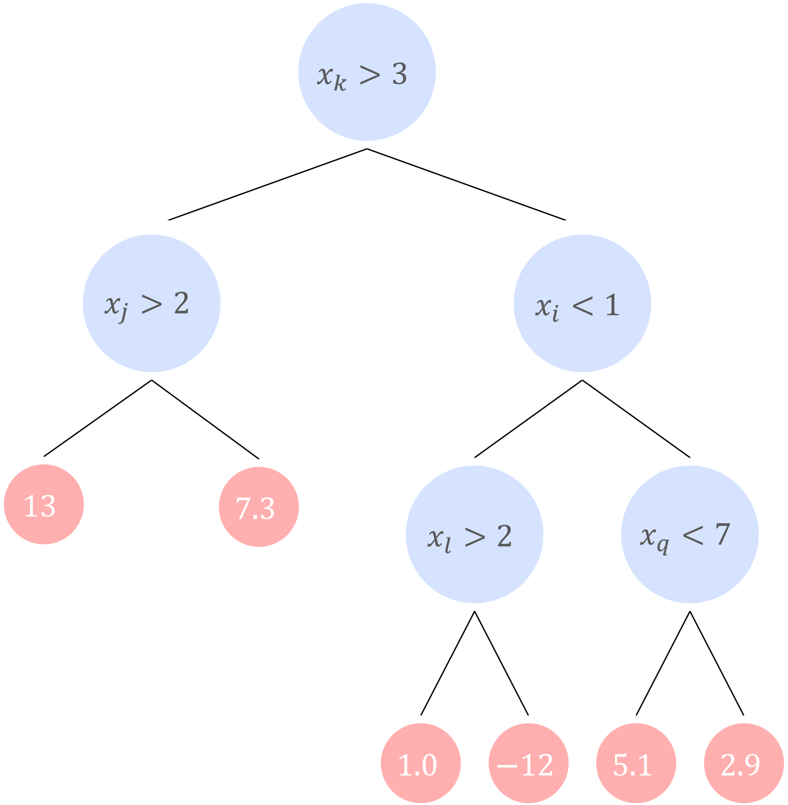
"Traditional" econometrics

Markov-switching

Quantile regression

OLS (benchmark)

Regression trees



- Set of **questions**
- Highly **flexible**
- Low bias, **high variance**
- Generally poor **predictive performances**

Boosting techniques

Gradient boosting	Gradient *linear* boosting
Boosting: **additive** method using trees as weak learners	Use of linear regression instead of trees as **weak learners**
Based on computation of the **gradient** of loss function	Regularization with **L1 and L2 penalty** terms
Overfitting controlled through **shrinkage**	Chen et al. (2016)

A detour by the forest

Random forest

- Average over un-correlated **decision trees**
- **De-correlation** of trees through
  - Bootstrapping samples (*bagging*)
  - Sub-sampling variables at each split (Breiman, 2001)

Macroeconomic random forest

- Canonical random forest **too flexible** for short time series
- Goulet-Coulombe (2020)
- Linear model between target variable and regressors but where coefficients can vary through time according to **random forest**

The macroeconomic random forest – illustrated

Can be **different** from regressors (X_t)

$$\begin{tikzcd} S_t \arrow[r, "\beta_t = \mathcal{F}(S_t)"] \arrow[rr, bend right, "Y_t = \mathcal{F}(S_t)"'] & \beta_t \arrow[r, "Y_t = \beta_t X_t"] & Y_t \end{tikzcd}$$

Horserace

Pre-selection

- Sure Independence Screening (Fan and Lv, 2008)
- t-stat-based (Jurado et al., 2015)
- LARS (Bai and Ng, 2008)
- Iterated Bayesian Model Averaging (Martinez-Martin and Rusticelli, 2021)

Factor extraction
(on pre-selected variables)

- PCA (Stock and Watson, 2002)
- 2-step (Doz et al., 2011)
- Quasi maximum likelihood (Doz et al., 2012)
- Generalized PCA (Forni et al., 2005)

Machine learning
(using factors on the RHS)

- *Tree*-based machine learning
  - Random Forest (RF)
  - Gradient boosting (GB)
- *Regression*-based machine learning
  - Macroeconomic RF
  - Linear GB
- “Traditional” econometrics
  - Markov-switching
  - Quantile regression
- OLS (benchmark)

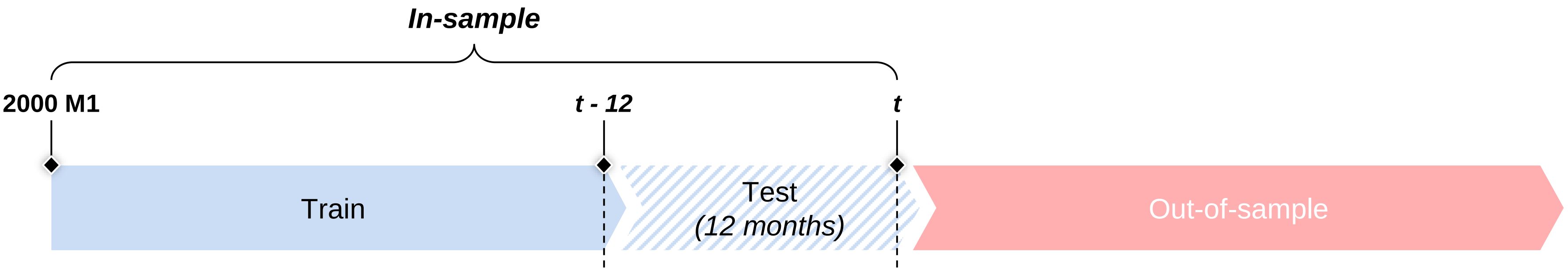
Rules of the horserace

- **Out-of-sample** predictions
- Over **Jan. 2012 – Apr. 2022**
- **Four horizons**:
  - *t-2* and *t-1* (back-casts)
  - *t* (nowcast)
  - *t+1* (forecast)
- Datasets at different **moments of the month**: 1st, 11th and 21st days

