Nowcasting GDP using machine learning methods

Dennis Kant

Andreas Pick*
Erasmus University Rotterdam
and De Nederlandsche Bank

Jasper de Winter
De Nederlandsche Bank
and Vrije Universiteit Amsterdam

May 18, 2022

Abstract

This paper compares the ability of several econometric and machine learning methods to nowcast GDP in (pseudo) real-time. The analysis takes the example of Dutch GDP over the years 1992-2018 using a broad data set of monthly indicators. It discusses the forecast accuracy but also analyzes the use of information from the large data set of regressors. We find that the random forest forecast provides the most accurate nowcasts while using the different variables in a relative stable and equal manner.

1 Introduction

GDP is published several weeks after the end of a quarter and initial releases are subject to substantial uncertainty. This fact led to the development of models to predict GDP of the past, current, and future quarters, a practice referred to as "nowcasting". Such nowcasting models use dimension reduction techniques to nowcast GDP from a large number of macroeconomic and financial variables.

This paper explores the nowcasting precision of econometrics and machine learning methods in comparison to the most commonly used nowcasting method, the dynamic factor model. We consider nowcasts, that is, predictions two quarters ahead to two months backwards, using the example of Dutch GDP with a broad data set of 83 macroeconomic and financial variables. We consider the performance of several methods: factor models,

*Corresponing author: andreas.pick@cantab.net. The research for this paper was conducted while the first author was at De Nederlandsche Bank. The opinions expressed are those of the authors and do not represent the views of DNB or the EuroSystem.

regularization methods, random subspace methods, and the random forest. We separately evaluate their performance over different periods and different states of the economy.

For policy purposes, the interpretation of nowcasts is often of interest and we, therefore, investigate how the different methods use the information from the large data set. This is relatively straightforward for the factor, regularization, and random subspace models. The random forest, in contrast, is highly nonlinear and we use Shapley values to calculate the importance of the different variables.

Our findings suggest that, until the financial crisis, the nowcasts of the dynamic factor model were as accurate as those of the machine learning models. Since the financial crisis, however, machine learning methods, in particular the random forest, have been more accurate and this is true across various forecast horizons. When considering the variables that the different models rely most heavily on, it emerges that the dynamic factor model has shifted away from surveys and relies more heavily on production, sales, and financial variables. The random forest, in contrast, has a fairly even and stable use of variables in all categories.

The benchmark model in our analysis is the dynamic factor model as this model has been widely used at central banks to nowcast GDP (Giannone et al., 2008; Bańbura and Rünstler, 2011; Jansen et al., 2016; Hindrayanto et al., 2016; Bok et al., 2018). Another popular class of models that we also consider are MIDAS based models (Marcellino and Schumacher, 2010; Kuzin et al., 2011; Foroni and Marcellino, 2014).

We compare these models to a range of models that are sometimes referred to as machine learning models. The first set of models are regularization methods, which estimate linear models where the parameters are subject to penalization terms. The nature of the penalization distinguishes the different models. We consider the LASSO of Tibshirani (1996) and the elastic net of Zou and Hastie (2005). We also considered other penalization models, namely the adaptive LASSO and ridge regression. However, the results were inferior to the LASSO and elastic net and for brevity we omit them from the discussion.

Next, we consider the random subspace methods of Elliott et al. (2013) and Boot and Nibbering (2019). These methods exploit the fact that model averaging tends to reduce the mean square forecast error by making a large number of predictions, where the model for each prediction combines subsets of the variables in the data set in a random manner. The predictions are then averaged to yield the final forecast.

Finally, we consider the random forest of Breiman (2001), which combines the model averaging feature of the random subspace methods with the nonlinear modeling inherent in regression trees. A downside of the random forest is that its nonlinearity complicates the interpretation of the role of the different variables in the forecasts and we work with Shapley values as discussed by Štrumbelj and Kononenko (2014) and Lundberg and Lee (2017) in our analysis of the importance of the variables in the random forest.

We also consider averaging the forecasts from the different models using a range of weights that have been discussed by Elliott and Timmermann (2016). Model averaging has been shown to be able to reduce the mean square forecast error as it can reduce the variance of the resulting forecast. However, the weights with which the forecasts are averaged introduce uncertainty, which can negatively impact the forecast precision. In our application, average forecasts have a relatively high precision but tend not to beat the best model, the random forest, over all horizons.

