

Nowcasting World Trade with Machine Learning: a Three-Step Approach

Menzie Chinn¹, Baptiste Meunier²
and Sebastian Stumpner³

July 2023, WP #917

ABSTRACT

We nowcast world trade using machine learning, distinguishing between *tree*-based methods (random forest, gradient boosting) and their *regression*-based counterparts (macroeconomic random forest, linear gradient boosting). While much less used in the literature, the latter are found to outperform not only the tree-based techniques, but also more “traditional” linear and non-linear techniques (OLS, Markov-switching, quantile regression). They do so significantly and consistently across different horizons and real-time datasets. To further improve performance when forecasting with machine learning, we propose a flexible three-step approach composed of (*step 1*) pre-selection, (*step 2*) factor extraction and (*step 3*) machine learning regression. We find that both pre-selection and factor extraction significantly improve the accuracy of machine-learning-based predictions. This three-step approach also outperforms workhorse benchmarks, such as a PCA-OLS model, an elastic net, or a dynamic factor model. Finally, on top of high accuracy, the approach is flexible and can be extended seamlessly beyond world trade.

Keywords: Forecasting, Big Data, Large Dataset, Factor Model, Pre-Selection

JEL classification: C53, C55, E37

¹ University of Wisconsin. mchinn@lafollette.wisc.edu

² European Central Bank, Aix-Marseille School of Economics (AMSE). baptiste.meunier@ecb.europa.eu

³ Banque de France. Sebastian.stumpner@banque-france.fr

We are very grateful to C. Marsilli, C. Jardet, S. Haincourt, Y. Kalantzis, an anonymous referee, and participants to various internal seminars, the 12th ECB conference on forecasting (June 2023), and the 43rd International Symposium on Forecasting (June 2023) for useful comments. We are indebted to O. Darné and C. Schumacher for sharing their codes on dynamic factor models. We thank F. Lebreton and A. Le Metayer for excellent support regarding the data. The views expressed in this paper are those of the authors, and do not necessarily represent those of the Banque de France, the European Central Bank, the university of Wisconsin, or the AMSE. This work was supported by the French National Research Agency Grant ANR-17-EURE-0020, and by the Excellence Initiative of Aix-Marseille University - A*MIDEX.

NON-TECHNICAL SUMMARY

Real-time economic analysis often faces the fact that indicators are published with significant lags. This problem is encountered for world trade in *volumes*: the earliest indicator is published by the Dutch Centraal Plan Bureau (CPB) roughly eight weeks after month end – meaning that March 2023 data is available around May 25th. Since these data are widely used among economists, this poses a challenge policy-wise as decisions should rely on timely information about the current business cycle. In the meantime, a number of early indicators are available. The purpose of this paper is to exploit such information to get advanced estimates of world trade ahead of the CPB releases.

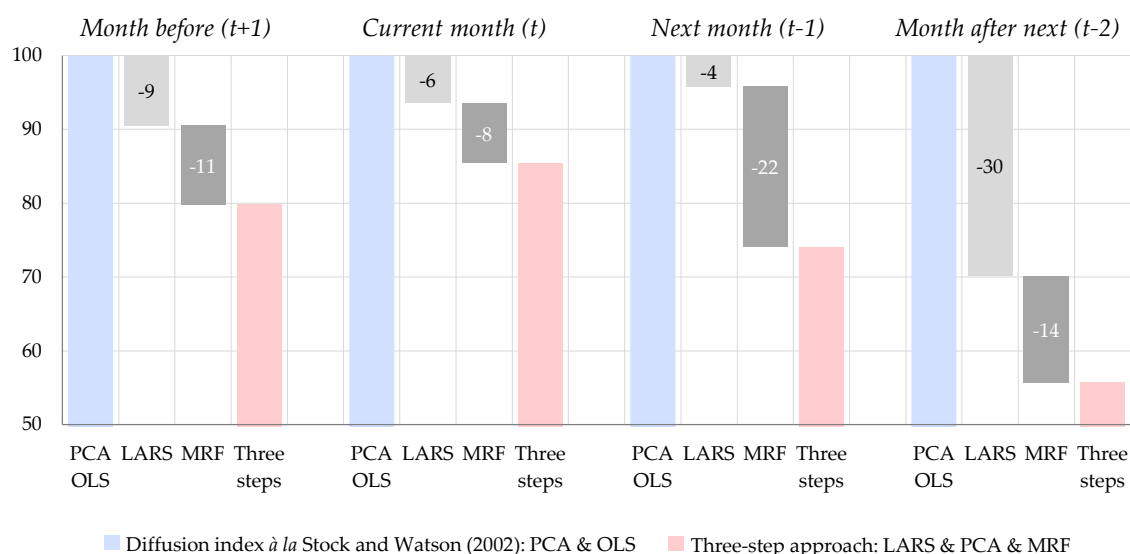
A key novelty of this paper is the use of machine learning techniques for nowcasting. We distinguish between *tree*-based and *regression*-based techniques. The first category – *tree*-based – includes random forest and gradient boosting and is the most popular in the literature. It is however found to perform poorly on our dataset, supporting recent evidence that such techniques might be ill-equipped to deal with the small samples of macroeconomic time series. In contrast, the *regression*-based techniques – macroeconomic random forest and linear gradient boosting – provide the most accurate predictions. They outperform all other techniques, not only *tree*-based machine learning but also more “traditional” non-linear techniques (Markov-switching and quantile regression) and Ordinary Least Squares (OLS). They do so significantly and consistently across different horizons, real-time datasets, and states of the economy.

A second key contribution is to propose a three-step approach for forecasting with machine learning and large datasets. The approach works sequentially: (*step 1*) a pre-selection technique identifies the most informative predictors among our dataset of 600 variables; (*step 2*) selected variables are summarized and orthogonalized into a few factors; and (*step 3*) factors are used as explanatory variables in the regression of world trade, using machine learning techniques. While such pre-selection and factor extraction have been already used in the literature, our contribution is to use them in a combined framework for machine learning. We compare different methods for each step: the best-performing triplet is formed by the Least Angle Regression (LARS, see Efron et al., 2004) for pre-selection, principal component analysis (PCA) for factor extraction, and the Macroeconomic Random Forest (MRF; Goulet-Coulombe, 2020) for prediction. LARS is similar to stepwise regression when dealing with a large set of potential regressors, variables are included step-by-step, but the method ensures that regression coefficients are similar in absolute value when the variables have the same correlation with the residuals.

The three-step approach outperforms benchmarks significantly and consistently. This three-step approach can be viewed as an extension of the widely used “diffusion index” of Stock and Watson (2002) who combine factor extraction by PCA and OLS regression. Compared to a model *à la* Stock and Watson (2002), the three-step approach delivers on average a 26% lower RMSE with accuracy gains coming both from the addition of a pre-selection step and from the use of the macroeconomic random forest (**Figure N1**). We finally check that the three-step approach outperforms workhorse nowcasting models such as a dynamic factor model.

In the end, the three-step approach can be viewed as a step-by-step method for forecasters willing to employ machine learning techniques in order to improve forecast accuracy. Aside from the use of innovative *regression*-based machine learning techniques, the contribution of this paper is the combination of those three steps. We show that each step improves accuracy: alternative approaches that excludes either pre-selection, factor extraction, or machine learning are found to underperform. Such findings contribute to the growing literature on machine learning by showing empirically that: (*i*) on short samples, machine learning techniques work best if data is summarized into factors instead of taking all of the individual series as explanatory variables, and (*ii*) accuracy is even greater if only a subset of the potential regressors is pre-selected.

Figure N1. Decomposition of accuracy gains relative to PCA-OLS



Notes: "PCA OLS": Diffusion index following Stock and Watson (2002); "LARS": pre-selection with Least Angle Regression; "ML": machine learning with macroeconomic random forest; "3-step": final accuracy with the three-step approach using: (1) LARS for pre-selecting the 60 most informative regressors, (2) factor extraction through PCA, and (3) regression with macroeconomic random forest (Goulet-Coulombe, 2020). Results are presented in terms of relative accuracy to PCA-OLS normalized to 100 for each month. Results are average gains over datasets at the 1st, 11th, and 21st days of the month.

Prévoir le commerce mondial en temps réel grâce au *machine learning* : une approche en trois étapes

RÉSUMÉ

Nous prévoyons le commerce mondial à l'aide des méthodes de *machine learning*, en distinguant entre celles fondées sur les *arbres de décision* (forêt aléatoire, *gradient boosting*) et leurs équivalents basés sur les *régressions* (forêt aléatoire macroéconomique, *gradient boosting* linéaire). Bien que moins utilisées dans la littérature, ces dernières s'avèrent plus performantes que les techniques basées sur les arbres, ainsi que d'autres techniques linéaires et non linéaires (MCO, *Markov-switching*, régression quantile). Leur performance est significative et constante à différents horizons. Nous proposons une approche flexible en trois étapes : (*étape 1*) pré-sélection des variables, (*étape 2*) extraction de facteurs et (*étape 3*) régression *machine learning*. Nous constatons que la pré-sélection et l'extraction de facteurs améliorent de manière significative la précision des prévisions de *machine learning*. Cette approche en trois étapes est également plus performante que les modèles de référence, tels qu'un modèle à facteurs dynamiques. En plus d'une précision accrue, l'approche est flexible et peut être appliquée de même manière au-delà du commerce mondial.

Mots-clés : prévision, modèles à facteurs, pré-sélection, larges bases de données

Les Documents de travail reflètent les idées personnelles de leurs auteurs et n'expriment pas nécessairement la position de la Banque de France. Ils sont disponibles sur publications.banque-france.fr

Introduction¹

Real-time economic analysis often faces the issue that indicators are published with significant lags. This problem is encountered for world trade: while information on trade in *values* is available with little delay, trade in *volumes* tends to be much less timely. At monthly frequency, the Dutch Centraal Plan Bureau (CPB) issues estimates for global trade in *volumes* which are widely used among economists, but which are published around eight weeks after month end – meaning Sept. 2022 data would be available around November 25th. In the meantime, a number of trade indicators are available, providing signals regarding the current stance of global trade.

The purpose of the paper is to exploit such early available information to provide advance estimates of trade in *volumes* ahead of the CPB releases. To this end, we assemble a large dataset of 600 variables based on the literature on nowcasting trade. Given publication delays for the CPB data, the purpose is not only to predict trade for the current month (“nowcasting”: prediction for month t at which the forecaster is) but also in previous months (“back-casting” at months $t - 2$ and $t - 1$ for which CPB data have not released yet). We also “forecast” at $t + 1$ to assess the informative content of our dataset about future developments.

Armed with such a large dataset, we explore the use of innovative machine learning techniques. One key interest of such methods lies in their ability to handle non-linearities, that can be central for trade forecasting given the prominence of crisis in the last decades and the inherently high volatility of trade (Bussiere et al., 2013). As we test over a range of different techniques, an important ingredient in our study is the distinction between machine learning techniques based on *trees* and those based on *regressions*. The first category (random forest, gradient boosting) is the most widely used in the literature and works by aggregating several *decision trees* together. The second category (macroeconomic random forest, linear gradient

¹ Another non-technical article on this work can be found as *Banque de France bulletin* (Chinn et al., 2023). Please note that the R code for running a simplified version of the three-step approach can be found at https://github.com/baptiste-meunier/NowcastingML_3step.

boosting) is much less used in the literature and is an adaptation of the first category but using *linear regressions* instead of, or in complement to, decision trees.²

Second, we propose a three-step approach composed of pre-selection, factor extraction, and machine learning regression. This framework aims at maximizing the accuracy of the machine-learning-based predictions. It is motivated by the literature: for instance, Goulet-Coulombe et al. (2022) suggest that machine learning techniques are more accurate when used in a factor model rather than when applied directly on all individual series. Doing pre-selection ahead of factor extraction responds to another literature (Bai and Ng, 2008) that found that selecting fewer but more informative regressors improve performances of factor models. Our framework combines and extends these two strands of the literature. In addition, we test a large number of different methods for pre-selection and factor extraction in order to assess the best-performing combination.

As regards machine learning techniques, we find that those based on *regressions* outperform other techniques significantly and consistently. They outperform in particular the *tree*-based methods which – despite their increasing popularity in the literature – are found to perform poorly in our setup. This supports recent evidence that such techniques might be ill-equipped to deal with the small samples of macroeconomic time series. More broadly, we find that *regression*-based machine learning techniques also outperform more “traditional” linear (OLS) and non-linear (Markov-switching, quantile regression) techniques, again significantly and consistently across different horizons, real-time datasets, and states of the economy. Individually, the best-performing method is found to be the macroeconomic random forest of Goulet-Coulombe (2020), an extension of the canonical random forest. Compared to the OLS benchmark, such a technique allows for significant accuracy gains, of a magnitude about 15-20% on average.

We also find empirically that the three-step approach significantly outperforms other workhorse nowcasting techniques. It outperforms in particular the widely used “diffusion index” method of Stock and Watson (2002) that uses two steps: factors extraction *via* Principal Components Analysis (PCA) and OLS regression on these factors. Our approach also

² It should be noted that despite the use of linear regressions in the framework, these machine learning techniques (linear gradient boosting, macroeconomic random forest) remain highly non-linear.

outperforms a dynamic factor model, a technique widely used in the nowcasting literature. We also show that both pre-selection and factor extraction improve the accuracy of machine learning techniques. For instance, adding a pre-selection enhances predictive accuracy by around 10-15%. This suggests that pre-selection can be instrumental also for machine learning techniques – despite the idea that such techniques can handle smoothly large datasets with irrelevant variables. We finally find that, for machine learning techniques, predictions using a few factors are more accurate than predictions using all individual variables as regressors, with accuracy gains also in the range of 10-15%.

This paper contributes to the nowcasting literature, in particular to the growing strand forecasting with machine learning. For trade, there have been relatively few efforts as most of the literature on nowcasting trade relied on dynamic factor models (Guichard and Rusticelli, 2011; Jakaitiene and Dees, 2012; Bahroumi et al., 2016; Martinez-Martin and Rusticelli, 2021). Our paper goes in line with recent efforts towards machine learning, such as Hopp (2021) using neural networks. In the more general field of nowcasting, our paper brings a new distinction between machine learning techniques based on *trees* and those based on *regressions*. A key contribution of our paper is to show that the techniques that are the most widely used in the literature (*tree*-based) have mixed performances, while the innovative *regression*-based techniques strongly outperform their competitors.

Another contribution to the literature is to lay out a practical step-by-step approach for forecasting with machine learning and a large dataset. Our findings suggest that pre-selecting fewer regressors and summarizing the regressors into a few factors improve the accuracy of machine-learning-based predictions. This contributes to the fast-growing literature on machine learning forecasting by showing that accuracy is improved (*i*) when summarizing data into a few factors – instead of feeding all individual series as explanatory variables, and (*ii*) if only a subset of the large dataset is kept. These findings add to the practitioners' guide on machine learning forecasting and extends similar efforts in the literature, most notably Goulet-Coulombe et al. (2022).

The rest of the paper is organized as follows: **section 1** describes the three-step approach and the different techniques tested; **section 2** details how the nowcasting is performed in real-time.

Section 3 provides the main results. **Section 4** highlights the benefits of each of the three steps of the approach. The final section concludes.