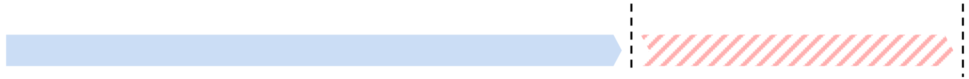
The real-time set-up

1 *Pre-selection based on the in-sample period*

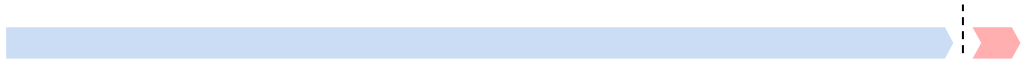
2 *Factor extraction using the pre-selected variables*



3 *Tuning of hyperparameters for machine learning: model fitted on "train" sample, validation on "test" sample*



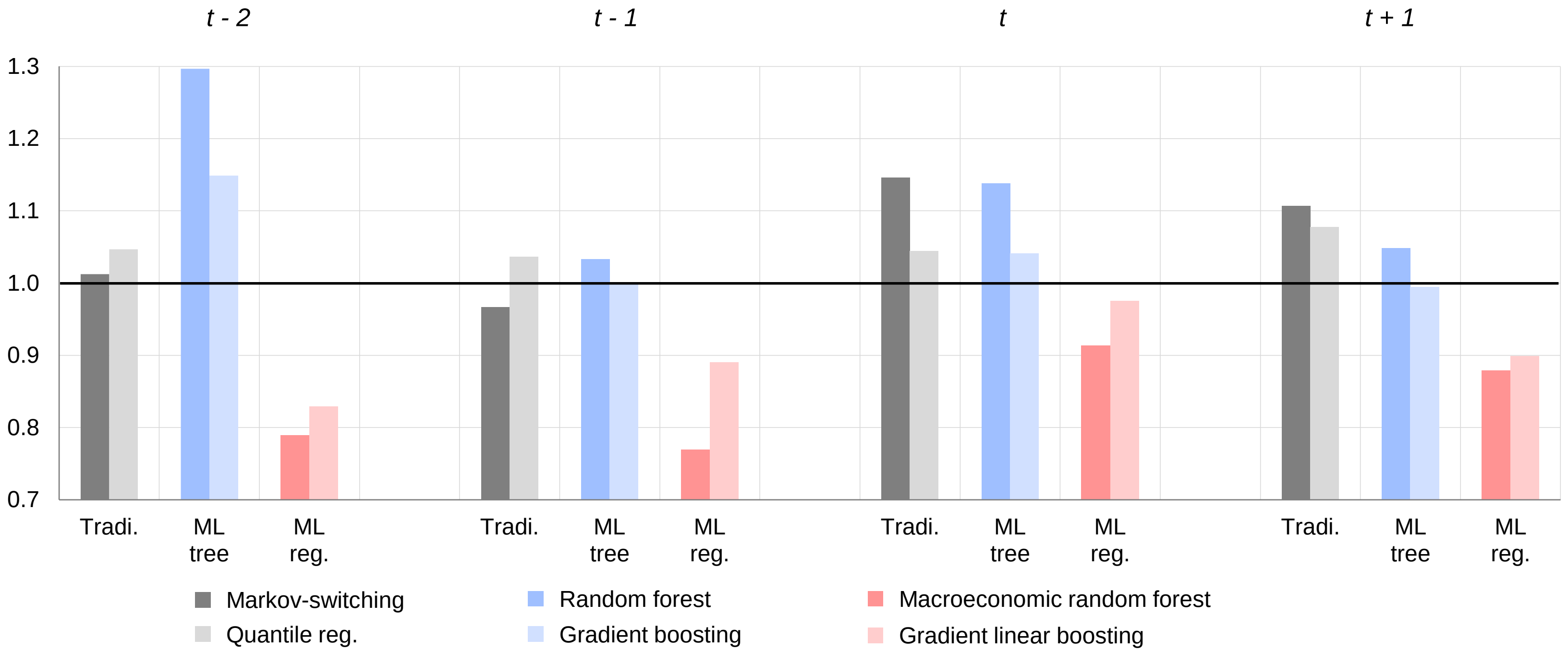
4 *Out-of-sample predictions (backcasts at t-2 and t-1, nowcast at t, and forecast at t+1)*