Our paper relates to the growing literature on forecasting with machine learning methods in macroeconomics. Richardson et al. (2018) evaluate nowcasts of New Zealand's GDP from a number of machine learning models using quarterly data. They find that machine learning methods, in particular support vector machines, improve over their benchmark, the univariate autoregressive model. They do, however, not include the dynamic factor model or random forest, which are the leading methods in our analysis. Jönsson (2020) evaluates the nowcasting performance of the nearest neighbour algorithm for Swedish GDP and finds that it compares well to standard linear sentiment index models that are commonly used in Sweden. Yoon (2022) nowcasts Japanese GDP using boosted trees and random forests and finds that these outperform nowcasts of the Bank of Japan and the International Monetary Fund. Finally, Jardet and Meunier (2022) nowcast world GDP using factor-augmented and LASSO MIDAS models. They find that factor-augmented models deliver the most precise nowcasts. An important difference between these papers and our work is that their analysis is restricted to average forecast performance measures. In contrast, the importance of the different variables in the nowcasts is not considered. In a policy context, however, such interpretations are important and for this reason we put particular emphasis on the role of the variables in our nowcasts.

Machine learning methods have also been used to predict other macroeconomic variables. Medeiros et al. (2021) use a range of methods to predict U.S. inflation between 1 and 12 month into the future. Similar to our findings, their results suggest that the random forest is the most precise method to forecast inflation. Gogas et al. (2021) predict euro area unemployment using machine learning methods and similarly find that the random forest delivers the most precise unemployment forecasts. Finally, Maehashi and Shintani (2020) forecast seven Japanese macroeconomic time series (but not GDP). Comparable to our results, they find that machine learning models tend to offer more precise forecasts than traditional time series models and that ensemble methods improve over individual models.

In the next section, we introduce the models that we use for nowcasting and discuss the interpretation of the role the different variables play in the nowcasts. Details of the data are in Section 3 and the results are discussed in Section 4. Finally, Section 5 concludes.

2 Nowcasting models

In this section, we give a brief overview of the different methods we use in this paper and the parameter choices we make. We relegate a more formal description to the Online Appendix.

We compare the nowcasts produced by these methods to those from two benchmark models: the prevailing mean model, which assumes that the mean in the estimation sample provides the best forecast, and an autoregressive model with lag length chosen by AIC. All parameters are re-estimated for each nowcast using an expanding window.

Dynamic factor model

The dynamic factor model is widely used by policy institutions. In order to facilitate comparison to these models, we use the specification in the literature (Giannone et al., 2008; Bańbura and Rünstler, 2011; Jansen et al., 2016; Bok et al., 2018). The dynamic factor model summarises the large set of regressors in a small number of factors. The factors and their loadings are estimated using principal components, and their dynamics subsequently estimated in a VAR(p). This procedure simplifies the analysis relative to estimation in state space models as the parameters of the VAR can be estimated via OLS.

The specification depends on the number of factors and the number of lags in the VAR. In line with the literature Kuzin et al. (2013); Jansen et al. (2016) we produce nowcasts for VAR models with between one and six lags for between one and six factors and then use an equal weighted average as the final nowcast of the dynamic factor model

Mixed-data sampling factor-augmented model

The MIDAS model of Ghysels et al. (2007) has been adapted for nowcasting by Marcellino and Schumacher (2010). In the factor-augmented MIDAS model, factors are extracted at a monthly frequency. These are then used to model the quarterly series. The monthly series can be averaged using the exponential Almon lag. Alternatively, in the unrestricted MIDAS model, the monthly series are included separately using skip sampling. In our analysis, the unrestricted MIDAS model produced the more precise nowcasts and we will therefore concentrate on this weighting scheme. For consistency with the dynamic factor model, we use the same number of factors and construct the factors using between one to six lags of the data. We then average the different forecasts. For brevity, we will refer to this factor-augmented MIDAS model as the MIDAS model.