Section 1: A three-step approach for back-, now-, and fore-casting

1. Overview

Figure 1 illustrates our three-step approach aimed at maximizing flexibility. Starting from a large dataset, the first step consists in selecting a few regressors with the highest predictive ability. The pre-selected dataset is then summarized in fewer factors, which are then used in non-linear regression models. Our approach follows and complements different strands of the nowcasting literature. A first strand combines pre-selection with factor models but remains with a linear regression set-up – e.g. the LARS-DFM of Runstler (2016) or the FA-MIDAS of Marcellino and Schumacher (2010). A second strand combines factor extraction and regression: the baseline framework of Stock and Watson (2002) uses an OLS regression upon factor extraction, but it has been adapted to non-linear techniques, and even machine learning techniques (Goulet-Coulombe et al., 2022). Building on this literature, our approach combines the different steps into an integrated set-up, and for each of the steps, tests a wide range of methods – notably innovative machine learning techniques. The different methods tested are laid out in the bottom of **Figure 1**. Although we test numerous methods for each of the steps, it should be noted that our goal is to select only the best-performing triplet consisting of one pre-selection technique, one factor extraction method, and one regression model.³

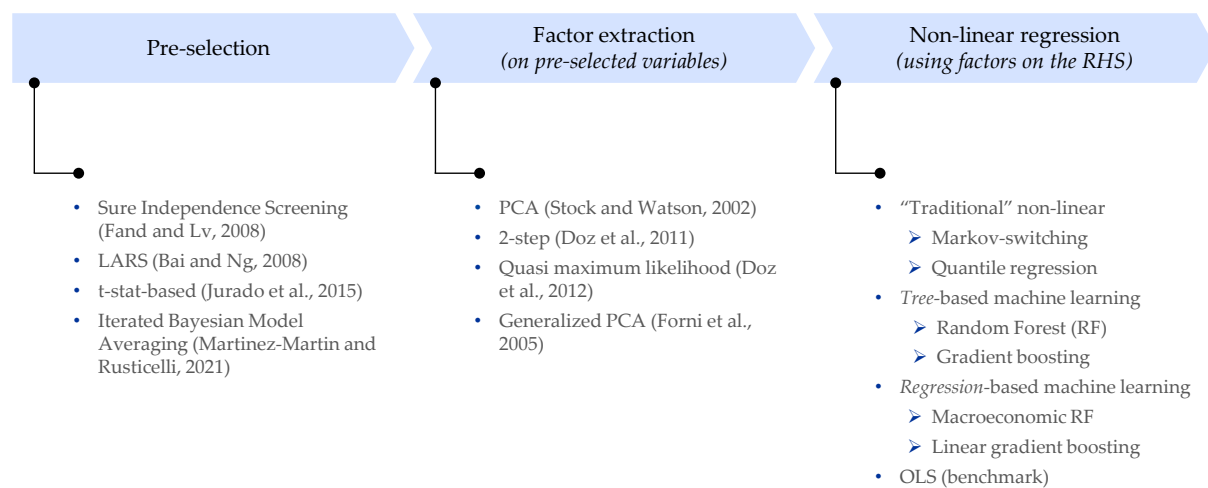
The approach is straightforward to use and highly flexible. It can be applied seamlessly to any dataset and, once operational, withstand data changes (e.g. inclusion of new data such as innovative datasets which emerged during the Covid-19 crisis,⁴ changes in release dates or other data settings). The separation between steps makes it easy to test a wide range of

³ We have also tested whether combining forecasts from different models would help – which would be in the spirit of adding a fourth step (model combination). We find however that an equally weighted pool of predictions *à la* Amisano and Geweke (2017) does not improve forecast accuracy. In addition, our suggested approach relying on a single model (the best-performing one) has the double advantage of lower computational time and simplicity.

⁴ Examples can include real-time marine traffic (Cerdeiro et al., 2020), credit card data (Carvalho et al., 2020), web-scraped housing listings (Bricongne et al., 2023), electricity consumption (Chen et al., 2020), or satellite data (Bricongne et al., 2021).

different techniques and select the ones tailored for each exercise. Another key advantage is that pre-selection is a data-driven step made automatically. Said otherwise, it lifts the burden of selecting variables from the forecaster, since data-driven pre-selection does not require any *a priori* knowledge. Instead of having to select variables by hand, the forecaster can provide the feed the full dataset into the framework – leaving this nitty-gritty task to the algorithm. A contribution of this approach is also to make the selection of variables explicit – while most of the literature does not provide any detail on this step, conducted *ex ante* and based on the own experience of the authors.

Figure 1. Summary of three-step approach



Beyond flexibility, the framework is also aimed at enhancing accuracy. Pre-selection has been shown in the literature to significantly enhance the performances when nowcasting with factor models (Bai and Ng, 2008; Runstler, 2016; Jardet and Meunier, 2022). Factor extraction has been shown to be a potent way to summarize an extensive amount of information in order to achieve parsimony and expel idiosyncratic noise from the data, ultimately leading to better performances in nowcasting (Stock and Watson, 2002; Bai and Ng, 2007). In addition, factor extraction produces orthogonal variables, thereby alleviating collinearity and enhancing the accuracy. Finally, non-linear techniques – including innovative machine learning techniques – have been used to improve accuracy relative to their linear counterparts (Goulet-Coulombe et al., 2022), particularly during crisis episodes or for volatile variables.

2. *Data*

Variables included in our dataset cover broad aspects of the trade outlook. Our target variable is the year-on-year growth rate of world trade from the CPB. Our set of explanatory variables is composed of 536 variables detailed in **Annex 2**. To build this dataset, we have taken all variables included at some point in the literature on nowcasting trade notably Keck et al. (2010), Guichard and Rusticelli (2011), Jakaitiene and Dees (2012), Stratford (2013), Bahroumi et al. (2016), d'Agostino et al. (2017), Martinez-Martin and Rusticelli (2021), Charles and Darné (2022). It includes early indicators for trade, e.g. trade values from various countries, shipping costs, freight volumes in several ports and trade routes, PMI “new export orders” for both manufacturing and services, or truck traffic. Variables for the broader macroeconomic outlook are also included, with the aim of covering both industrial activity (e.g. steel production, industrial production) and households’ consumption (e.g. retail sales, car registrations). Finally, commodity prices are included (oil and other non-energy prices) as well as financial indicators (e.g. S&P 1200 global) given the capacity of the later to act as a catch-all proxy for other developments in the outlook (Hasenzagl et al., 2020).

3. *Pre-selection techniques*

When forecasting with a high-dimensional dataset, the literature generally concludes that the factor models are significantly more accurate when selecting fewer but more informative predictors (Bai and Ng, 2008). On a more theoretical ground, Boivin and Ng (2006) find that larger datasets lead to poorer forecasting performances when idiosyncratic errors are cross-correlated or when the variables with higher predictive power are dominated.

Against this background, our first step consists in pre-selecting the regressors with the highest predictive power. Formally, the initial dataset is $X_t = (x_{1,t}, x_{2,t}, \dots, x_{N,t})'$ with $t = 1, \dots, T$ ($T = 271$) and N variables ($N = 536$). The idea underlying pre-selection is to rank regressors $x_{i,t}$ based on a measure of their predictive power with respect to the target variable. We consider four techniques from the literature:

- The “Sure Independence Screening” (SIS) of Fan and Lv (2008): regressors are ranked based on their marginal correlation with the target predictor. Fan and Lv (2008) provide theoretical ground for their approach by demonstrating that it has the sure screening property that “*all important variables survive after applying a variable screening procedure with probability tending to 1*”. This approach has been used for nowcasting in Ferrara and Simoni (2019) or Proietti and Giovannelli (2021).
- T-stat-based: each regressor $x_{i,t}$ is ranked based on the absolute value of the t-statistic associated with its coefficient estimates in a univariate regression of $x_{i,t}$ on the target variable y_t . The univariate regression also includes four lags of the dependent variable to control for endogenous dynamics. While originating in genetic studies (Bair et al., 2006), this technique has found its way to economics for example in Jurado et al. (2015).
- Least-Angle Regression (LARS) as in Bai and Ng (2008): while the two methods above are based on univariate relationships of regressors with the target variable, this one accounts for the presence of the other predictors. The LARS (Efron et al., 2004) is an iterative forward selection algorithm. Starting with no predictors, it adds the predictor x_i most correlated with the target variable y and then move the coefficient β_i in the direction of its least-squares estimate so that the correlation of x_i with the residual ($y - \beta_i x_i$) gets lower. It does so until another predictor x_j has similar correlation with $y - \beta_i x_i$ than x_i . At this point, x_j is added to the active set and the procedure continues with 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 $y - \beta_i x_i - \beta_j x_j$). This approach has been used in nowcasting notably in Schumacher (2010), Bulligan et al. (2015) or Falagiardia and Sousa (2015).
- Iterated Bayesian Moving Averaging (BMA) which also accounts for the presence of other regressors. This technique works by making repeated calls to a BMA procedure (Raftery, 1995). BMA applies a Bayesian framework on all possible models M_i using the set of variables; the Bayes rule then allows to compute the posterior mode probability for each model $p(M_i|\Omega_t)$ with Ω_t the data available at time t , according to

equation 1.⁵ The BMA returns the set of models whose posterior model probability is the highest.⁶ Beyond model selection, BMA allows to also compute posterior inclusion probability for each variable – based on the posterior model probabilities of models in which this variable is included. When used for pre-selection, BMA runs iteratively through regressors by groups of n and following a pre-determined pecking order.⁷ Starting with the first n regressors in pecking order, the BMA determines the posterior inclusion probability of each regressor. Those with probabilities higher than a threshold are kept while the others are replaced by the next regressors in the pecking order. The BMA is then run on this new batch of regressors, and so on and so forth until all regressors have been assessed. Initially developed for gene selection by Yeung et al. (2005), this approach has been used for nowcasting – notably trade in Martinez-Martin and Rusticelli (2021).

$$(1) \quad p(M_i|\Omega_t) = \frac{p(\Omega_t|M_i) \cdot p(M_i)}{p(\Omega_t)} = \frac{p(\Omega_t|M_i) \cdot p(M_i)}{\sum_k p(\Omega_t|M_k) \cdot p(M_k)}$$

4. Factor extraction

The econometric framework relies on a factor model. Formally, we assume that the pre-selected dataset X_t can be represented by a factor structure with a r -dimensional factor vector F_t , a loadings matrix Λ and an idiosyncratic component ξ_t unexplained by the common factors

⁵ Computing the *posterior* model probability therefore requires assigning a *prior* model probability $p(M_i)$ and calculating, for each data, the probability to have the data available Ω_t given parameters of the model M_k $p(\Omega_t|M_k)$. In general, prior model probability are set all equal.

⁶ Used for regression, BMA accounts for model uncertainty by averaging the predictions across selected models. In practice, predictions are obtained by averaging over the selected models, weighting individual predictions of each model by their posterior model probability. On top of accounting for model uncertainty, it can also be shown that BMA yields optimal predictions under the squared error loss function (Hoeting et al., 1999). In a real-time set-up, BMA also generally induce smoother updates in regression estimates when data changes.

⁷ The reason to call iterative BMA on groups of n variables (with $n < N$) is that BMA has to test all possible models based on the group of n variables. For instance, there exists 2^n possible models with 2 variables. If n is too large, the procedure becomes un-tractable computationally. Hence the calls to iterated BMA. As per the pre-determined pecking order, it is also necessary to ensure that the algorithm sees first variables with the *a priori* highest informative power. In practice, this order is based on the R^2 of a univariate regression of the regressor on the target variable.

– as in equation 2.⁸ The common ($\Lambda \cdot F_t$) and idiosyncratic components ξ_t are assumed to be mutually orthogonal.

$$(2) \quad X_t = \Lambda \cdot F_t + \xi_t$$

Following the canonical Stock and Watson (2002) framework, static factors are extracted *via* Principal Components Analysis (PCA). PCA assumes that F_t and ξ_t are independent and identically distributed (i.i.d.). Factors can be estimated through maximum likelihood and are consistent estimators as long as factors are pervasive and the idiosyncratic dependence and cross-correlation in ξ_t is weak. However, if common factors can no longer be assumed to be i.i.d. – most notably when they are serially dependent – the PCA might not be the most efficient factor extraction method as it will ignore this serial dependence. For this reason, alternative techniques for factor extraction are evaluated in **Annex 3** – it shows that accuracy is relatively similar across different techniques. Therefore, we elect PCA which has the double advantage of simplicity and lower computational need.

5. Regression techniques

Following the diffusion index method of Stock and Watson (2002), factors enter as explanatory variables for world trade. We produce direct forecast for different horizons h – which can be null (nowcasting) or take negative (back-casting) or positive (forecasting) values – following equation 3. Important to note is that (i) t is the current date for the forecaster but not the last available observation of y , and (ii) the set-up differs if the horizon is strictly positive.

$$(3) \quad \begin{cases} y_{t+h} = f(F_{t+h}) + \varepsilon_t & \text{if } h \leq 0 \\ y_{t+h} = f(F_t) + \varepsilon_t & \text{if } h > 0 \end{cases}$$

The main purpose of the paper is to test several functions f , comparing performances of the standard OLS technique with non-linear approaches – notably machine learning techniques.

⁸ The number of factors r is determined through the Bai and Ng (2002) information criteria, informed by Kaiser (1960)'s criterion which sets the higher admissible value r^{up} for the number of factors. Bai and Ng (2002) criterion then elects a number of factors in the interval $[0, r^{up}]$.

The focus on non-linear techniques is motivated both by the high volatility of trade and by the recent advances in econometrics.

The first category we explore are “traditional” non-linear techniques:

- Markov-switching (MS) that allows model parameters to differ across regimes.⁹ MS assumes that unobserved states are determined by a Markov-chain. The framework is characterized by transition probabilities describing the likelihood to stay in the same regime or to switch to another. Model parameters are estimated using maximum likelihood as in Engel and Hamilton (1990) based on expectation-maximization. In the first step, the path of the unobserved variable (latent variable) is estimated. In the second, given the unobserved regime estimated in first step, model parameters and transition probabilities are estimated. Both steps are iterated until convergence. Owing to its capacity to estimate the state of the business cycle, MS has been widely used in nowcasting, for example by Boot and Pick (2014) and Carstensen et al. (2020).
- The quantile regression (QR; Koenker and Bassett, 1978) in which the non-linearity comes from the fact that it estimates conditional quantiles of interest of the dependent variable. The framework differs from the OLS in two main ways: (i) coefficients $\beta_i(\tau)$ depend on the quantile τ , and (ii) rather than the sum of squared residuals $\sum(y_i - \beta_i \cdot x_i)^2$ as in OLS, the QR minimizes the expression 4 below where the check function ρ_τ gives asymmetric weights to the error depending on the quantile τ and the sign of the error. This is an extension of OLS which can be used when some conditions of the linear regression are not met (e.g. homoscedasticity, independence, normality). This method is notably employed in the growth-at-risk framework (Adrian et al., 2019) with applications to nowcasting (Mazzi and Mitchell, 2020).

$$(4) \quad \sum_i (y_i - \beta_i(\tau) \cdot x_i)^2, \quad \text{with } \rho_\tau(u) = \tau \cdot \max(u, 0) + (1 - \tau) \cdot \max(-u, 0)$$

⁹ In our applications, we set the number of regimes to two following the general practice (Hamilton, 1989).

Beyond these, we also test innovative machine learning methods. We first turn to techniques based on *trees* which are the most widely used in the literature:

- Random forest (RF; Breiman, 2001) is an ensemble method using a large number of decision trees. The underlying idea is to build a large number of un-correlated trees. Then, by averaging predictions over multiple noisy trees, the variance of the aggregate prediction is reduced. And since trees can also have relatively low bias, the aggregate prediction can exhibit both low variance and low bias. Key in this technique is the low correlation among trees: this is ensured by (i) growing each tree on a bootstrapped subsample of the initial dataset,¹⁰ and (ii) by restricting the number of variables considered at each node – only a random subset of variables is allowed, forcing even lower correlation amongst trees. Such a technique is increasingly used in nowcasting (Soybilgen and Yazgan, 2021; Medeiros et al., 2021).
- Gradient Boosting (GB-T; Friedman, 2001)¹¹ is another class of tree-based method which uses a combination of these un-correlated weak learners. But contrary to random forest which averages multiple trees, GB-T works by adding a tree at each iteration. More specifically, the tree is added following the direction that minimizes the loss from the prior model (i.e. following the gradient). A pre-determined number of trees are added, or the algorithm stops once the loss falls below a threshold or no longer improves. Overfitting is alleviated by constraining trees and on stochastic gradient descent in which, at each iteration, only a subsample is used. While GB-T remain less known than RF, the usage is nevertheless developing in nowcasting (e.g. Yoon (2021) for US GDP).

In the machine learning realm, we distinguish between the tree-based models above and their extensions which are rather based on (*linear*) *regressions*:

- For the random forest framework, we use the “Macroeconomic Random Forest” (MRF) of Goulet-Coulombe (2020). The MRF exploits the idea that canonical random forests

¹⁰ This method (bagging, for bootstrap aggregation) suits decisions trees since these are highly sensitive to the in-sample data, meaning that small changes on the estimation sample can result in significantly different trees.