Results across regression techniques

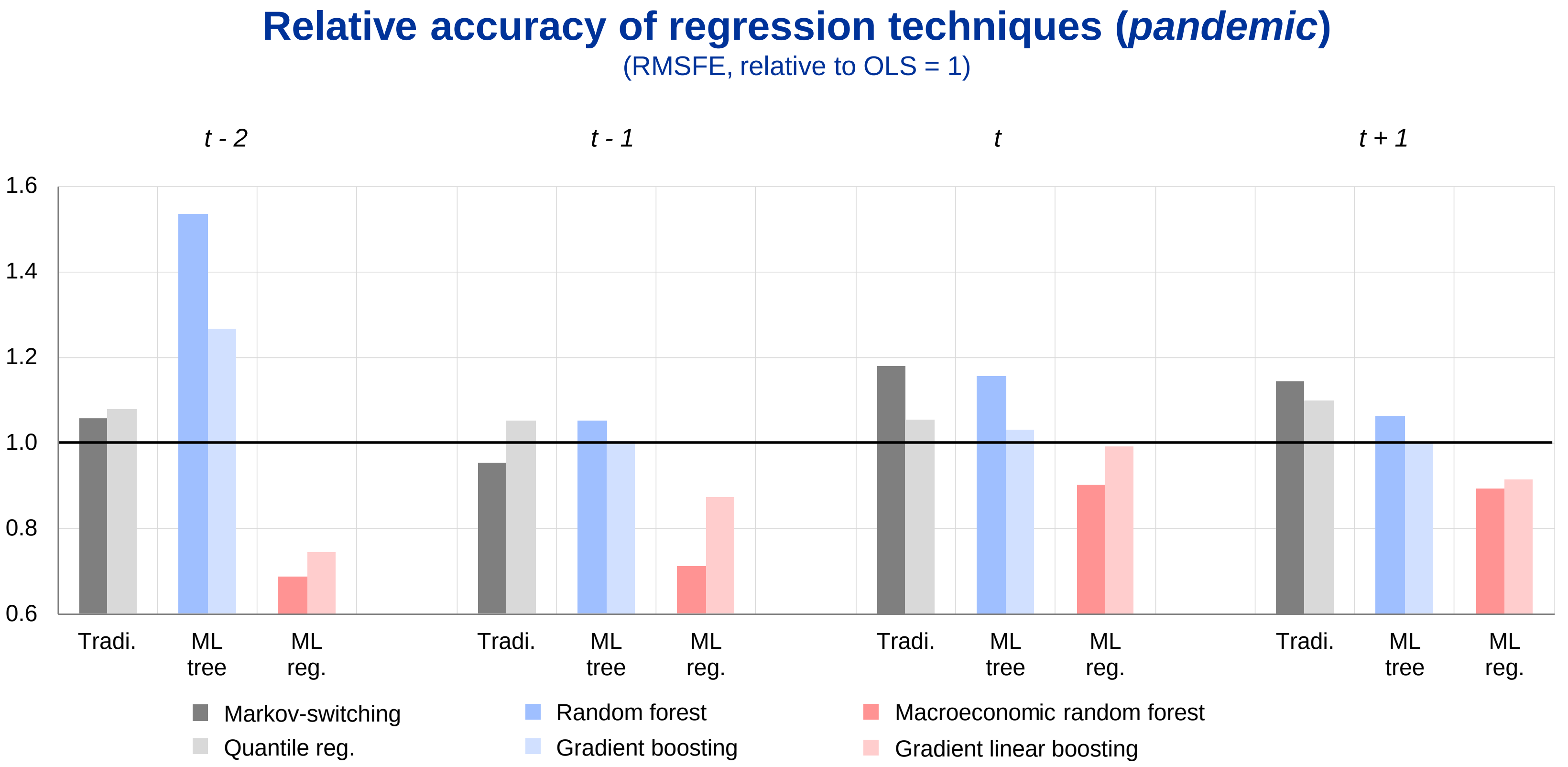
Relative accuracy of regression techniques

(RMSFE, relative to OLS = 1)



Notes: Accuracy is measured by the out-of-sample RMSE over Jan. 2012 – Apr. 2022. Performances are presented relative to the OLS benchmark (black line). Results are obtained for the average of the datasets mirroring data available to a forecaster at the 1st, 11th, and 21st days of the month, using a LARS for pre-selecting the 60 most informative regressors, with factors extracted through PCA on the pre-selected set. “Tradi.” = traditional, “ML tree” = machine learning techniques based on *decision trees*, “ML reg.” = machine learning techniques based on *linear regressions.*