Regularization techniques

We use the least absolute shrinkage and selection operator (LASSO) and the elastic net in this paper. We also obtained results for ridge regression and adaptive LASSO. However, the forecasts from the latter two were strictly dominated by those from LASSO and elastic net and, for brevity, we therefore omit these results.

The LASSO of Tibshirani (1996) performs both regularization and variable selection by imposing an ℓ_1 penalty in the estimation of the coefficients. The elastic net of Zou and Hastie (2005) imposes a combination of ℓ_1 and ℓ_2 penalties. Similar to the LASSO, the ℓ_1 norm selects parameters by shrinking some to zero but through the use of the ℓ_2 norm it also shrinks the remaining coefficients towards zero.

The amount of shrinkage in the LASSO is determined by a scalar parameter, λ. We determine this parameter via cross-validation. Importantly, the cross-validation is done for each forecast separately and uses only data in the respective estimation sample, and this is the case for all empirical methods that use cross-validation. Next to the shrinkage parameter, the elastic net has a second parameter that determines the relative importance of the ℓ_1 and ℓ_2 penalties. Again, we determine these parameters via cross-validation.

Random subspace regression

Random subspace methods encompass methods that reduce the regressor space by averaging forecasts from either random combinations or random subsets of the data. The idea of randomly selecting smaller models and averaging (with equal weights) their forecast is based on complete subset regression of Elliott et al. (2013). They build on the idea of model averaging to combine forecasts obtained from all possible combinations of smaller linear models that can be produced from a large data set. However, the number of possible combinations can quickly become prohibitively large. A solution is to take a smaller number, say R, of randomly chosen subsets of regressors. Boot and Nibbering (2019) show that this approximates the complete subset regression for finite R, such as $R = 1000$.

A tuning parameter of this method is the size of each regressor subset, k. Theoretical results by Boot and Nibbering (2019) suggest that k should be chosen relatively large at about 30. The experience of Pick and Carpay (2022) suggests that smaller k can deliver more precise forecasts. We initially experimented with different choices of k up to 30 and our experience confirms that smaller choices of k deliver better nowcasts. As a result, we average nowcasts over those obtained using $k = 2, 3, 4, 5$.

An alternative to selecting regressors would be to combine the regressors with random weights. Boot and Nibbering (2019) discuss this option and call it 'random projection'. The large regressor set is combined using a random weighting matrix to yield k combined regressors. For Gaussian random projections, the weights are independently drawn from a standard normal distribution. Multiple realizations of the weights are drawn and the resulting forecasts are averaged. Again, the choice variable k needs to be determined and as above we average nowcasts over those obtained using $k = 2, 3, 4, 5$.

Other weighting schemes are possible, for example, the weights used by Guhaniyogi and Dunson (2015). The analysis of Pick and Carpay (2022), however, suggests that random subset selection and random projection deliver superior forecast performances and we therefore limit our attention to these two methods.

Random forest

The random forest forecast averages the forecasts of multiple decision trees. To grow a decision tree, the space of predictor values is partitioned with the aim of minimizing the in-sample squared error. At each partition, the algorithm chooses a split based on one of the predictors that realizes the largest decrease in squared error.

Trees are designed to have a high degree of independence of each other by randomly drawing a subset of predictor variables and a subset of observations to grow any given tree. Averaging the forecasts from the trees in the random forest therefore minimizes the variance of the average forecast.

In order to reduce overfitting, each estimation sample is divided in a training and a validation set. The share of the data in the training set, ω, in all estimation samples is varied with $\omega = 0.6, 0.7, 0.8, 0.9$ and we average forecasts over the results from the ω. Within each training set size, ω, we cross-validate the number of skip-sampled predictors to split the tree, where the possible values of the number of predictors is $k = 1, 2, \ldots, 249$. We use the prediction of the 400 trees in each random forest.

2.1 Nowcast combinations

Given that the models above have distinct ways of incorporating the information of the monthly indicators, combining the forecasts could be beneficial. Combining forecasts from different sources has a long track record in the forecasting literature (Timmermann, 2006). We start by examining whether the models have marginal value in comparison to the best model and subsequently investigate forecast combination strategies.