¹¹ More specifically, we use the “Extreme Gradient Boosting” (XGBoost) method developed by Chen and Guestrin (2016) which has the advantages of being faster than other gradient boosting algorithms – notably by resorting to parallel processing.

are too flexible and therefore might be inefficient for macroeconomic time series with a limited number of observations. Instead of applying trees on the full sample – as in a random forest – the MRF sets a linear regression $y_t = X_t\beta_t$ where y_t is the target variable, X_t the vector of explanatory variables, and β_t the associated coefficients. But unlike in linear regression, coefficients β_t of the linear part can vary through time according to a random forest. Formally, $\beta_t = \mathcal{F}(S_t)$ where \mathcal{F} refers to the random forest algorithm and is based on S_t , a set of variables potentially different from X_t . This can be viewed as a way to discipline the flexibility of the random forest by ensuring some linearity in the model. This adaptation combining random forest with linear regression can also be interpreted as a “generalized time-varying parameters”.¹²

- On the gradient boosting, we use the “linear gradient boosting” (GB-L) version. The framework is the same as above for the GB-T, but a linear regression is used as the basic weak learner instead of a decision tree. To prevent over-fitting – that could arise more quickly with a linear regression than with a decision tree – the algorithm can include L1 and L2 regularizations.¹³

Section 2: The real-time set-up

1. The management of real-time data flow

In real-time, asynchronous publication dates across the different variables lead to a “ragged-edge” pattern at the bottom of the dataset (see the left-hand side of **Figure 2**). To address this issue, we apply the “vertical realignment” technique of Altissimo et al. (2006) to variables that

¹² The “generalized” comes from the fact that no law of motion (random walk, Markov process) has to be assumed *a priori* by the forecaster for the time-varying parameters.

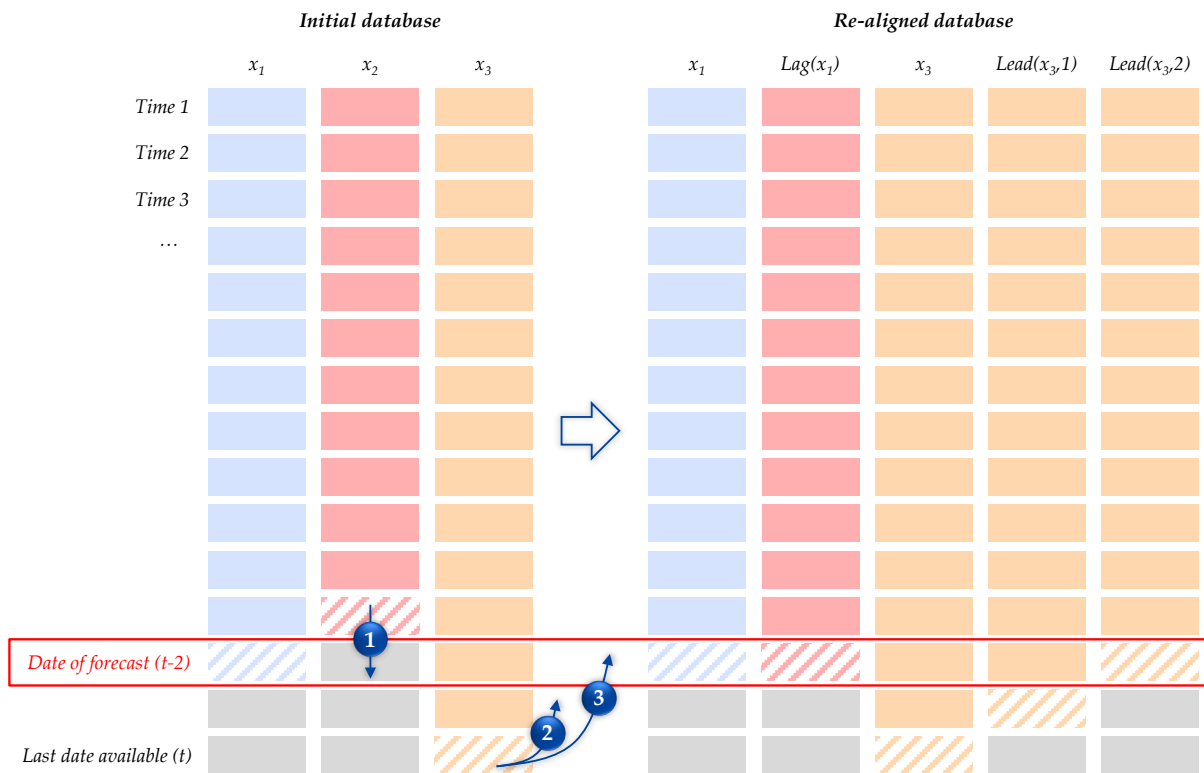
¹³ L1 and L2 regularizations have been popularized in economics by penalized regressions, such as the LASSO (Tibshirani, 1996). Instead of minimizing the sum of squared residuals $\sum (y_i - \beta_i \cdot x_i)^2$ as in an OLS, coefficients minimize the sum of squared residuals with a penalty which is the L1 norm – in the LASSO regression – or the L2 norm – in the RIDGE regression. It can also be a combination of both in the Elastic Net as in expression 11.1 below. The advantages over OLS are that penalized regressions ensure parsimony and avoid over-fitting. An application of these methods for forecasting trade can be found in Charles and Darne (2022).

$$(11.1) \quad \sum_i (y_i - \beta_i \cdot x_i)^2 + \lambda \cdot [\alpha \cdot \|\beta\|_2^2 + (1 - \alpha) \cdot \|\beta\|_1]$$

do not have values at the intended date of the forecast (e.g. x_2 below). For each variable, the last available point is taken as the contemporaneous value and the entire series is realigned accordingly. Formally, for a series x_t whose last observation at time T is for $T - k$, the realigned series is $\tilde{x}_t = x_{t-k}$. This straightforward procedure has been used in various nowcasting applications (Ferrara and Marsilli, 2019; Jardet and Meunier, 2022) and has been shown in Marcellino and Schumacher (2010) to perform as well as other techniques.¹⁴ In addition to the baseline vertical realignment of Altissimo et al. (2006), we adjust for variables with observations available after the intended date of the forecast – e.g. x_3 below, a situation which can arise due to the long publication delay of our target variable. We apply to these variables a symmetric approach to vertical realignment in which we consider *leads* instead of *lags*. But unlike in Altissimo et al. (2006) where the re-aligned series replace the old ones, here new series are created to avoid removing contemporaneous correlations. More precisely, if a series has k values after the date of the forecast, then k new series are created: the first new series is $\tilde{x}_t^1 = x_{t+1}$, the second is $\tilde{x}_t^2 = x_{t+2}$, and so on. The key advantage is to avoid losing the “excess” information – i.e. information related to after the date of the forecast – which can still contain valuable predictive power.

¹⁴ While some issues may arise with this method – most notably that the availability of data determines cross-correlation between variables and then can change over time – Marcellino and Schumacher (2010) empirically test alternative methods (EM-algorithm of Stock and Watson (2002) and the Kalman smoother of Doz et al. (2012)) and find no substantial changes on nowcasting performances across the different methods.

Figure 2. Re-alignment of dataset



Notes: Coloured cells represent available observations, grey cells missing values. The stripped cell account for the last available observation.

The procedure is graphically illustrated in **Figure 2**. For a given date of prediction (in red), no change is applied for series whose last available observation coincides with this date – as is the case for x_1 (in blue). Vertical realignment *à la* Altissimo et al. (2006) applies to series with missing values at the date of the prediction (x_2 – in red): the lagged series replaces the old one. For series with observations after the date of the forecast (x_3 – in orange), new series are created which take incremental lead of the old series. In this example with two leads, two new series (respectively $lead(x_3)$ and $lead(x_3, 2)$) are created. We check empirically the value-added of this method *vs.* the plain Altissimo et al. (2006) – where such observations after the date of forecast would be discarded – by computing the accuracy over Jan. 2012 to Apr. 2022 with both methods. Accuracy is very similar across methods – notably at longer horizons where there can be no leading data – but is marginally higher, by 3 to 7%, for short horizons with the method suggested in this paper.

2. *Pseudo real-time set-up*

The comparison across techniques is run out-of-sample on post-GFC trade, from January 2012 to April 2022 in a close-to-real-time set-up, as pre-selection, factor extraction and model parameters are re-estimated at each point. This aims at mimicking what a forecaster would have been capable to achieve with the information at his disposal at the time T of the forecast.¹⁵ Hence, pre-selection is performed only with the in-sample data. It means that variables in the pre-selected set X_t^T can change from T_1 to T_2 . Factors F_t^T can also change from T_1 to T_2 , not only because the pre-selected set can differ but also because of the incorporation of new data – i.e. factors at T_1 are estimated with observations up to T_1 , and factors at T_2 with data up to T_2 .¹⁶ Finally, the models are estimated in-sample. We then produce out-of-sample predictions at four horizons: two back-casts at $T - 2$ and $T - 1$, a nowcast at T , and a forecast at $T + 1$. These horizons follow the publication lags of CPB trade data, which are released around 2 months after month end (e.g. data for September 2022 is available around the 25th November 2022).

It should be noted that for machine learning methods, such a real-time set-up entails also the re-calibration of hyper-parameters – i.e. parameters that are not estimated by the model, but instead set by the forecaster, e.g. the number of trees in a random forest – at each date T . To do so, we perform a cross-validation adapted for time series: the in-sample period is split between a “test” sample (the last 12 monthly observations) and a “train” sample (the rest of the data). We then train models with different hyperparameters on the train sample to predict observations of the test sample. At the end, we select the set of parameters that minimizes the RMSE on the test sample.¹⁷ All-in-all, our pseudo-real-time set-up follows the procedure shown in **Figure 3**. To further mimic the real-time set-up and explore the consistency of our

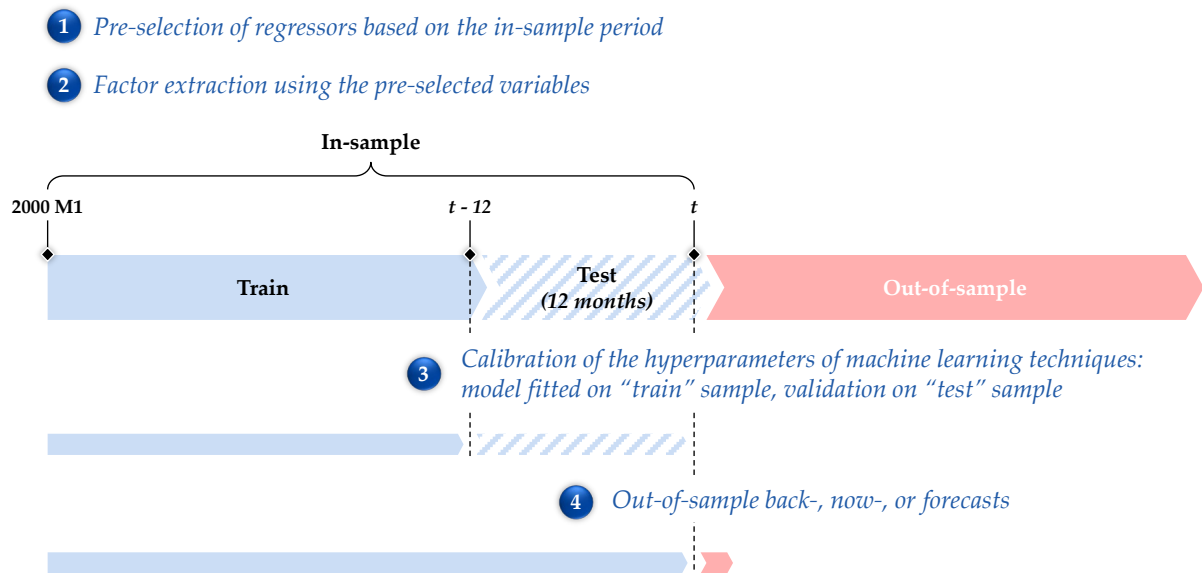
¹⁵ One caveat is that real-time vintages of the dataset were not available. Data were extracted in July 2022 and therefore incorporate all revisions between their publication and this date. While this means our set-up is rather “pseudo” real-time, it should be noted that our dataset is mostly composed of trade statistics which generally undergo little revisions after publication.

¹⁶ This is in contrast with the literature generally proceeding to pre-selection and factor extraction only once on the full sample – therefore incorporating information over the full sample and then being less close than us to what could have been achieved in real-time.

¹⁷ In practice, the number of hyper-parameters depend on the method: number of variables allowed when growing a tree and minimum size of terminal node for Random Forest (2); number of iterations, learning rate, maximum tree depth, minimum child weight, and minimum loss reduction required to make a further partition for Gradient Boosting (5); number of iterations, learning rate, L1 penalty terms for Linear Gradient Boosting (3); number of trees, number of variables in linear part, L2 norm penalty term, size of the block for sub-sampling, and sampling method for the Macroeconomic Random Forest (5). We set the range of admissible values for each hyper-parameter based on the literature (e.g. Hastie et al., 2008), empirical evidence, and considerations of computational time.

findings, we apply this methodology with datasets mirroring data available at three different days of the month: namely on the 1st, 11th, and 21st days. More specifically, our baseline results always relate to the median dataset (at 11th day) with other days of months as consistency and robustness checks.

Figure 3. Pseudo real-time set-up



Section 3: Results

1. Pre-selection techniques

We first test the accuracy of the four different pre-selection techniques. Using one technique at a time, we produce out-of-sample predictions over Jan. 2012 to Apr. 2022 and compare the RMSFE with each technique. We use the same empirical set-up – PCA and OLS at respectively steps *ii* and *iii* – to ensure that differences in accuracy arise only from pre-selection.¹⁸ To check if differences in accuracy are consistent, we run such a comparison:

- Over different values of the number of regressors (n) selected by the pre-selection technique, ranging from 5 to 70 by step of 5.

¹⁸ While we elect LARS based on PCA and OLS, **Annex 4** shows that results are similar across other regression techniques: LARS is consistently found to be the best-performing pre-selection technique.

- Over different horizons, running it at our four horizons of interest: $t - 2$, $t - 1$ (backcasts), t (nowcast), and $t + 1$ (forecast).
- Over the three datasets corresponding to data available to a forecaster at the 1st, 11th, and 21st days of the month.

Results are reported in **Figure 4** for data at the 11th day:¹⁹ each panel refers to a given horizon (from $t - 2$ to $t + 1$) with the number of variables pre-selected (from 5 to 70) on the x-axis. The y-axis represents the accuracy, measured by the out-of-sample RMSFE over Jan. 2012 to Apr. 2022, relative to the no pre-selection benchmark (black dotted line). The LARS is generally the best-performing technique, with accuracy gains relative to the no pre-selection benchmark reaching up to 40%. For each horizon, the best accuracy is generally reached with LARS – a feature which remains generally consistent whatever the number n of variables pre-selected, with only the iterated BMA performing better for few n at few horizons. Tested more formally with Diebold and Mariano (1995) tests, differences in predictive accuracy between LARS and other techniques – including the absence of pre-selection – are significant at 10% at $t - 2$, but less as the horizon grows.