Results across regression techniques – pandemic period

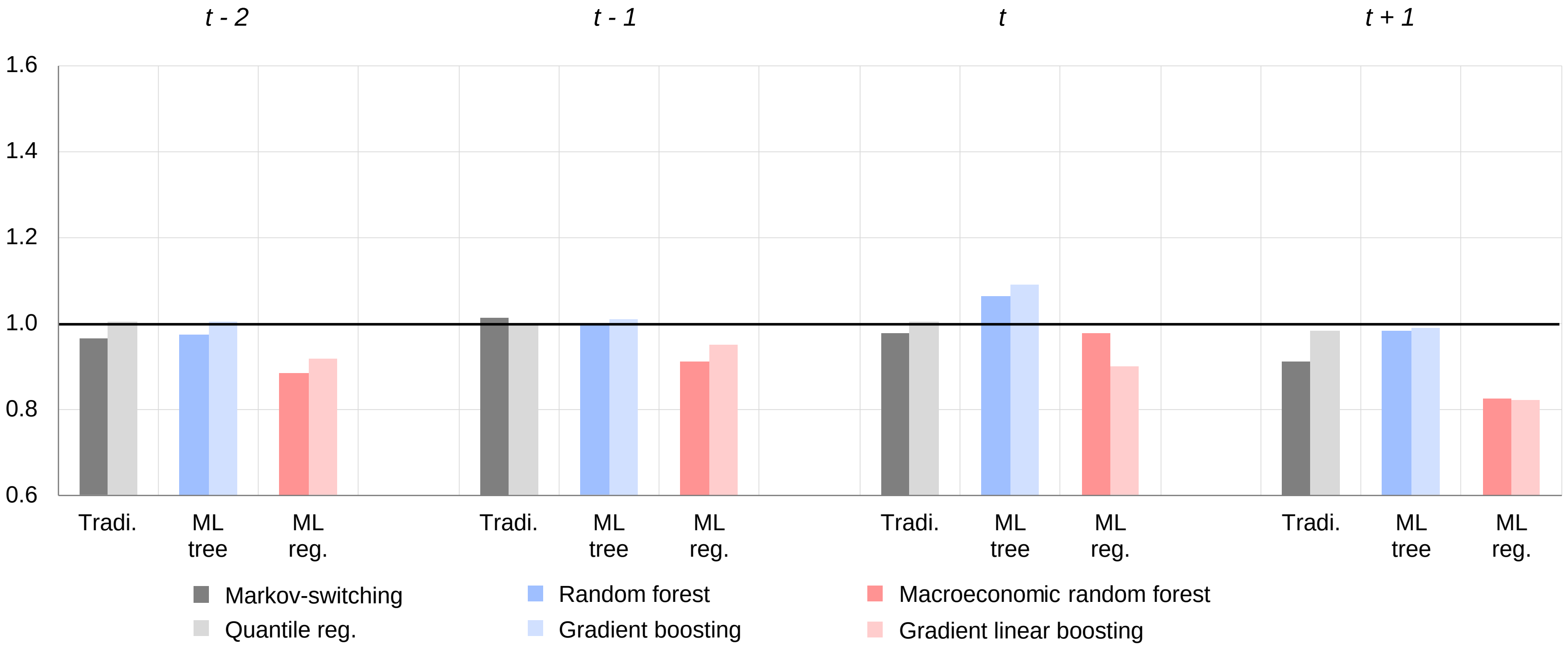


Notes: Accuracy is measured by the out-of-sample RMSE over Jan. 2020 – Dec. 2021. Performances are presented relative to the OLS benchmark (black line). Results are obtained for the average of the datasets mirroring data available to a forecaster at the 1st, 11th, and 21st days of the month, using a LARS for pre-selecting the 60 most informative regressors, with factors extracted through PCA on the pre-selected set. "Tradi." = traditional, "ML tree" = machine learning techniques based on *decision trees*, "ML reg." = machine learning techniques based on *linear regressions.*

Results across regression techniques – normal times

Relative accuracy of regression techniques (*normal times*)

(RMSFE, relative to OLS = 1)

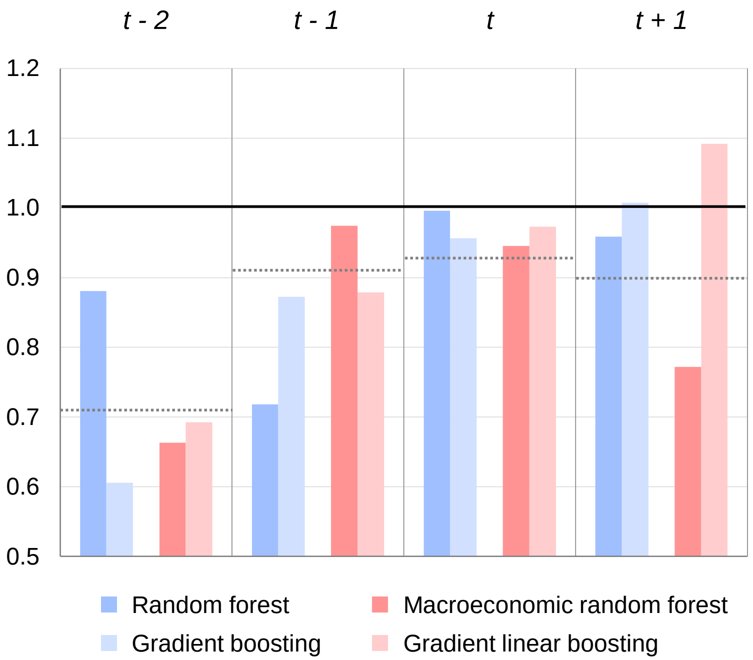


Notes: Accuracy is measured by the out-of-sample RMSE over Jan. 2012 – Dec. 2019 and Jan. 2022 – Apr. 2022. Performances are presented relative to the OLS benchmark (black line). Results are obtained for the average of the datasets mirroring data available to a forecaster at the 1st, 11th, and 21st days of the month, using a LARS for pre-selecting the 60 most informative regressors, with factors extracted through PCA on the pre-selected set. "Tradi." = traditional, "ML tree" = machine learning techniques based on *decision trees*, "ML reg." = machine learning techniques based on *linear regressions.*

Does every step really counts? The case of pre-selection

Accuracy relative to no pre-selection

(RMSFE, relative to no pre-selection = 1)



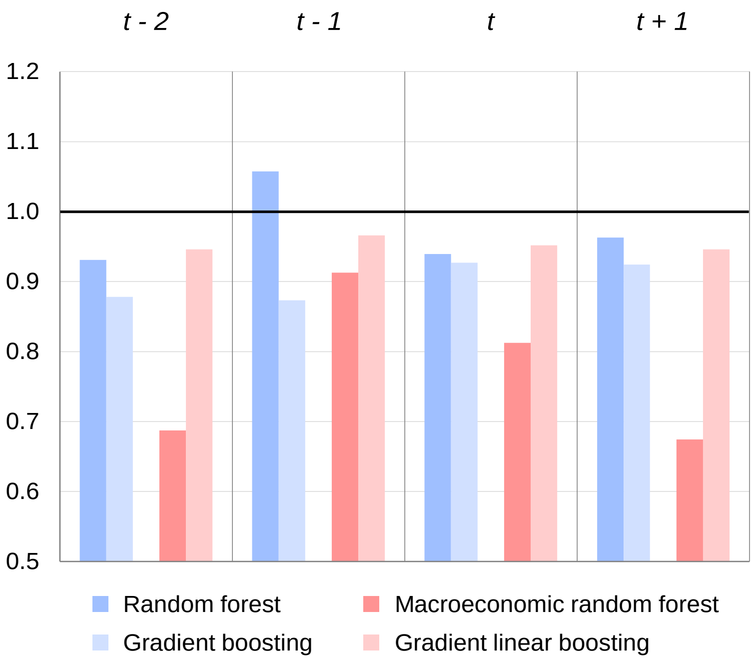
Notes: Accuracy is measured by the out-of-sample RMSE over Jan. 2012 – Apr. 2022. Performances are presented relative to the benchmark of no pre-selection (black line). Results are obtained for the average of the datasets mirroring data available to a forecaster at the 1st, 11th, and 21st days of the month, using PCA to extract factors.