The marginal value of a model can be established via an encompassing test, where the marginal value of a model relative to the model with the lowest RMSFE is assessed by estimating the parameters of the following regression model,

$$y_{t+h} = \alpha + \delta \hat{y}_{a,t+h|t} + (1 - \delta)\hat{y}_{b,t+h|t} + \varepsilon_{t+h}$$

Table 1
Hyperparameter choices

Method	hyperparam.	choice
DFM	no. factors	average over $q = 1, 2, \ldots, 6$ factors
	lag length	average over $p = 1, 2, \ldots, 6$ lags
MIDAS	no. factors	average over $q = 1, 2, \ldots, 6$ factors
	lag length	average over $p = 1, 2, \ldots, 6$ lags
LASSO	weight ℓ_1 penalty	cross-validation
EN	weight ℓ_1 penalty	cross-validation
	weight ℓ_1 vs ℓ_2 penalty	cross-validation
RS	size models	average over $k = 2, 3, 4, 5$
	no. of models	1000
RP	size models	average over $k = 2, 3, 4, 5$
	no. of models	1000
RF	share training set	average over $\omega = 0.6, 0.7, 0.8, 0.9$
	no. of regressors per tree	cross-validate per ω for $k = 1, 2, \ldots, 249$
	no. trees	400

The methods are: DFM the dynamic factor model, MIDAS the mixed-data sampling factor-augmented model, LASSO the least absolute shrinkage and selection operator, EN the elastic net, RS the random subset selection, RP the random projection, RF the random forest.

where $\hat{y}_{a,t+h|t}$ and $\hat{y}_{b,t+h|t}$ are two of our forecasts, subscript a denotes the alternative forecast and b the most precise forecast *ex post*, and for the weights we have that $\delta \in [0, 1]$. The alternative model has marginal value next to the best model if, significantly, $\delta > 0$.

The first forecast combination would then be to use δ as the model weight. As we have multiple forecasts at our disposal, we consider different strategies.

Combining more than two forecasts requires weights for each of the forecasts. A simple solution is to give the forecasts equal weight, which turns out to be a difficult to beat benchmark (Timmermann, 2006). An advantage of equal weighted forecast combinations is that fixed weights avoids estimation uncertainty that would translate into forecast uncertainty. The downside, however, is that if the unknown optimal weights are far from equal weights, the forecast combination may suffer. We therefore also consider ways to estimate the weights.

The first method to estimate the forecast weights is to use weights that are inversely proportional to the MSFE. We measure the cumulative square forecast error, $\upsilon^2_{j,t|t-h}$, and calculate the weights as

$$w_{j,t} = \frac{(\upsilon_{j,t|t-h})^{-2}}{\sum_{j=1}^{m}(\upsilon_{j,t|t-h})^{-2}}$$

where m is the number of forecasts to combine. We calculate the cumulative error, $\upsilon_{j,t|t-h}$, using either an expanding window or a rolling window of ten quarters. Prior to the first forecast, no weights can be determined and we take an equally weighted average as the first forecast combination.

The weights above do not address potential biases of the forecasts. If biases are suspected, Granger and Ramanathan (1984) suggest estimating weights in a linear regression

$$y_{t+h|t} = \beta_0 + \sum_{j=1}^{m} \beta_j \hat{y}_{j,t+h|t} + \epsilon_{t+h}$$

where the estimated coefficients are then used as forecast weights in addition to the intercept that estimates the bias. Again, we estimate the coefficient using an expanding and a rolling window of size 40. For the first 40 forecasts we use equal weights.

2.2 Interpreting nowcasts

In policy environments, such as central banks, the interpretation of nowcasts is important. We will therefore illustrate the contributions of the underlying time series to the nowcasts over time. For the dynamic factor model, the weights that are implicitly assigned to predictors in a model are determined using the methods of Koopman and Harvey (2003) and Bańbura and Rünstler (2011).

The MIDAS, LASSO, elastic net, random subset regression, and random projection are all linear in the predictors. This means that it is straightforward to extract the contributions of the time series to the forecasts. For the shrinkage methods, the contributions of the time series to the forecasts follow from the linear relationship the selected variables have with GDP. The random subspace methods also fall into the class of linear models conditional on the selection regressors and the projection matrices.