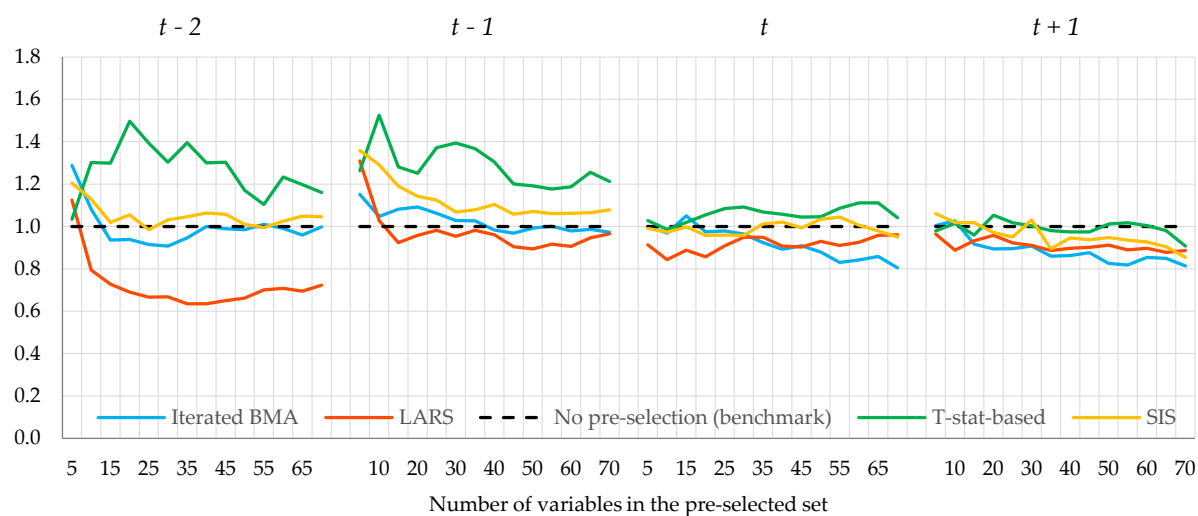
Finally, apart from comparing accuracy between pre-selection techniques, this exercise also allows determining the optimal number n of variables maximizing the accuracy for a given technique. This changes with the horizon but for simplicity we take $n = 60$ which generally correspond to the highest accuracy point for the LARS (red curve) and conveniently represent a pre-selection of roughly 10% of the initial dataset.^{20,21}

¹⁹ Results at other days of the month yield similar conclusions (see **Figures A1.1** and **A1.2** in **Annex 1**).

²⁰ The dataset has 536 raw series as detailed in **Annex 2** but for PMIs, we include the series both in levels and in cumulation over 12 months, increasing the size of the set by 63 series – so 599 in total.

²¹ Interestingly, pre-selection with LARS ends up with including variables that a forecaster might have discarded *a priori* such as the production of steel in Japan (but not in other countries), car registrations in Denmark (but not in other countries), and PMI new export orders from Brazil rather than for European countries. While their pairwise correlation with the target variable (CPB trade) might be low, these variables are likely selected because they bring value with respects to other variables in the pre-selected dataset.

Figure 4. Relative accuracy of pre-selection techniques



Notes: Accuracy is measured by the out-of-sample RMSE over Jan. 2012 - April 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. Upon pre-selection, factors are obtained through PCA and the regression is performed through OLS using the three first factors as independent variables.

2. Regression techniques

We now turn to the comparison across different regression techniques, also looking at out-of-sample predictions over Jan. 2012 to April 2022. We hold pre-selection (LARS for 60 variables) and factor extraction (PCA) fixed so that differences arise only from regression.²² To check if differences in accuracy are consistent, we again run the comparison (i) over the four horizons of interest: $t - 2$, $t - 1$ (back-casts), t (nowcast), and $t + 1$ (forecast), and (ii) over the three datasets corresponding to data available to a forecaster at the 1st, 11th, and 21st days of the month.

Results show that regression-based ML methods perform best at all horizons, with accuracy around 20% on average relative to the OLS benchmark. Accuracy gains can reach up to 33% and are consistent across horizons and datasets at other days of the month (see **Figures A1.3** and **A1.4** in **Annex 1** for respectively 1st and 21st days). An interesting feature is that regression-based ML methods significantly outperform tree-based methods despite sharing a similar

²² As mentioned above, we elect PCA for factor extraction. The comparison of accuracy across a range of factor extraction techniques (PCA, 2-step, quasi maximum likelihood, dynamic PCA) can be found in **Annex 3**. Accuracy is very close across techniques, in line with similar empirical findings in Kapetanios and Marcellino (2009). In the end, we apply Occam's razor (if different methods have equivalent performance, pick the simplest) and use PCA.

framework, as the regression-based ML methods are generally an adaptation of their tree-based counterparts. This suggests that in the short time samples common in macroeconomics, forecasts might be better when relying on regression-based ML methods than on the most widely known tree-based ML methods. In that sense, the fact that most of the literature on machine learning reports little to no gains when using ML methods compared with more traditional frameworks (e.g. Richardson et al., 2019) might be due to the fact that those studies generally use such tree-based methods, which are also found to perform relatively poorly in our set-up. Among the regression-based ML methods, the macroeconomic random forest (dark red) is outperforming the linear gradient boosting (light red) consistently at all horizons, albeit by a small margin. Finally, looking at panels B (middle) and C (bottom) shows that the outperformance of these models relative to OLS and to other models is greater during crisis periods. But these models still outperform others during normal times.

More formally, we test the significance of the differences in predictive accuracy using Diebold-Mariano tests. Results are provided in **Table 1** for horizon $t - 2$ – with results for other horizons in **Tables A1.1 to A1.3** in **Annex 1**.²³ While all models significantly outperform the AR benchmark, only the regression-based machine learning techniques do so relative to the OLS benchmark. Moreover, these techniques also outperform non-linear competitors, both “traditional” (Markov-switching and quantile regression) and tree-based machine learning (random forest and linear gradient boosting). As a complement, we also run Model Confidence Set (MCS) tests – provided in **Table A1.4** in **Annex 1** – suggesting the same outperformance of the macroeconomic random forest as generally the only model in the 90% confidence set.²⁴

²³ Results are given for data at the 11th day of the month. Results for other days of the month are similar.

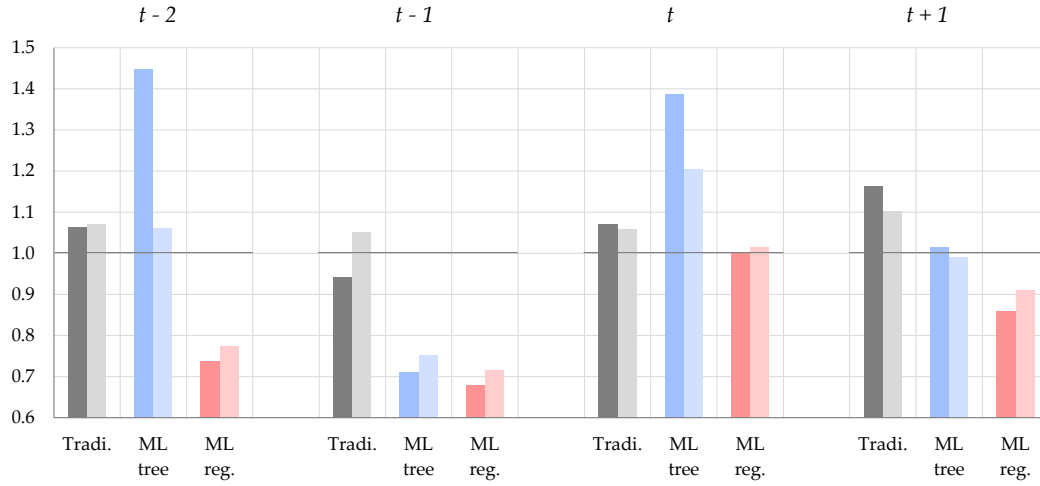
²⁴ Because nested models can weaken the inference in Diebold-Mariano tests (Clark and McCracken, 2001), we complement Diebold-Mariano with a MCS test of Hansen et al. (2011). Results of the MCS test relates to the dataset at the 11th day of the month. MCS test results for datasets at other days of the month are similar.

Figure 5. Accuracy of regression techniques (relative to OLS)

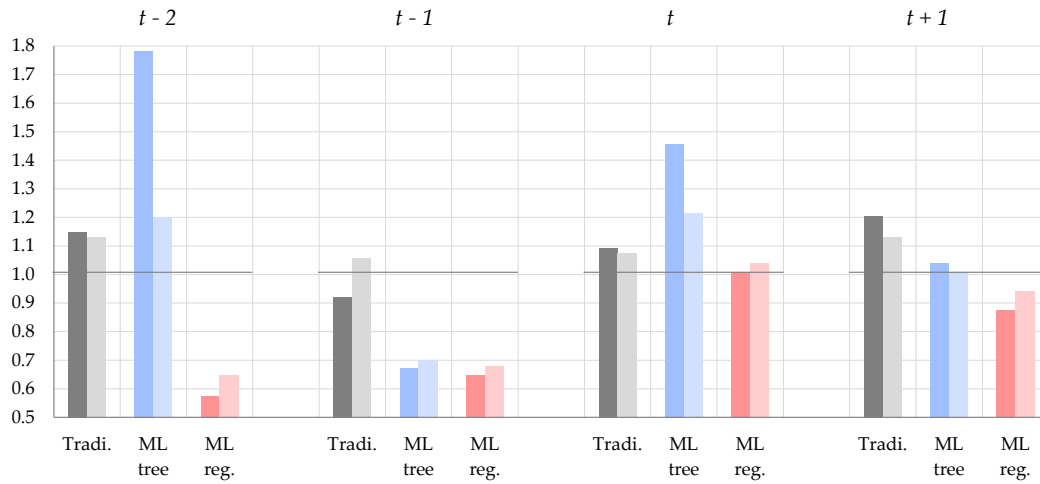
Markov-switching
 Random forest
 Macroeconomic random forest

 Quantile reg.
 Gradient boosting
 Gradient linear boosting

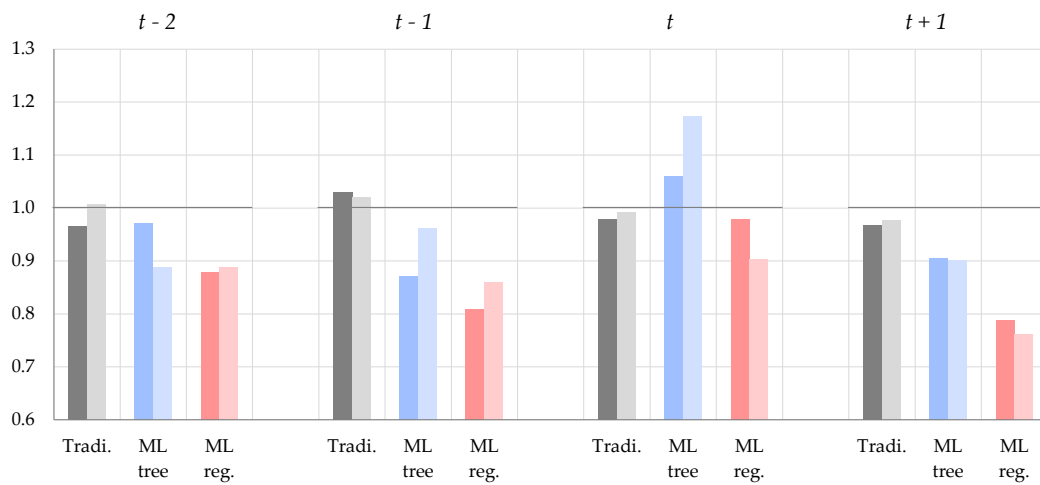
Panel A. Full sample (Jan. 2012 - Apr. 2022)



Panel B. Crisis sample (Jan. 2020 - Apr. 2022)



Panel C. Normal sample (Jan. 2012 - Dec. 2019)



Notes: Accuracy is measured by the out-of-sample RMSE over Jan. 2012 - April 2022. Performances are presented relative to the OLS benchmark (dark grey line). Results are obtained for the dataset mirroring data available to a forecaster at the 11th day of the month, using a LARS for pre-selecting the 60 most informative regressors, with factors extracted through PCA on the pre-selected set.

Table 1: Diebold-Mariano test results (horizon = $t - 2$)								
	Benchmarks		"Traditional"		ML tree-based		ML regression-based	
	AR	OLS	MS	QR	RF	XGB-T	MRF	XGB-L
AR		0.98	0.98	0.98	0.98	1.00	0.99	0.99
OLS	0.02		0.14	0.07	0.07	0.37	0.98	0.95
MS	0.02	0.86		0.45	0.10	0.51	0.98	0.96
QR	0.02	0.93	0.55		0.10	0.53	0.98	0.96
RF	0.02	0.93	0.90	0.90		0.95	0.99	0.98
GB-T	0.00	0.63	0.49	0.47	0.05		0.96	0.94
MRF	0.01	0.02	0.02	0.02	0.01	0.04		0.19
GB-L	0.01	0.05	0.04	0.04	0.02	0.06	0.81	

Notes: MS = Markov-switching; QR = quantile regression; RF = random forest; GB-T = gradient-boosting; MRF = Macroeconomic Random Forest, GB-L = linear gradient boosting. Tests are based on one-period-ahead out-of-sample predictions over Jan. 2012 to Apr. 2022. Results are obtained for a dataset mirroring the data available to a forecaster at the 11th day of the month. The table reports *p*-values for the one-sided test of Diebold and Mariano (1995) with the small sample correction introduced by Harvey et al. (1997). Light grey cells indicate significance at 10%.

Section 4: Quantifying the gains from the three-step approach

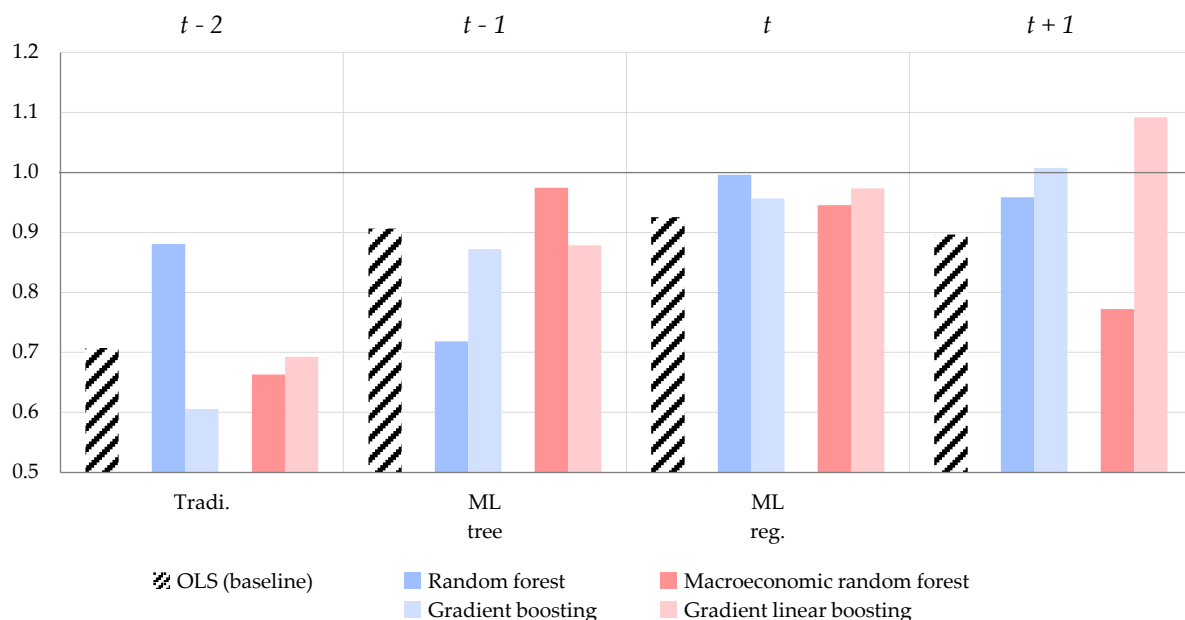
Our first robustness check relates to what extent pre-selection – step 1 in our proposed three-step approach – consistently improves accuracy. We conduct two tests to assess whether the results of section 3.1 are contingent to the set-up used.

- With results of section 3.1 based only on OLS, we first check whether pre-selection also entails gains for other regression techniques. In particular for machine learning techniques, pre-selection might be less useful as these techniques are to some extent designed to accommodate for large amount of data. We test this formally by running such regressions with no prior pre-selection *vs.* with a LARS. Results are reported in **Figure 6** relative to the no-pre-selection alternative, meaning that a value below 1 (dark grey line) indicates outperformance of the approach with pre-selection.²⁵ It is important to note that each bar compares within one technique: in other words, a bar for "random forest" at 0.9 and a bar for "OLS" at 0.7 means that pre-selection improves accuracy by 10% for the random forest and by 30% for OLS, but does not mean that

²⁵ Results are given for data at the 11th day of the month. Results for other days of the month are similar.