- Each model is compared with the **same model WITHOUT pre-selection** – meaning a factor extraction on the full dataset
- **Accuracy gains** from pre-selection consistent across models and horizons – up to 40%

Does every step really counts? The case of factor extraction

Accuracy relative to no factors

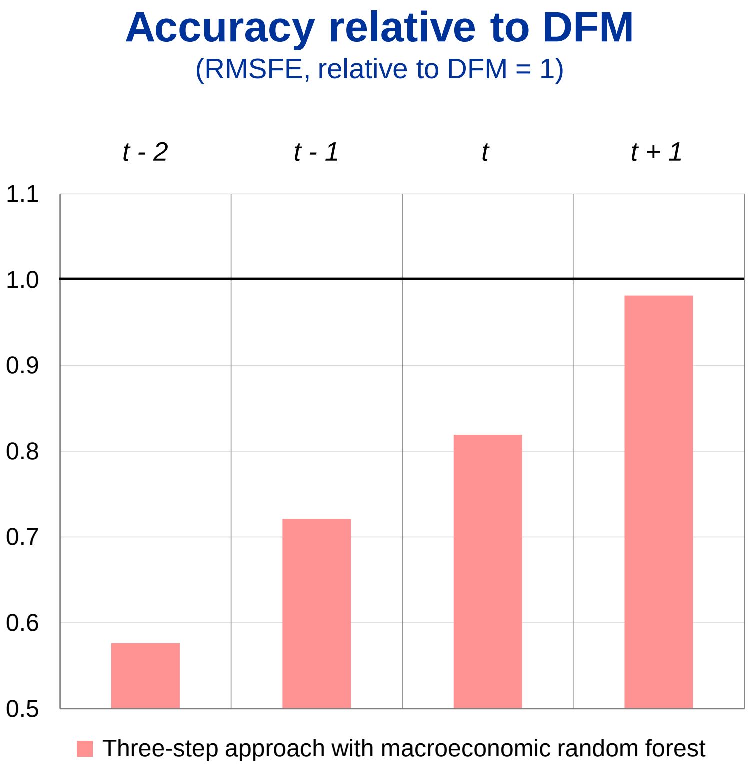
(RMSFE, relative to no factors = 1)



Notes: Accuracy is measured by the out-of-sample RMSE over Jan. 2012 – Apr. 2022. Performances are presented relative to the benchmark of no factor extraction (black line). Results are obtained for the average of the datasets mirroring data available to a forecaster at the 1st, 11th, and 21st days of the month, using LARS for pre-selecting the 60 most informative regressors.

- Each model is compared with the **same model WITHOUT factor extraction** – meaning a regression directly on the selected variables
- **Accuracy gains** from pre-selection consistent across models and horizons – around 10-15% on average

Comparison with workhorse model



Notes: Accuracy is measured by the out-of-sample RMSE over Jan. 2012 – Apr. 2022. Performances are presented relative to the DFM (black line). Results are obtained for the average of the datasets mirroring data available to a forecaster at the 1st, 11th, and 21st days of the month, using LARS for pre-selecting the 60 most informative regressors. Three-step approach uses PCA for factor extraction.

- **Dynamic factor model** (DFM) based on the quasi maximum likelihood estimator of Banbura and Modugno (2014)

- Using similar dataset of variables **pre-selected** by LARS

Key take-aways

- ***Regression*-based** machine learning outperforming significantly and consistently *tree*-based ML as well linear and non-linear benchmarks

- Performances of machine learning techniques significantly enhanced by doing first **pre-selection** and **factor extraction**

- Three-step approach outperforming a workhorse **dynamic factor model**

THANK YOU

Paper: https://www.nber.org/papers/w31419

Code: https://github.com/baptiste-meunier/NowcastingML_3step

A word on pre-selection techniques

Univariate	Multivariate

Univariate

- **Correlation-based** (SIS): ranking based on pairwise correlation with target variable (CPB trade)
- **T-stat-based**: ranking based on the t-stat of an univariate regression on the target variable (CPB trade) and lags of the endogenous variable