The random forest, in contrast, is a highly nonlinear model, which makes interpreting the role of different predictors considerably more complex. We use the concept of Shapley values (Shapley, 1953) to interpret the role of predictors, which has been developed further by Štrumbelj and Kononenko (2014) and Lundberg and Lee (2017).

The Shapley value measures the average difference in the loss of trees that include the variable in question to the loss of trees that do not include that variable. Denote the loss of a tree that includes a given variable i by $L(\mathcal{S} \cup \{i\})$ and that of another tree with the same variables except variable i by $L(\mathcal{S})$, where $\mathcal{S}$ is the set of variables in the tree except variable i, and $\mathcal{S} \subseteq \mathcal{F}$ with $\mathcal{F}$ denoting the complete set of variables. As the effect of variable i likely depends on the other variables in the tree, the loss differential is computed for all possible variables in a tree. The contribution of variable

i, ϕ_i, is then computed as the weighted average of the differences

$$\phi_i = \frac{1}{N_M} \sum_{\mathcal{S} \subseteq \mathcal{F}} [L(\mathcal{S} \cup \{i\}) - L(S)]$$

where N_M is the number of possible combinations of variables in trees excluding variable i. With many predictors, this procedure is computationally burdensome and we therefore use the approximation proposed by Štrumbelj and Kononenko (2014).

3 Data

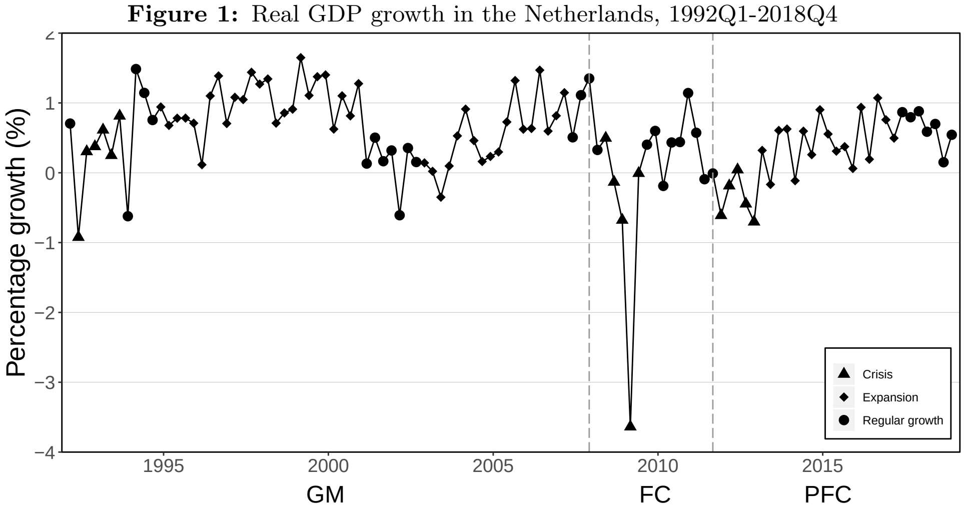
The data set consists of 83 monthly time series and quarterly GDP that were downloaded on the 26$^{\text{th}}$ of March 2019. The statistical monthly information set reflects the public knowledge at the end of the month. The time series were obtained from Statistics Netherlands, the central banks of Belgium and the Netherlands, Datastream, the European central bank, Eurostat, and the Hamburg Institute of International Economics.

The series can be grouped into five categories. The first category is hard, quantitative information on production and sales, such as industrial production, car sales, retail sales, exports, imports, and unemployment. The second category is soft, qualitative information on expectations derived from surveys among consumers, retailers, and firms. The third category contains financial variables, both quantities (monetary aggregates) and prices (interest rates and stock prices), which determine financing conditions for firms and consumers. Moreover, financial market prices partly reflect financial market expectations on output developments in the near future. The fourth category contains information on prices, i.e. consumer prices, producer prices, the housing price, and commodity prices. The fifth category contains some miscellaneous series, i.e. bankruptcies and issued vehicle registrations.

Most monthly data are seasonally (and calendar effects) adjusted at the source, except for prices and financial variables. If data are not seasonally adjusted we apply the US Census' X12-method. All monthly series are made stationary by differencing, log-differencing or double log-differencing (in the case of prices). Moreover, all variables are standardized by subtracting the mean and dividing by the standard deviation. This normalization is necessary to avoid overweighting of high-variance series in the extraction of common factors. Details can be found in Table A.1 in the Appendix.