OLS outperforms random forest. Gains in accuracy from pre-selection are similar across methods and are similar to those for the OLS. This shows that gains from the pre-selection step are not contingent to the type of regression used, suggesting the robustness of the three-step approach. Beyond this, the results suggest that pre-selection can enhance nowcasting performances also for machine learning techniques. This complements Goulet-Coulombe et al. (2022) who found evidence that the best-performing nowcasting method is to perform a PCA, then use factors in machine learning regressions. Our paper adds to it that a pre-selection before PCA can further improve the accuracy, by up to 40% at some horizons.

Figure 6. Accuracy of LARS relative to no pre-selection

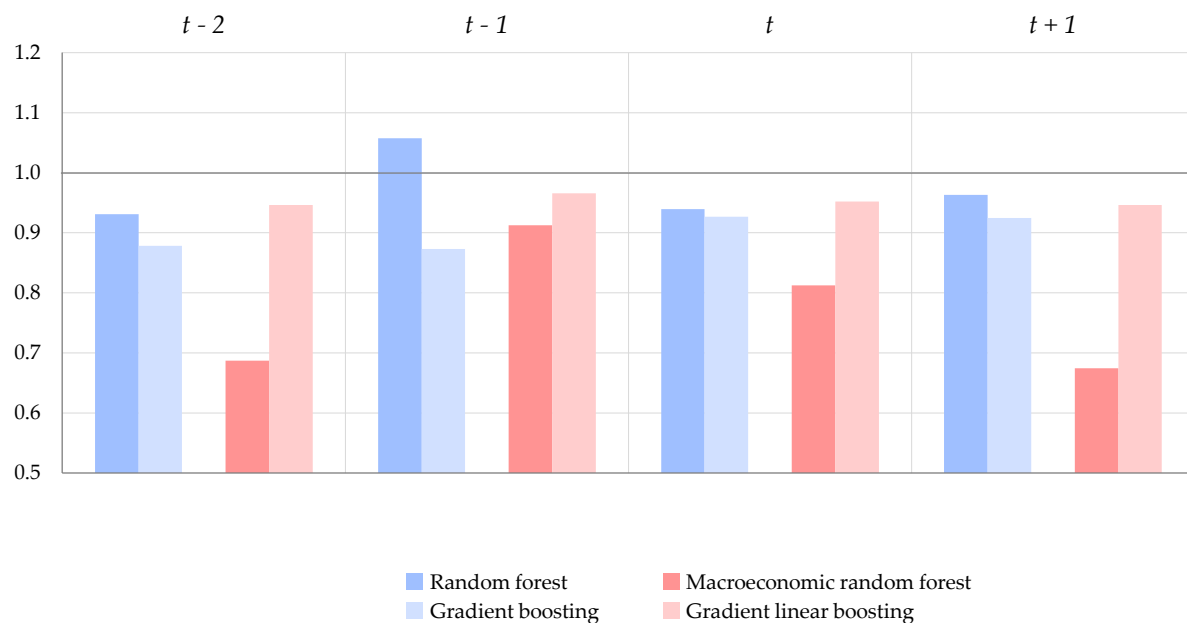


Notes: Accuracy is measured by the out-of-sample RMSE over Jan. 2012 - April 2022. Performances are presented relative to the benchmark of no pre-selection (dark grey line). Results are obtained for the dataset mirroring data available to a forecaster at the 11th day of the month, using a LARS for pre-selecting the 60 most informative regressors, with factors extracted through PCA on the pre-selected set.

- We also check whether accuracy gains from pre-selection are contingent to the bespoke re-alignment of data. In particular, the addition of new series for regressors with “excess” data could give an advantage to pre-selection since it potentially adds more uninformative series. We therefore compare accuracy of LARS *vs.* no-pre-selection for a dataset re-aligned using the baseline Altissimo et al. (2006) method. Accuracy gains from LARS are very similar in this case, suggesting that the value-added of pre-selection is not contingent on the realignment strategy.

We then check whether factor extraction improves performances. In particular, while factor extraction has been proven handy for OLS (e.g. Jolliffe, 2002), it might be less useful for machine learning techniques designed to handle high-dimensional dataset. We therefore compare the accuracy of our three-step approach – in which factors are extracted from the pre-selected dataset – with an alternative without factor extraction – in which pre-selected variables are fed directly into the regression.²⁶

Figure 7. Accuracy relative to the alternative with no factors



Notes: Accuracy is measured by the out-of-sample RMSE over Jan. 2012 - April 2022. Performances are presented relative to the benchmark without factor extraction (dark grey line). Results are obtained for the dataset mirroring data available to a forecaster at the 11th day of the month, using a LARS for pre-selecting the 60 most informative regressors.

Results are reported in **Figure 7** relative to the alternative without factor extraction, meaning that a value below 1 (dark grey line) indicates outperformance of the approach with factors.²⁷ The same caveat as for **Figure 6** applies: each bar is a comparison with the same technique so that no conclusion can be drawn on the performances of techniques *between* them, only on the gains of factor extraction *for a given technique*. Across all techniques and all horizons, extracting factors enhance the performances by up to 33%. This suggests that summarising information into a few factors helps. For machine learning techniques, this can be explained by: (i) the fact

²⁶ This robustness check is conducted only for machine learning techniques. The results are the same for other techniques, however these methods (OLS, Markov-switching, quantile regression) are known in the literature to perform badly if the number of independent variables is large. We therefore focus on machine learning techniques, for which gains from factor extraction are less straightforward and have been less explored in the literature so far.

²⁷ Results are given for data at the 11th day of the month. Results for other days of the month are similar.

that parsimony can be still key given the limited number of observations, in that respect having fewer variables could help the model learn more accurately the non-linearities and patterns in the data, and (ii) factor extraction not only summarises information but also ensures that the resulting factors are orthogonal, which can be important to speed the convergence of machine learning techniques in particular those relying on the independence across weak learners.²⁸ These results are also in line with Goulet-Coulombe et al. (2022) finding that using PCA-extracted factors in machine learning regressions is the best-performing method for nowcasting, compared to feeding directly a high-dimensional dataset in machine learning techniques.

Having tested the benefits of each step of the proposed three-step approach, we now turn to checking whether treating these steps *sequentially* brings value compared to methods able to perform such steps *simultaneously*. A first benchmark is the elastic net (Zou and Hastie, 2005) which handles variable selection and regression simultaneously. The elastic net is done on the full dataset (i.e. without pre-selection) with hyper-parameters optimized at each period.²⁹ Another benchmark relates to dynamic factor models (DFM) which handle factor extraction and regression in an integrated framework. To make the DFM set-up more comparable with the three-step approach, the DFM is based on the pre-selected dataset. The benchmark is then close to Runstler (2016) who recommends a LARS pre-selection combined with DFM. The DFM is based on quasi-maximum likelihood estimation, which has the advantage of being more general as it can accommodate for *approximate* factor models where the assumption that idiosyncratic components are mutually uncorrelated at all lags is relaxed. More specifically, we use the algorithm developed in Banbura and Modugno (2014) and widely used in nowcasting application (e.g. New York Fed’s US nowcasting).

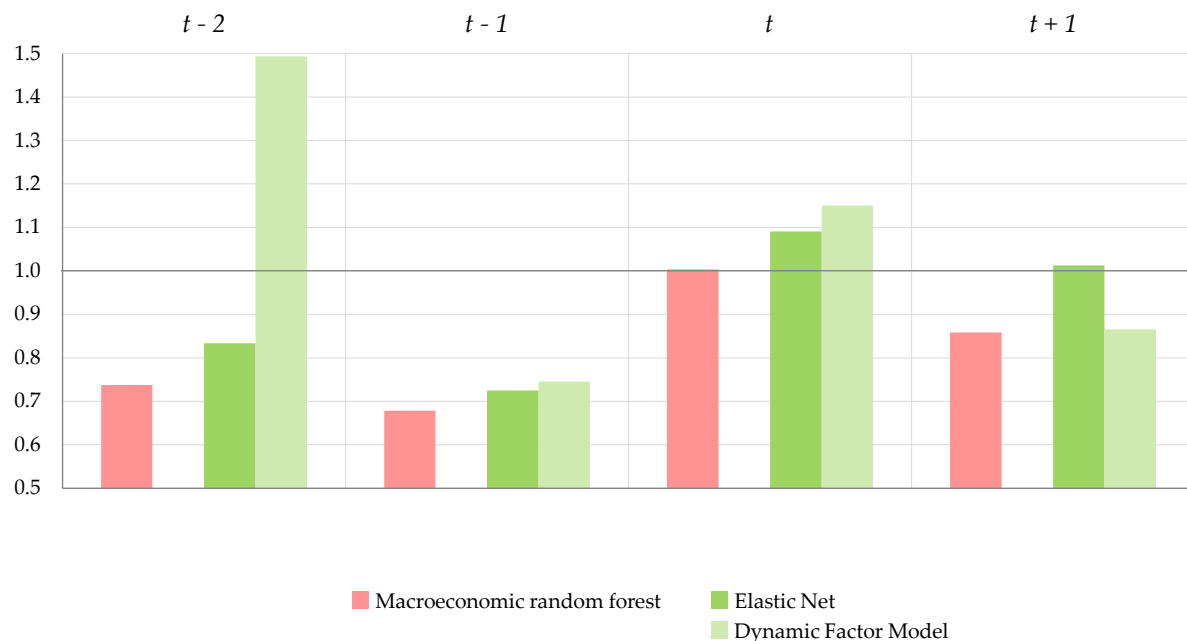
²⁸ This might be the case in particular for random forest techniques where the underlying idea is to build trees that are independent from each other. In that respect, having orthogonal variables could help.

²⁹ Hyper-parameters for the elastic net are λ (penalty term) and α (weight of L2 norm relative to L1 norm in the penalty). They are set based on 10-fold cross-validation. As a reminder, elastic net coefficients β_i are determined by minimizing the expression 25.1 below – which is sum of squared residuals with a penalty:

$$(25.1) \quad \sum_i (y_i - \beta_i \cdot x_i)^2 + \lambda \cdot [\alpha \cdot \|\beta\|_2^2 + (1 - \alpha) \cdot \|\beta\|_1]$$

Results are provided in **Figure 8** relative to the OLS benchmark.³⁰ A value below 1 indicates outperformance relative to the OLS (black line). As the benchmark remains the same, conclusions can be drawn on the relative accuracy of the different techniques. The three-step technique using macroeconomic random forest, LARS and PCA (red bar), is the best-performing techniques across all horizons.³¹ It is also the only technique to consistently outperform the OLS benchmark – as indicated by a bar always below 1. This suggests that following a step-by-step approach as proposed in our framework pays off compared to techniques encompassing several steps (variable selection, factor extraction, regression) in an integrated framework.

Figure 8. Accuracy relative to OLS



Notes: Accuracy is measured by the out-of-sample RMSE over Jan. 2012 - April 2022. Performances are presented relative to the OLS benchmark (dark grey line). Results are obtained for the dataset mirroring data available to a forecaster at the 11th day of the month, using LARS for pre-selecting the 60 most informative regressors.

³⁰ Results are given for data at the 11th day of the month. Results for other days of the month are similar.

³¹ It can be also noted that our results up to t confirm Charles and Darne (2022) finding that penalized regressions (such as the elastic net) can outperform a DFM for back-casting CPB trade.

Conclusion

This paper uses a three-step approach composed of pre-selection, factor extraction, and non-linear regression for nowcasting trade with a large dataset. Such an approach outperforms other methods, notably a standard diffusion index approach (Stock and Watson, 2002) and a DFM. Looking at the gains step by step, we show that (i) pre-selecting regressors can enhance performances by around 10-15% on average and up to 40%; (ii) factor extraction entails around 10-15% further gains, including also for machine learning techniques despite the ability of such methods to accommodate for high-dimensional datasets; and (iii) using machine learning techniques can further improve the accuracy by around 15-20%. We check the consistency of accuracy gains across different prediction horizons and different datasets mirroring the data available to the forecaster at different days of the month. The three-step approach is in addition highly flexible and can adapt various methods for pre-selection, factor extraction, or regression. The interest of the three-step is that it can also be seamlessly adapted to any other forecast exercise with a large dataset. In particular, having an automated pre-selection alleviates to some extent the need for the forecaster to have prior knowledge about the data and to hand-pick variables first.

A second contribution of the paper relates to the test of several techniques for each of the steps: pre-selection, factor extraction, and regression. Starting with the latter, a key finding is to draw a distinction between machine learning techniques based on *trees* and those based on *regressions*. We show that machine learning based on trees, which are the most common in the literature, have generally limited performances compared to a simpler OLS. In contrast, machine learning techniques based on regressions – macroeconomic random forest and linear gradient boosting that are much less used in the literature – are found to outperform significantly and consistently all competitors: not only OLS but also other non-linear techniques (Markov-switching, quantile regression) as well as the machine learning techniques based on trees (random forest, gradient boosting). In that respect, this paper sheds some light on these techniques and demonstrates their significant accuracy gains in nowcasting. In another endeavour, this paper also tests several pre-selection techniques: in line with the literature (e.g. Bai and Ng, 2008; Runstler, 2016; Jardet and Meunier, 2022), we find that LARS outperforms other techniques albeit marginally at some horizons. Testing

across different factor extraction techniques, we finally find that differences in accuracy are generally not significant and therefore elects the most straightforward method (PCA). In the end, our best-performing nowcasting model (i.e. a triplet of one pre-selection technique, one factor extraction method, and one regression model) is achieved using LARS for pre-selection, PCA for factor extraction, and macroeconomic random forest (Goulet-Coulombe, 2020) for regression.

While this paper focuses solely on CPB trade data, an avenue for future research could be the generalization of the approach to other macroeconomic variables. In addition, macroeconomic random forest has been shown to have the ability to derive contributions from the different variables in the linear part: while the focus on our paper is solely on accuracy, such a feature could be exploited to open the “black box” of nowcasting and interpret the predictions.

References

- Adrian, T., Boyarchenko, N., and Giannone, D. (2019). "Vulnerable growth", *American Economic Review*, 109, pp. 1263–1289
- d'Agostino, A., Modugno, M., and Osbat, C. (2017). "A Global Trade Model for the Euro Area", *International Journal of Central Banking*, 13(4), pp. 1–34
- Altissimo, F., Cristadoro, R., Forni, M., Lippi, M., and Veronese, G. (2006). "New Eurocoin: Tracking Economic Growth in Real Time", *CEPR Discussion Papers*, No 5633
- Amisano, G., and Geweke, J. (2017). "Prediction using several macroeconomic models", *The Review of Economics and Statistics*, 99(5), pp. 912-925
- Bai, J., and Ng, S. (2002). "Determining the Number of Factors in Approximate Factor Models", *Econometrica*, 70(1), pp. 191–221
- Bai, J., and Ng, S. (2007). "Determining the Number of Primitive Shocks in Factor Models", *Journal of Business & Economic Statistics*, 25(1), pp. 52–60
- Bai, J., and Ng, S. (2008). "Forecasting economic time series using targeted predictors", *Journal of Econometrics*, 146(2), pp. 304–317
- Bair, E., Hastie, T., Paul, D., and Tibshirani, R. (2006). "Prediction by supervised principal components", *Journal of the American Statistical Association*, 101(473), pp. 119–137
- Banbura, M., and Modugno, M. (2014). "Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data", *Journal of Applied Econometrics*, 29(1), pp. 133–160
- Barhoumi, K., Darné, O., and Ferrara, L. (2016). "A world trade leading index (WTLI)", *Economic Letters*, 146, pp. 111–115.
- Boivin, J. and Ng, S. (2006). "Are more data always better for factor analysis", *Journal of Econometrics*, 132, pp. 169–194
- Boot, T., and Pick, A. (2018). "Optimal Forecasts from Markov Switching Models", *Journal of Business & Economic Statistics*, 36(4), pp. 628–642