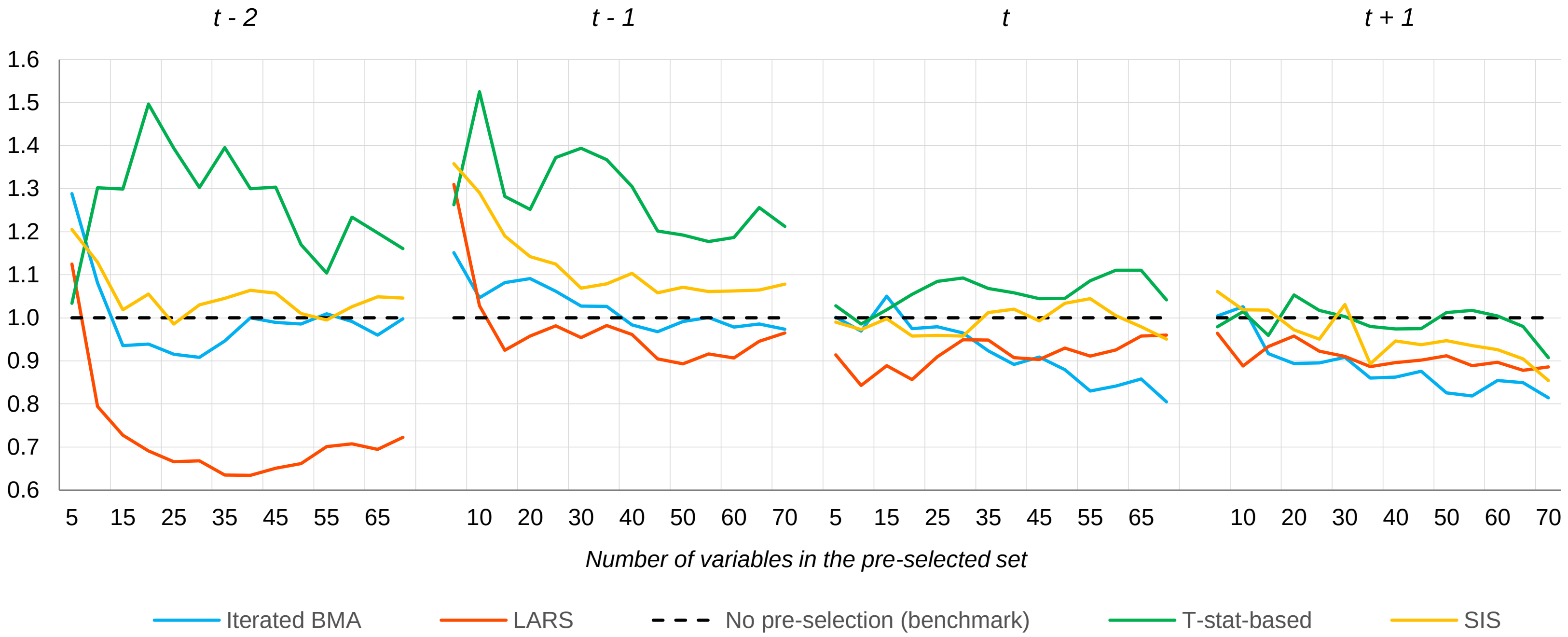
Multivariate

- **LARS**: iterative forward selection method
- **Iterated BMA**: repeated calls to Bayesian Model Averaging which delivers posterior inclusion probability for each variable – depending on inclusion on “best” models

Results across pre-selection techniques

Relative accuracy of pre-selection techniques

(RMSFE, relative to the no-preselection benchmark = 1)

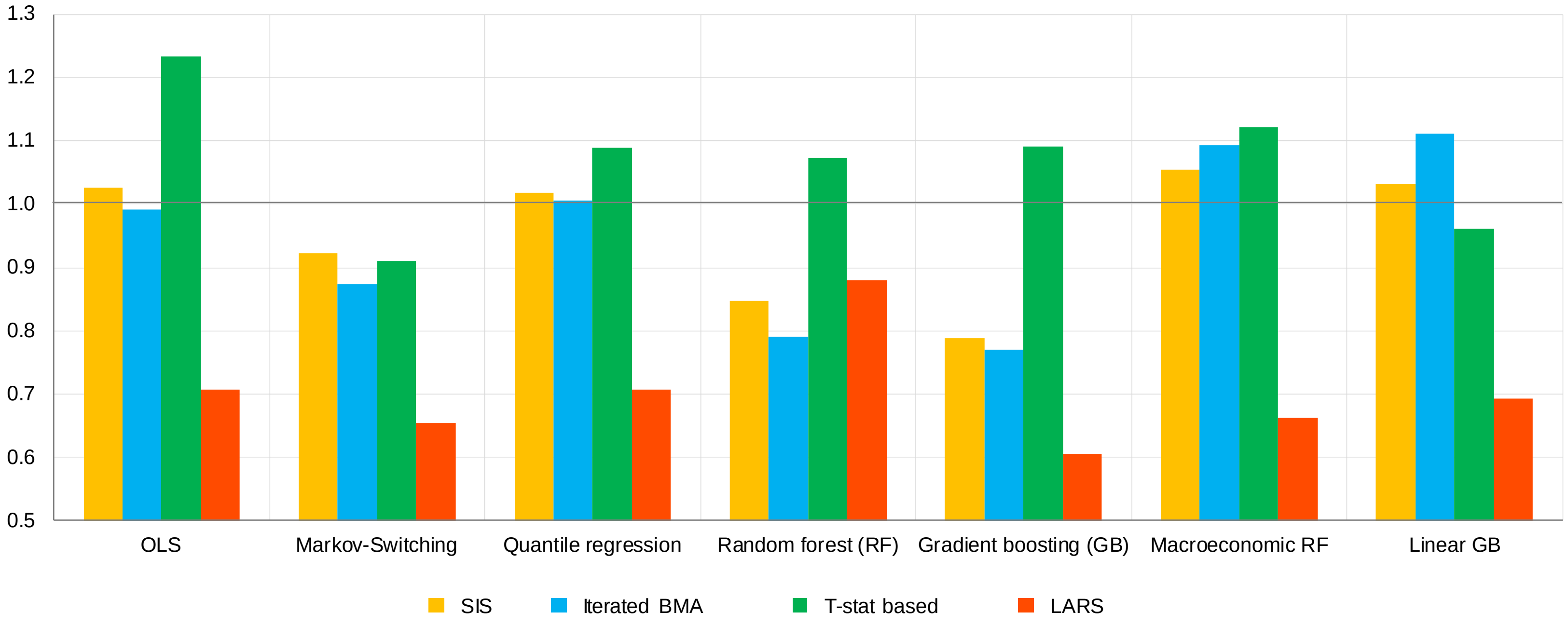


Notes: Accuracy is measured by the out-of-sample RMSE over Jan. 2012 – Apr. 2022. Performances are presented relative to the benchmark of no pre-selection (black dotted line). Results are obtained for the dataset mirroring data available to a forecaster at the 11th day of the month. Factors are obtained through PCA and the regression is performed through OLS.