All monthly indicator series start in January 1985, while the quarterly GDP series start in the first quarter of 1985. The forecast evaluation period runs from 1992Q1 to 2018Q4. We produce pseudo-out-of sample forecasts for the observations in the forecast period using an expanding estimation sample, that is, model selection and the estimation of the parameters in the

Figure 1: Real GDP growth in the Netherlands, 1992Q1-2018Q4



Note: Three time periods are depicted: GM: Great Moderation (1992Q1-2007Q4), FC: Financial Crisis (2008Q1-2011Q3) and PFC: Post-Financial Crisis (2011Q4-2018Q4). Economic crises, expansions and regular growth are depicted in the graph.

models relies on the data of the estimation sample and, after each forecast, the estimation sample is expanded to include the additional observations.

We analyze the performance of the model during periods of economic contraction as defined by the Centre for Economic Policy Research (CEPR), normal growth, and expansion. We use the dating of crises by the CEPR in the euro zone as a proxy for crises in the Netherlands. The CEPR identifies three crises in the evaluation period: 1992Q2-1993Q3, 2008Q2-2009Q2, and 2011Q4-2013Q1. Similarly, based on the findings of the International Monetary Fund (IMF), periods of economic expansion are identified as upward trends in annual real GDP growth in the time periods 1994Q4-2000Q4, 2002Q4-2007Q1 and 2013Q1-2017Q1. GDP growth is plotted in Figure 1. In 17 quarters the economy contracted, in 31 quarters the economy grew moderately and in 60 quarters the economy expanded.

The ragged edge

The different indicators are published with varying delays. Generally, the hard variables are accompanied with a significant publication delay, whereas soft variables (surveys) are published on a more timely basis. The varying availability of the indicators is commonly referred to as the "ragged-edge" (Wallis, 1986).

Our modelling strategy needs to account for data availability at the moment when the nowcast would be made in real time. We take an approach here that is akin to the direct forecasting approach: rather than predict variables that are still unreleased, we nowcast GDP using the latest available data. This has been called "vertical realignment" by Marcellino and Schumacher (2010). We have experimented with other methods to deal with the ragged-edge, such as using univariate methods to predict the missing observations. However, these methods, while computationally much more costly, did not lead to improvements in nowcast precision and, for brevity, we therefore omit these results.

We employ a pseudo real-time design, which takes data publication delays into account, but ignores the possibility of data revisions for GDP and some indicators, such as retail trade. The latter might implies that we overestimate the forecasting accuracy of statistical models. However, it is also quite likely that the effects of data revisions on the final forecast will largely cancel out because statistical methods typically attempt to eliminate noise from the process by either extracting factors from a large data set or pooling large numbers of indicator-based forecasts. For example, using real-time data vintages for Germany, Schumacher and Breitung (2008) did not find any clear impact of data revisions on the forecast errors of factor models. Moreover, the effect on the relative performances of models, which is the main focus of this paper, is likely to be quite small (see Bernanke and Boivin, 2003).

4 Results

4.1 Model Performance

We report the forecast performance as measured by the root mean square forecast error (RMSFE) of each method relative that of the prevailing mean forecast

$$\text{relative RMSFE} = \frac{\sqrt{\frac{1}{T_f}\sum_{t=1}^{T_f}(y_t - \hat{y}^{(a)}_{t|t-h})^2}}{\sqrt{\frac{1}{T_f}\sum_{t=1}^{T_f}(y_t - \hat{y}^{(pm)}_{t|t-h})^2}}$$

where superscript a denotes the respective forecasts, superscript pm the prevailing mean forecast, and T_f denotes the number of forecasts. For the prevailing mean, we report the levels of the RMSFE.

We also use the test of Diebold and Mariano (1995) to evaluate significance of the forecasts of the different models against those from the dynamic factor model. The Diebold-Mariano test results should be interpreted with caution since we are using an expanding window, which could imply that the assumptions of the Diebold-Mariano test are violated. Significance can therefore be interpreted as a sign of improved forecast performance relative to insignificant forecasts but not necessarily at the stated significance level.