- Breiman, L. (2001). "Random Forests", *Machine Learning*, 45(1), pp. 5–32
- Bricongne J-C., Meunier B., and Pical, T. (2021). "Can satellite data on air pollution predict industrial production?", *Banque de France Working Papers*, No. 847
- Bricongne J-C., Meunier B., and Pouget, S. (2023). "Web Scraping Housing Prices in Real-time: the Covid-19 Crisis in the UK", *Journal of Housing Economics*, 101906
- Bulligan, G., Marcellino, M., and Venditti, F. (2015). "Forecasting economic activity with targeted predictors", *International Journal of Forecasting*, 31(1), pp. 188–206
- Bussière, M., Callegari, G., Ghironi, F., Sestieri, G., and Yamano, N. (2013). "Estimating Trade Elasticities: Demand Composition and the Trade Collapse of 2008-2009", *American Economic Journal: Macroeconomics*, 5(3), pp. 118–151
- Carstensen, K., Heinrich, M., Reif, M., and Wolters, M. (2020). "Predicting ordinary and severe recessions with a three-state Markov-switching dynamic factor model: An application to the German business cycle", *International Journal of Forecasting*, 36(3), pp. 829–850
- Carvalho V., Garcia J., Hansen S., Ortiz Á., Rodrigo T., Rodríguez Mora J., and Ruiz J. (2020). "Tracking the COVID-19 Crisis with High-Resolution Transaction Data", *Cambridge-INET Working Papers*, No 2016
- Charles, A., and Darné, O. (2022). "Backcasting world trade growth using data reduction methods", *The World Economy*, 45(10), pp. 3169–3191
- Chen, T., and Guestrin, C. (2016). "XGBoost: A Scalable Tree Boosting System", in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794
- Chen S., Igan D., Pierri N., and Presbitero A. (2020). "Tracking the Economic Impact of COVID-19 and Mitigation Policies in Europe and the United States", *International Monetary Fund Working Papers*, No 20/125
- Chinn, M., Meunier, B., and Stumpner, S. (2023). "Nowcast world trade in goods with machine learning: a three-step approach", *Bulletin de la Banque de France*, forthcoming

- Cerdeiro D., Komaromi A., Liu Y., and Saeed M. (2020). "World Seaborne Trade in Real Time: A Proof of Concept for Building AIS-based Nowcasts from Scratch", *International Monetary Fund Working Papers*, No 20/57
- Clark, T., and McCracken, M. (2001). "Tests of equal forecast accuracy and encompassing for nested models", *Journal of Econometrics*, 105(1), pp. 85–110
- Doz, C., Giannone, G., and Reichlin, L. (2011). "A two-step estimator for large approximate dynamic factor models based on Kalman filtering", *Journal of Econometrics*, 164(1), pp. 188–205
- Doz, C., Giannone, D., and Reichlin, L. (2012). "A quasi maximum likelihood approach for large approximate dynamic factor models", *The Review of Economics and Statistics*, 94(4), pp. 1014–1024
- Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). "Least angle regression", *Annals of Statistics*, 32(2), pp. 407–499
- Engel, C., and Hamilton, J. (1990). "Long Swings in the Dollar: Are They in the Data and Do Markets Know It?", *American Economic Review*, 80(4), pp. 689–713
- Falagiarda, M., and Sousa, J. (2015). "Forecasting euro area inflation using targeted predictors: is money coming back?", *European Central Bank Working Paper Series*, No 2015
- Fan, J., and Lv, J. (2008). "Sure independence screening for ultrahigh dimensional feature space", *Journal of the Royal Statistical Society Series B*, 70(5), pp. 849–911
- Ferrara, L., and Marsilli, C. (2019). "Nowcasting global economic growth: A factor-augmented mixed-frequency approach", *The World Economy*, 42(3), pp. 846–875
- Ferrara, L., and Simoni, A. (2019). "When are Google data useful to nowcast GDP? An approach via pre-selection and shrinkage", *CREST Working Papers*, No 2019-04
- Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005). "The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting", *Journal of the American Statistical Association*, 100(471), pp. 830–840

- Friedman, J. (2001). "Greedy Function Approximation: A Gradient Boosting Machine", *The Annals of Statistics*, 29(5), pp. 1189–1232
- Giannone, D., Reichlin, L., and Sala, L. (2005). "Monetary Policy in Real Time", NBER Chapters, in: *NBER Macroeconomics Annual 2004*, 19, pp. 161–224
- Giannone, D., Reichlin, L., and Small, D. (2008). "Nowcasting: The real-time informational content of macroeconomic data", *Journal of Monetary Economics*, 55(4), pp. 665–676
- Goulet-Coulombe, P. (2020). "The Macroeconomy as a Random Forest", *arXiv pre-print*
- 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), pp. 920–964
- Guichard, S., and Rusticelli, E. (2011). "A Dynamic Factor Model for World Trade Growth", *OECD Economics Department Working Papers*, No 874
- Hallin, M., and Liška, R. (2007). "Determining the Number of Factors in the General Dynamic Factor Model", *Journal of the American Statistical Association*, 102(478), pp. 603–617
- Hamilton, J. (1989). "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", *Econometrica*, 57(2), pp. 357–384
- Hansen, P., Lunde, A., and Nason, J. (2011). "The model confidence set", *Econometrica*, 79(2), pp. 453–497
- Harvey, D., Leybourne, S., and Newbold, P. (1997). "Testing the equality of prediction mean squared errors", *International Journal of Forecasting*, 13(2), pp. 281–291
- Hasenzagl, T., Plagborg-Møller, M., Reichlin, L., and Ricco, G. (2020). "When is growth at risk?", *Brookings Papers on Economic Activity*, Spring, pp. 167–229
- Hastie, T., Tibshirani, R., and Friedman, J. (2008). *The Elements of Statistical Learning*, Springer Series in Statistics, 2nd edition
- Hoeting, J., Madigan, B., Raftery, A., and Volinsky, C. (1999). "Bayesian Model averaging: A tutorial", *Statistical Science*, 14(4), pp. 382–417

- Jardet, C., and Meunier, B. (2022). "Nowcasting World GDP Growth with High-Frequency Data", *Journal of Forecasting*, 41(6), pp. 1181–1200
- Jakaitiene, A., and Dees, S. (2012). "Forecasting the world economy in the short term", *The World Economy*, 35(3), pp. 331–350.
- Jolliffe, I. (2002). "Principal Component Analysis for special types of data", *Principal components analysis*, pp. 338–372
- Jurado, K., Ludvigson, S., and Ng, S. (2015). "Measuring Uncertainty", *American Economic Review*, 105(3), pp. 1177–1216
- Hopp, D. (2021). "Economic Nowcasting with Long Short-Term Memory Artificial Neural Networks (LSTM)", *arXiv pre-print*
- Kaiser, H. (1960). "The application of electronic computers to factor analysis", *Educational and Psychological Measurement*, 20, pp. 141–151
- Kapetanios, G., and Marcellino, M. (2009). "A parametric estimation method for dynamic factor models of large dimensions", *Journal of Time Series Analysis*, 30(2), pp. 208–238
- Keck, A., Raubold, A., and Truppia, A. (2010). "Forecasting international trade: A time series approach", *OECD Journal: Journal of Business Cycle Measurement and Analysis*, vol. 2009/2
- Koenker, R., and Bassett, G. (1978). "Regression quantiles", *Econometrica*, 46, pp. 33–50
- Marcellino, M., and Schumacher, C. (2010). "Factor-MIDAS for now- and forecasting with ragged-edge data: A model comparison for German GDP", *Oxford Bulletin of Economics and Statistics*, vol. 72(4), pp. 518–550
- Martínez-Martín, J., and Rusticelli, E. (2021). "Keeping track of global trade in real time", *International Journal of Forecasting*, 37(1), pp. 224–236
- Mazzi, G., and Mitchell, J. (2020). "New methods for timely estimates: nowcasting euro area GDP growth using quantile regression", *Statistical working papers*, Eurostat

- Medeiros, M., Vasconcelos, G., Veiga, A., and Zilberman, E. (2021). "Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods", *Journal of Business & Economic Statistics*, 39(1), pp. 98–119
- Proietti, T., and Giovannelli, A. (2021). "Nowcasting monthly GDP with big data: A model averaging approach", *Journal of the Royal Statistical Society: Series A*, 184(2), pp. 683–706
- Miranda, K., Poncela, P., and Ruiz, E. (2022). "Dynamic factor models: Does the specification matter?", *SERIEs*, 13, pp. 397–428
- Raftery, A. (1995). "Bayesian model selection in social research (with Discussion)", *Sociological Methodology 1995*, Peter Marsden (ed.), pp. 111-196, Cambridge, Blackwells
- Richardson, A., van Florenstein Mulder, T., and Vehbi, T. (2019). "Nowcasting New Zealand GDP using machine learning algorithms", IFC Bulletins chapters, in: Bank for International Settlements (ed.), *The use of big data analytics and artificial intelligence in central banking*, 50
- Rünstler, G. (2016). "On the Design of Data Sets for Forecasting with Dynamic Factor Models", *Advances in econometrics*, Eric Hillebrand and Siem Jan Koopman (eds.), Dynamic Factor Models, 35, pp. 629–662, Emerald Publishing Ltd.
- Schumacher, C. (2010). "Factor forecasting using international targeted predictors: The case of German GDP", *Economics Letters*, 107(2), pp. 95–98
- Soybilgen, B., and Yazgan, E. (2021). "Nowcasting US GDP Using Tree-Based Ensemble Models and Dynamic Factors", *Computational Economics*, 57, pp. 387–417
- Stratford, K. (2013). "Nowcasting world GDP and trade using global indicators", *Bank of England Quarterly Bulletin*, vol. 53(3), pp. 233–242
- Stock, J., and Watson, M. (2002). "Forecasting using principal components from a large number of predictors", *Journal of the American Statistical Association*, 97(460), pp. 1167–1179
- Tibshirani, R. (1996). "Regression shrinkage and selection via the Lasso", *Journal of the Royal Statistical Society: Series B*, 58, pp. 267–288

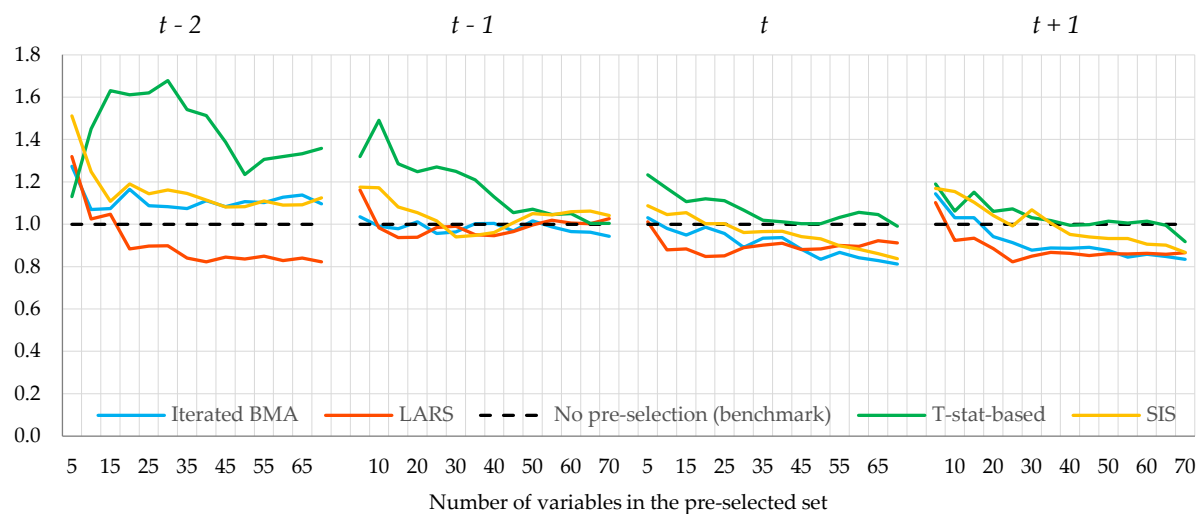
Yeung, K., Bumgarner, R., and Raftery, A. (2005). "Bayesian Model Averaging: Development of an improved multi-class, gene selection and classification tool for microarray data", *Bioinformatics*, 21(10), pp. 2394–2402

Yoon, J. (2021). "Forecasting of Real GDP Growth Using Machine Learning Models: Gradient Boosting and Random Forest Approach", *Computational Economics*, 57, pp. 1–19

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

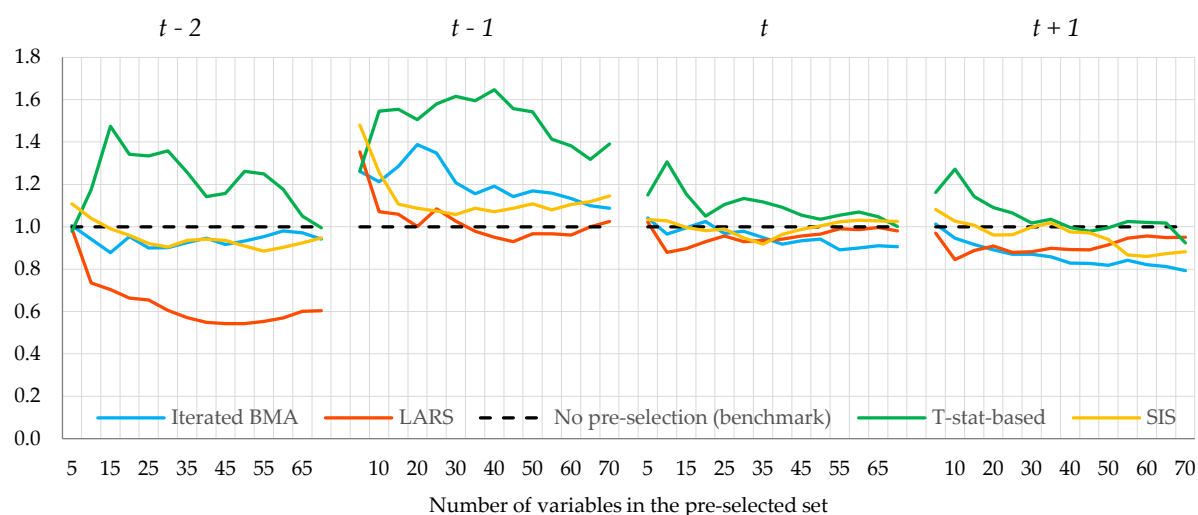
Annex 1: Background charts and tables

Figure A1.1. Relative accuracy of pre-selection techniques (1st day)



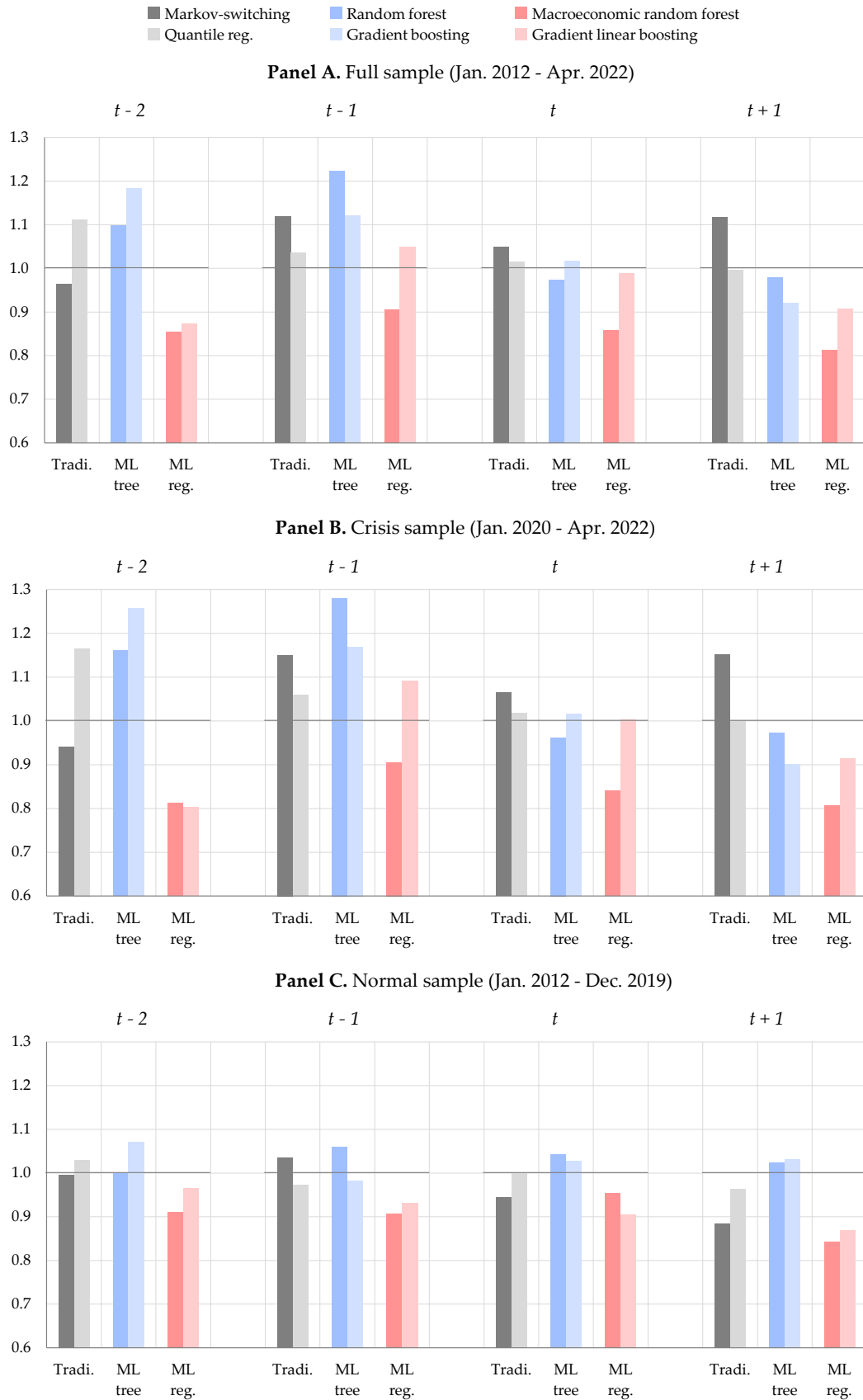
Notes: Accuracy is measured by the out-of-sample RMSE over Jan. 2012 - April 2022. Performances are presented relative to the benchmark of no pre-selection (black dotted line). Results are obtained for the dataset mirroring data available to a forecaster at the 1st day of the month. Upon pre-selection, factors are obtained through PCA and the regression is performed through OLS using the three first factors as independent variables.