Pre-selection with alternative regression techniques

Relative accuracy of pre-selection techniques

(RMSFE, relative to the no-preselection benchmark = 1)

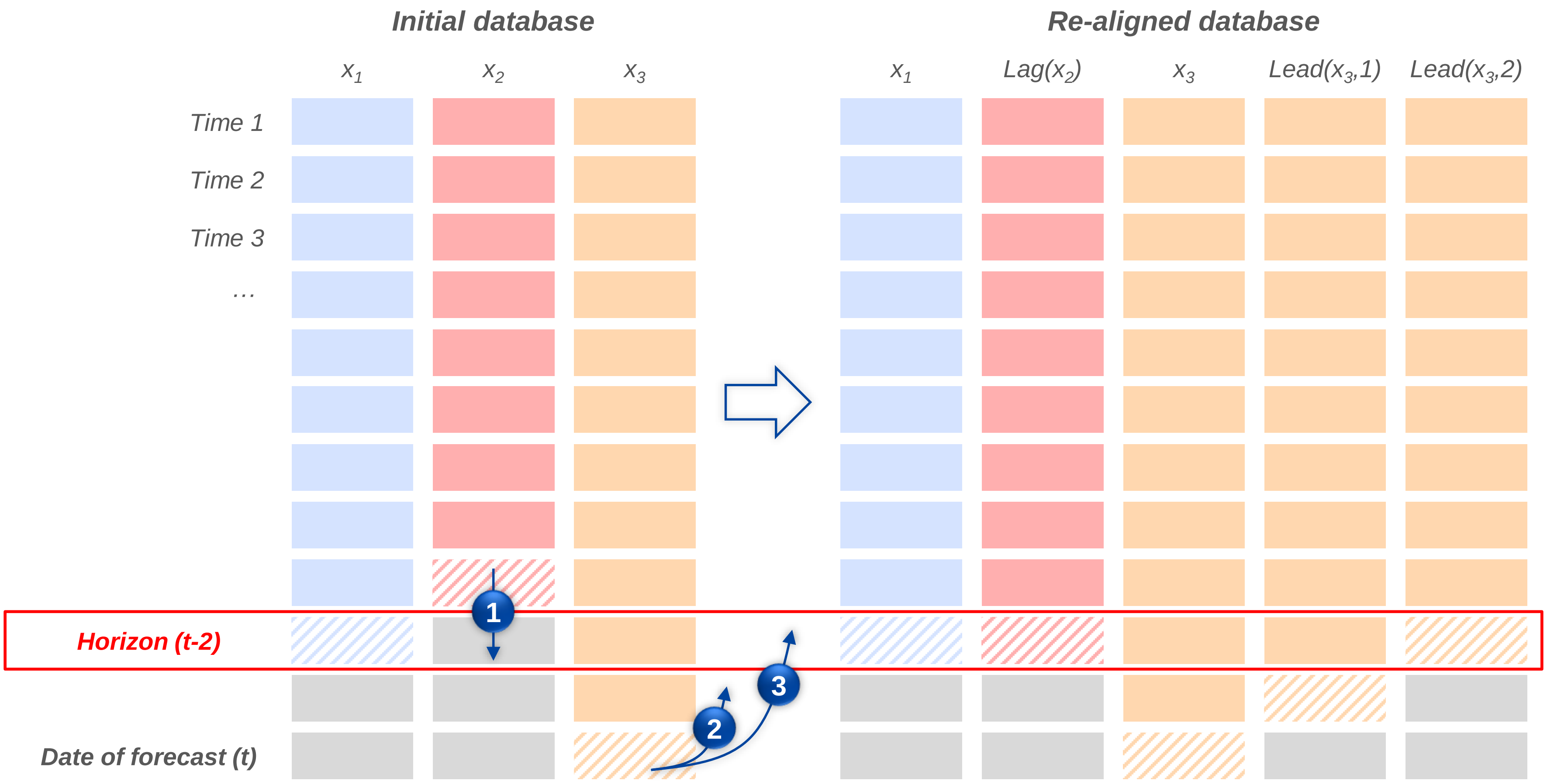


Notes: Accuracy is measured by the out-of-sample RMSE over Jan. 2012 – Apr. 2022. Performances are presented relative to the benchmark of no pre-selection (dark grey line). Results are obtained for the dataset mirroring data available to a forecaster at the 11th day of the month with 60 variables pre-selected by the technique under consideration. Factors are obtained through PCA and the regression is performed through OLS. Results are presented for horizon *t-2.*

LARS algorithm

- Starting with no predictors, add the predictor x_i **most correlated** with the target variable (y)
- Move the coefficient β_i in the direction of its **least-squares estimate**
- The correlation of x_i with the **residual** ($y - \beta_i x_i$) gets lower
- Continue increasing β_i coefficient until another predictor x_j has **similar correlation** with $y - \beta_i x_i$ than x_i
- **Add** x_j to the active set
- Continue by now moving both coefficients β_i and β_j **equiangularly** in the direction of their least-squares estimates, until another predictor x_k has as much correlation with the residual (now $y - \beta_i x_i - \beta_j x_j$)

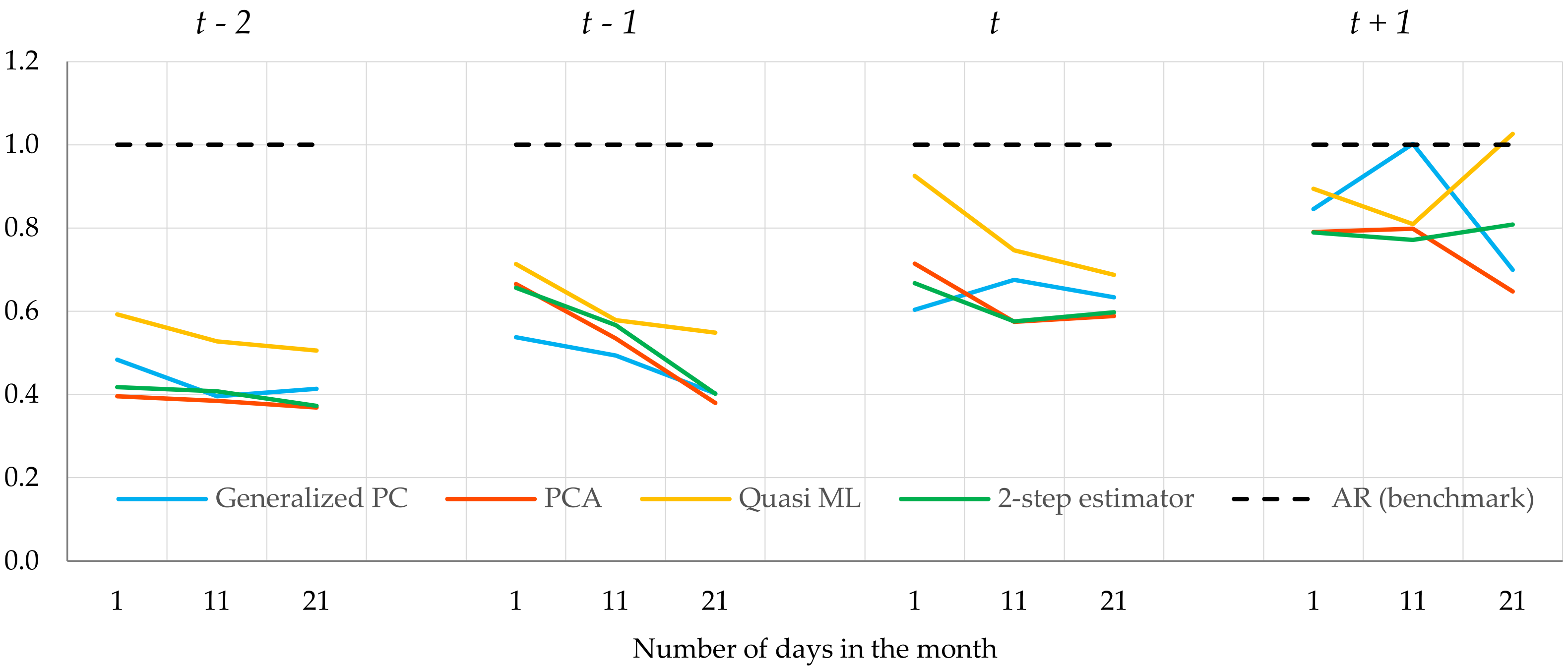
The management of real-time data flow



Results across factor extraction methods

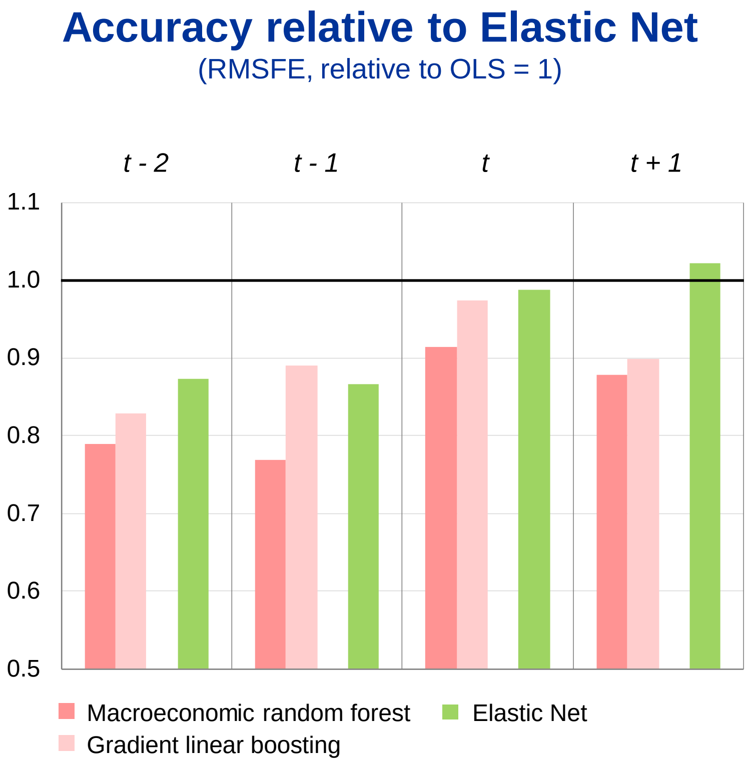
Relative accuracy of factor extraction techniques

(RMSFE, relative to the no factor benchmark = 1)



Notes: Accuracy is measured by the out-of-sample RMSE over Jan. 2012 – Apr. 2022. Performances are presented relative to the benchmark of no factor extraction (black dotted line).

Comparison with elastic net



Notes: Accuracy is measured by the out-of-sample RMSE over Jan. 2012 – Apr. 2022. Performances are presented relative to the OLS (black line). Results are obtained for the average of the datasets mirroring data available to a forecaster at the 1st, 11th, and 21st days of the month, using LARS for pre-selecting the 60 most informative regressors and PCA for factor extraction.

- As one would expect, performances of **gradient linear boosting** overall close to elastic net
- Still gains from **macroeconomic random forest** – notably significant in forecasting

General Gradient Boosting algorithm

- **Initialize** the model $F_0(x) = 0$
- For m going from ***1 to M*** (defined by the user)
    - Compute $(\beta_m, \gamma_m) = \boldsymbol{argmin}_{\beta,\gamma} \sum_{i=1}^{N} L(y_i, F_{m-1}(x_i) + \beta b(x_i, \gamma))$
    - Set the **new model** $F_m(x) = F_{m-1}(x_i) + \epsilon \beta_m b(x, \gamma_m)$
- ϵ is a **shrinkage** parameter that slows the building of the model to prevent the overfitting and generally lead to better predictive performances