Figure A1.2. Relative accuracy of pre-selection techniques (21st day)



Notes: Accuracy is measured by the out-of-sample RMSE over Jan. 2012 - April 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. Upon pre-selection, factors are obtained through PCA and the regression is performed through OLS using the three first factors as independent variables.

Figure A1.3. Accuracy of regression techniques (relative to OLS) - 1st day

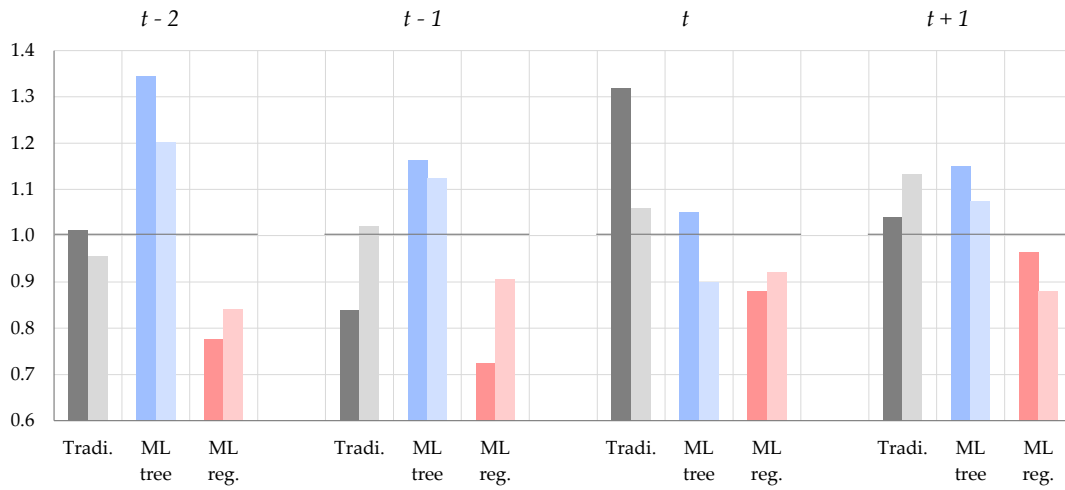


Notes: Accuracy is measured by the out-of-sample RMSE over Jan. 2012 - April 2022. Performances are presented relative to the OLS benchmark (dark grey line). Results are obtained for the dataset mirroring data available to a forecaster at the 1st day of the month, using a LARS for pre-selecting the 60 most informative regressors, with factors extracted through PCA on the pre-selected set.

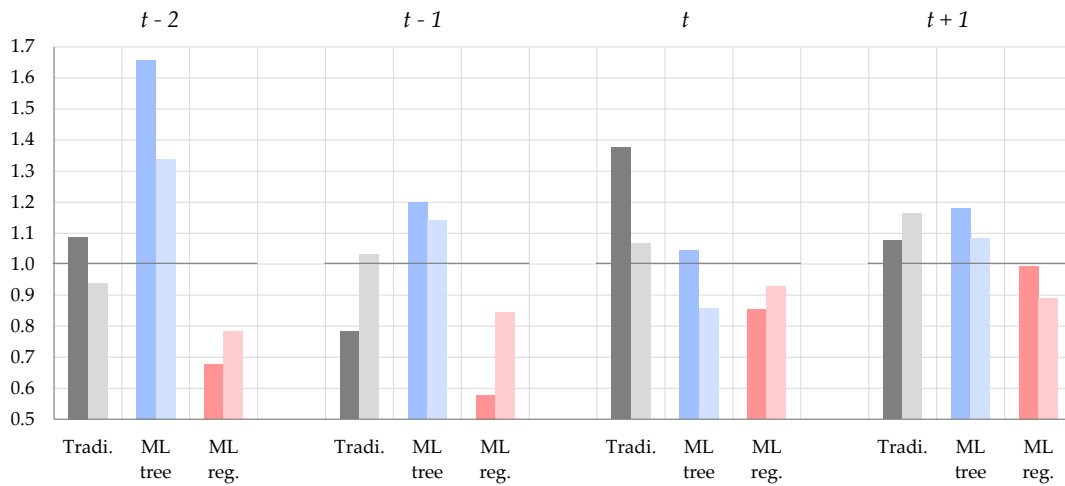
Figure A1.4 Accuracy of regression techniques (relative to OLS) - 21st day

Markov-switching
 Random forest
 Macroeconomic random forest
 Quantile reg.
 Gradient boosting
 Gradient linear boosting

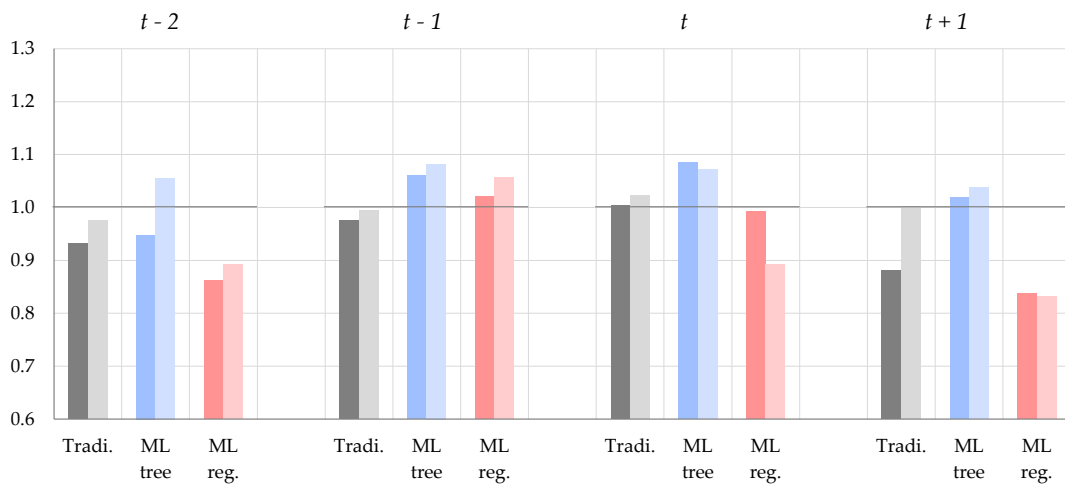
Panel A. Full sample (Jan. 2012 - Apr. 2022)



Panel B. Crisis sample (Jan. 2020 - Apr. 2022)



Panel C. Normal sample (Jan. 2019 - Dec. 2019)



Notes: Accuracy is measured by the out-of-sample RMSE over Jan. 2012 - April 2022. Performances are presented relative to the OLS benchmark (dark grey line). Results are obtained for the dataset mirroring data available to a forecaster at the 21st day of the month, using a LARS for pre-selecting the 60 most informative regressors, with factors extracted through PCA on the pre-selected set.

Table A1.1: Diebold-Mariano test results (horizon = $t - 1$)								
	Benchmarks		"Traditional"		ML tree-based		ML regression-based	
	AR	OLS	MS	QR	RF	XGB-T	MRF	XGB-L
AR		0.91	0.99	0.82	0.99	0.99	0.99	0.99
OLS	0.09		0.77	0.04	0.86	0.83	0.89	0.88
MS	0.01	0.23		0.14	0.90	0.86	0.93	0.92
QR	0.18	0.96	0.86		0.89	0.86	0.91	0.90
RF	0.01	0.14	0.10	0.11		0.17	0.68	0.48
GB-T	0.01	0.17	0.14	0.14	0.83		0.92	0.76
MRF	0.01	0.11	0.07	0.09	0.32	0.08		0.09
GB-L	0.01	0.12	0.08	0.10	0.52	0.24	0.91	

Notes: MS = Markov-switching; QR = quantile regression; RF = random forest; GB-T = gradient-boosting; MRF = Macroeconomic Random Forest, GB-L = linear gradient boosting. Tests are based on one-period-ahead out-of-sample predictions over Jan. 2012 to Apr. 2022. Results are obtained for a dataset mirroring the data available to a forecaster at the 11th day of the month. The table reports *p*-values for the one-sided test of Diebold and Mariano (1995) with the small sample correction introduced by Harvey et al. (1997). Light grey cells indicate significance at 10%.

Table A1.2: Diebold-Mariano test results (horizon = t)								
	Benchmarks		"Traditional"		ML tree-based		ML regression-based	
	AR	OLS	MS	QR	RF	XGB-T	MRF	XGB-L
AR		1.00	1.00	1.00	0.86	0.99	1.00	1.00
OLS	0.00		0.08	0.09	0.02	0.04	0.47	0.24
MS	0.00	0.92		0.57	0.03	0.13	0.81	0.83
QR	0.00	0.91	0.43		0.02	0.04	0.90	0.90
RF	0.14	0.98	0.97	0.98		0.94	0.99	0.98
GB-T	0.01	0.96	0.87	0.96	0.06		0.98	0.96
MRF	0.00	0.53	0.19	0.10	0.01	0.02		0.39
GB-L	0.00	0.76	0.17	0.10	0.02	0.04	0.61	

Notes: MS = Markov-switching; QR = quantile regression; RF = random forest; GB-T = gradient-boosting; MRF = Macroeconomic Random Forest, GB-L = linear gradient boosting. Tests are based on one-period-ahead out-of-sample predictions over Jan. 2012 to Apr. 2022. Results are obtained for a dataset mirroring the data available to a forecaster at the 11th day of the month. The table reports *p*-values for the one-sided test of Diebold and Mariano (1995) with the small sample correction introduced by Harvey et al. (1997). Light grey cells indicate significance at 10%.

Table A1.3: Diebold-Mariano test results (horizon = $t + 1$)								
	Benchmarks		"Traditional"		ML tree-based		ML regression-based	
	AR	OLS	MS	QR	RF	XGB-T	MRF	XGB-L
AR		0.86	0.51	0.73	0.94	0.94	0.98	0.98
OLS	0.14		0.15	0.08	0.44	0.54	0.90	0.98
MS	0.49	0.85		0.65	0.80	0.83	0.94	0.93
QR	0.27	0.92	0.35		0.85	0.87	0.97	1.00
RF	0.06	0.56	0.20	0.15		0.71	0.95	0.92
GB-T	0.06	0.46	0.17	0.13	0.29		0.93	0.86
MRF	0.02	0.10	0.06	0.03	0.05	0.07		0.28
GB-L	0.02	0.02	0.07	0.00	0.08	0.14	0.72	

Notes: MS = Markov-switching; QR = quantile regression; RF = random forest; GB-T = gradient-boosting; MRF = Macroeconomic Random Forest, GB-L = linear gradient boosting. Tests are based on one-period-ahead out-of-sample predictions over Jan. 2012 to Apr. 2022. Results are obtained for a dataset mirroring the data available to a forecaster at the 11th day of the month. The table reports *p*-values for the one-sided test of Diebold and Mariano (1995) with the small sample correction introduced by Harvey et al. (1997). Light grey cells indicate significance at 10%.

Table A1.4: Model confidence set (Hansen et al., 2011) test results								
	Benchmarks		"Traditional"		ML tree-based		ML regression-based	
	AR	OLS	MS	QR	RF	XGB-T	MRF	XGB-L
$t - 2$	0.12	0.24	0.54	0.33	0.30	0.50	1.00	0.37
$t - 1$	0.11	0.45	0.18	0.30	0.83	0.45	1.00	0.82
t	0.26	1.00	0.63	0.62	0.37	0.25	0.94	0.89
$t + 1$	0.60	0.74	0.63	0.47	0.60	0.58	1.00	0.27

Notes: MS = Markov-switching; QR = quantile regression; RF = random forest; GB-T = gradient-boosting; MRF = Macroeconomic Random Forest, GB-L = linear gradient boosting. Tests are based on one-period-ahead out-of-sample predictions over Jan. 2012 to Apr. 2022. Results are obtained for a dataset mirroring the data available to a forecaster at the 11th day of the month. The table reports Results report *p*-values of the statistic T_{max} of Hansen et al. (2011) based on a squared error loss function. Grey cells indicate models in the 10% confidence set.

Annex 2: Data

Table A2.1. Description of data

	Number of series	Publication delay	Sources
<i>Monthly dependent variable</i>			
Global trade growth	1	8 weeks	Centraal Plan Bureau
<i>Trade indicators (total = 158)</i>			
PMI “new export orders” (manuf.)	29	1 day	IHS Markit
PMI “new export orders” (services)	29	1 day	IHS Markit
“New export orders” indices	5	1 day	NBS, ISM, CBI
Container throughput index	3	3 weeks	RWI/ISL
Port traffic	31	1-3 weeks	Thomson Reuters, port authorities
National trade statistics (values)	51	1-3 weeks	Thomson Reuters
Baltic dry indices (shipping costs)	6	Current	Thomson Reuters
Harpex (shipping costs)	1	1 week	Thomson Reuters
Truck traffic	3	Current	Destatis
<i>Broad macroeconomic outlook (total = 285)</i>			
Steel production (volumes)	69	3 weeks	Thomson Reuters
Semi-conductor billings	1	5 weeks	Thomson Reuters
US Tech Pulse index	1	3 weeks	Thomson Reuters
Industrial production	61	4-5 weeks	Thomson Reuters
Composite Leading indicator (CLI)	86	6 weeks	OECD
Retail sales	23	4-5 weeks	Thomson Reuters
Car registrations	35	4-5 weeks	OECD, Thomson Reuters
Business climate surveys	8	1 day	HIS Markit, Thomson Reuters
<i>Financial indicators and commodity prices (total = 93)</i>			
Stock market indices	29	Current	Thomson Reuters, S&P
Commodity prices	8	Current	Thomson Reuters, S&P
Exchange rates	56	1 day	BIS

Note: Publication delay is expressed in terms of delay after the end of the month, e.g. a 8-week delay for CPB data means that data for September 2022 is available around the 25th November 2022.

Annex 3: Alternative factor extraction methods

Given the potential shortcomings of PCA – notably in case of serial correlation – we test alternative approaches assuming factor dynamics as in equation A3.1. In this case, the number of shocks q (dimension of u_t below) can differ from the number r of factors.

$$(A3.1) \quad F_t = A_1 F_{t-1} + \dots + A_p F_{t-p} + B u_t, \quad u_t \sim i. i. d.$$

A first alternative technique is the 2-step estimator of Doz et al. (2011). In the first step, factors \hat{F}_t are estimated through PCA on the data X_t and a VAR of order p is estimated by OLS on the \hat{F}_t to estimate the A_i matrices. In a second step, the Kalman smoother starts with estimates of first step and yields new factor estimates. This procedure has been used in various nowcasting applications following the seminal work of Giannone et al. (2005) and Giannone et al. (2008).

A second alternative is the quasi-maximum likelihood (QML) of Doz et al. (2012) which iterates on the two-step approach. While the previous method is more suited for *exact* factor model – which assumes that the idiosyncratic components are mutually uncorrelated at all lags: $E[\xi_{i,t}, \xi_{j,s}] = 0, \forall i \neq j$ – the quasi-maximum likelihood suits for an *approximate* factor model where this assumption is relaxed. The idiosyncratic component ξ_t is an n -dimensional stationary process with mean zero and non-null covariance matrix and can notably follow an AR process as in equation A3.2.

$$(A3.2) \quad \xi_t = Z_1 \xi_{t-1} + \dots + Z_s \xi_{t-s} + e_t, \quad e_t \sim i. i. d.$$

A maximum likelihood estimation is obtained by iterating until convergence between the two steps of an expectation-maximization algorithm. In the first step, the expectation of the log-likelihood conditional on the data is computed using the parameters estimated at the previous iteration. In the second step, new parameters are re-estimated through the maximization of the expected log-likelihood. The initialization is done with the 2-step approach. This approach is suited for mixed frequency nowcasting as Banbura and Modugno (2014) adapted it to cope with arbitrary patterns of missing data. It has been widely used in nowcasting (e.g. New York Fed's US nowcasting) including for trade (Guichard and Rusticelli, 2011; Martinez-Martin and Rusticelli, 2021).

The third alternative relies on Forni et al. (2005)'s generalized dynamic factor model (GDFM) where the estimation is performed in the frequency domain. For the estimation of common factors, time observations are weighted according to their signal-to-noise ratio. The method starts by estimating the density spectral matrix of the common components ($\Lambda \cdot F_t$) and of the idiosyncratic components ξ_t . Based on this, the inverse Fourier transformation gives the time-domain autocovariances $\Gamma(k)$ for both common and idiosyncratic components. The estimation of factors is then performed by finding the r linear combinations of X_t that maximise the contemporaneous covariance $\Gamma(0)$. This later method has been used for nowcasting trade in Charles and Darne (2022).

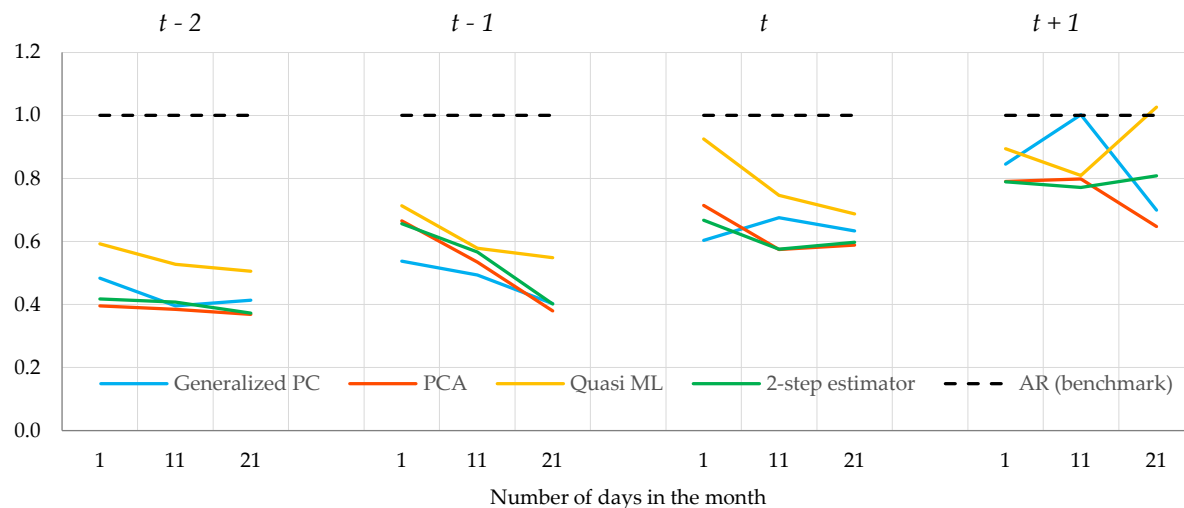
We turn to comparing these alternative factor extraction techniques with PCA. As in the main text, we explore the out-of-sample predictive performances over Jan. 2012 to Apr. 2022. To mimic real-time set-up, factor extraction is performed at any date T using the dataset X_t^T of variables pre-selected through LARS. The four different factor extraction techniques (*method* = PCA, 2-step, QML, Generalized PC) are applied to X_t^T to extract factors $F_t^{T,method}$. These factors are then used as independent variables in an OLS regression. As in the main text, for robustness, we test across different horizons (-2 , -1 , 0 , and $+1$) and across different datasets corresponding to data available at different days of the month (1st, 11th, and 21st). Results are reported in **Figure A3.1**: each panel refers to a given horizon (from $t - 2$ to $t + 1$) with the day of the month (1, 11, or 21) on the x-axis. The y-axis represents the accuracy, measured by the out-of-sample RMSE over Jan. 2012 to Apr. 2022, relative to the AR benchmark (black dotted line).

Results in **Figure A3.1** report overall broadly similar accuracy across the different techniques, consistently across all horizons and days of the month. If anything, PCA is generally on the upper bound in terms of accuracy (lower RMSFE) while quasi maximum likelihood stands generally on the lower bound.³² But differences in predictive accuracy are generally not significant. This is in line with the literature, notably Kapetanios and Marcellino (2009) finding

³² We interpret these differences, albeit small, as a possible consequence of the design of the dataset. Quasi-ML is intended to deal with missing data; here, as the dataset is balanced, this technique might lose some of its interest. In addition, while some dynamics in the factors can be expected *a priori*, the vertical realignment might impair such dynamics *a posteriori*.

similar performances across factor extraction techniques. In the end, we apply Occam’s razor (if different methods have equivalent performance, pick the simplest) and use PCA.

Figure A3.1. Relative accuracy of factor extraction techniques



Notes: Accuracy is measured by the out-of-sample RMSE over Jan. 2017 - April 2022. Performances are presented relative to the AR benchmark (black dotted line). Regression performed through OLS using factors as independent variables. "Generalized PC" refers to the one-sided method developed in Forni et al. (2005), "2-step estimator" refers to Doz et al. (2011), and "Quasi ML" to Doz et al. (2012).

Annex 4: Best-performing pre-selection with alternative regressions

In the main part of the paper, the choice of the pre-selection technique is based on PCA and OLS. The interest to fix factor extraction and regression (respectively steps *ii* and *iii*) is that the difference in accuracy would only come from the pre-selection technique (step *i*). However, this can result in sub-optimality as it is not guaranteed that LARS, which is found to be the best-performing technique for PCA and OLS, would still be the best-performing for different methods in steps *ii* and *iii*. This annex checks whether LARS remains the best-performing pre-selection across different regression techniques.³³

LARS is generally found to be the best-performing techniques across all regression techniques. **Figure A4.1** reports the accuracy across pre-selection techniques measured by the out-of-sample RMSFE over Jan. 2012 to Apr. 2022, relative to the no pre-selection benchmark (dark grey line). Each panel represents a different horizon, from $t - 2$ to $t + 1$. Accuracy is reported relatively for each of the regression techniques (on the x-axis). Results are reported for data at the 11th day.³⁴ The LARS is in almost all cases the best-performing technique, with accuracy gains relative to the no pre-selection benchmark reaching up to 40% at $t - 2$. At some specific horizons and regression techniques, LARS might not be the best-performing – for example, OLS at t – but in those cases, the difference in accuracy between LARS and the other best-performing technique is generally found to be small and non-significant. On average, LARS also exhibits low variation of performances across regression techniques and horizons: this is in contrast for instance with the t-stat-based technique which can be the best-performing technique (for example for random forest at t) while also resulting in the worst performances (for example for gradient boosting at $t - 1$).

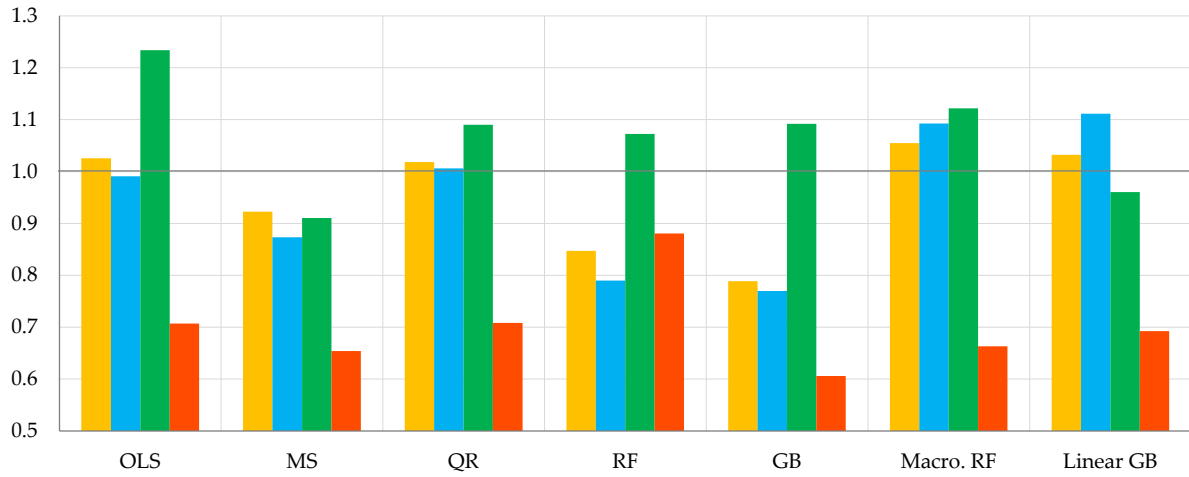
³³ For simplicity, we fix PCA as factor extraction. As shown in **Annex 3**, changing factor extraction technique results in relatively close performances.

³⁴ Results at 1st and 21st days of the month yield similar conclusions. Results are reported for a pre-selection of 60 variables; other numbers of variables yield also similar conclusions.

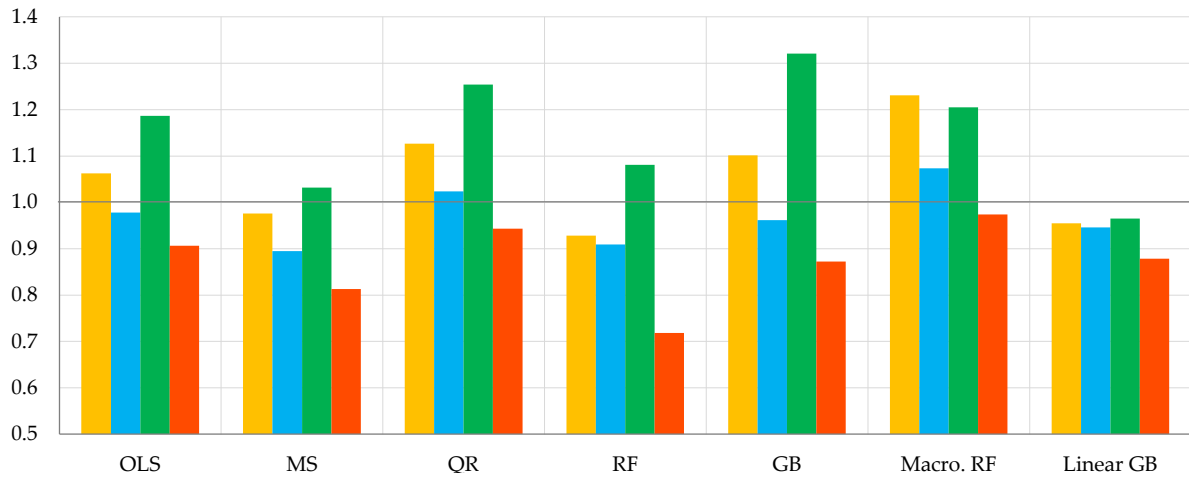
Figure A4.1. Accuracy relative to no pre-selection

■ SIS ■ Iterated BMA ■ T-stat based ■ LARS

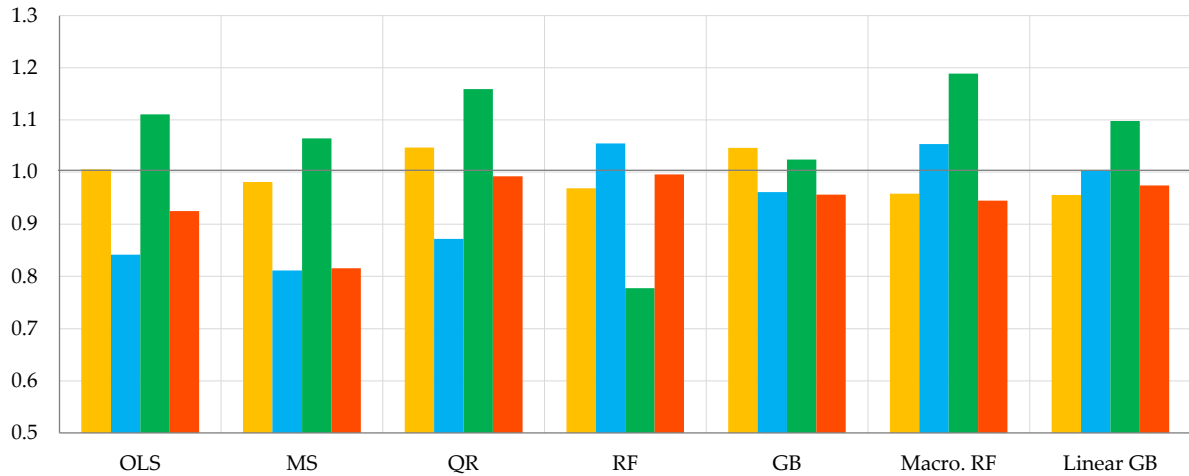
Panel A. $t - 2$



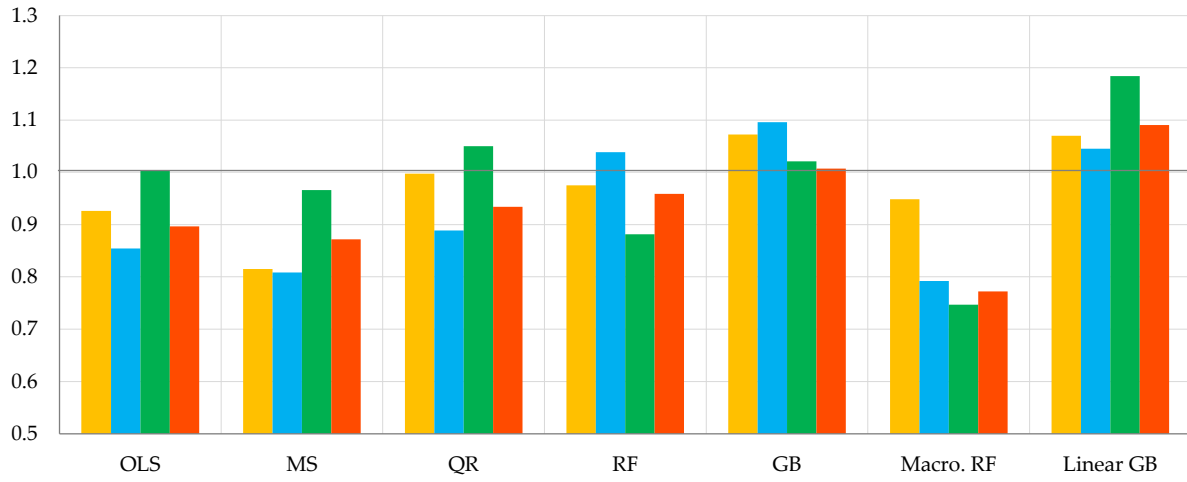
Panel B. $t - 1$



Panel C. t



Panel D. $t+1$



Notes: Accuracy is measured by the out-of-sample RMSE over Jan. 2012 - April 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, and factors extracted through PCA on the pre-selected set. MS = Markov-switching; QR = Quantil regression; RF = Random forest; GB = Gradient boosting; Macro. RF = Macroeconomic random forest, Linear GB = Gradient linear boosting