Table 2 reports the RMSFE over the entire forecast sample over four forecast horizons: two-quarter ahead, one-quarter ahead, and nowcasting, where each is the average of the three months of nowcasts in this quarter,

Table 2
Average forecast precision over the period 1992Q1-2018Q4

Model	Benchmark			Alternative					
	PM	AR	DFM	MIDAS	LASSO	EN	RS	RP	RF
2Q ahead	*0.69*	1.01	0.99	0.99	0.98	0.98	0.97	1.01	0.94^{*}
1Q ahead	*0.69*	1.00	0.97	1.01	0.97	0.95	0.96	0.99	0.92^{*}
Nowcast	*0.69*	0.96	0.90	0.99	0.88	0.88	0.90	0.92	0.88^{**}
Backcast	*0.69*	0.97	0.77	0.89	0.82	0.83	0.88	0.92^{**}	0.83

Note: The table reports the results over the entire forecast sample. For the prevailing mean (PM) the entries (in italics) refer to the level of the RMSFE; for all other models the entries refer to the RMSFE relative to the RMSFE of the PM model. Grey cells indicate the model with the lowest RMSFE. AR denotes the autoregressive model, DFM the dynamic factor model, MIDAS the mixed-data sampling factor-augmented model, LASSO the least absolute shrinkage and selection operator, EN the elastic net, RS the random subset selection, RP the random projection, RF the random forest. *, **, *** indicate statistical significance at 10, 5 and 1 percent levels in a two-sided Diebold-Mariano test relative to the DFM (the significance levels should be interpreted with caution due to the expanding window estimation).

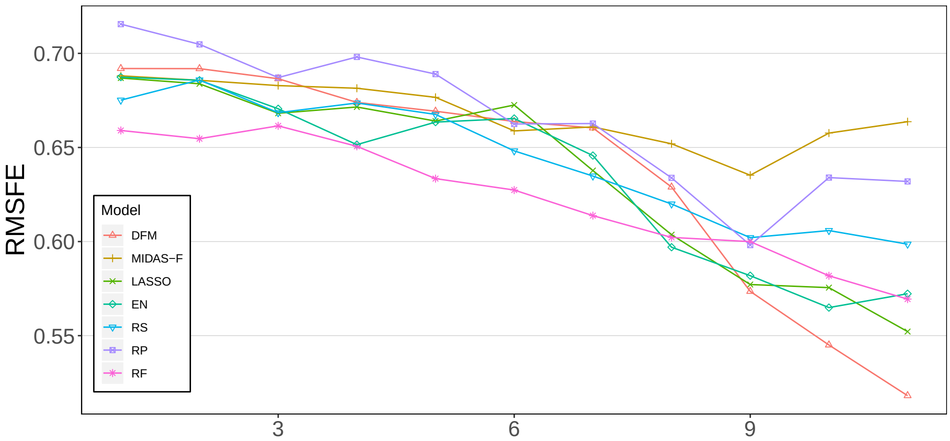
and backcasts, which we do for two months after the quarter in question. The table shows that in the short-term forecasts and nowcasts, the random forest provides the most precise predictions. The Diebold-Mariano test is statistically significant. The advantage of the random forest over the dynamic factor model decreases with the forecast horizon and the dynamic factor model becomes more precise when backcasting.

The AR, MIDAS model, and random projection are generally less precise than the dynamic factor model. The LASSO, elastic net, and random subset regression are roughly on par with the dynamic factor model when forecasting and nowcasting but less precise when backcasting. This suggests that over the entire forecast sample, the non-linearity of the random forest does well when fore- and nowcasting. Other methods that rely more on linear combinations of the data cannot substantially improve on the dynamic factor model. For backcasting, the dynamic factor model clearly is the most precise approach.

Figure 2 shows the RMSFE at monthly frequency, where 1 to 3 are two-month ahead forecasts, 4 to 6 one-month ahead, 7 to 9 are nowcasts and 10 and 11 backcasts. As expected, the RMSFE generally decreases as more information becomes available. The random forest has consistently lower RMSFE then the other methods for the forecasting months. For the first nowcasting month it is also the most precise but then the improvements to the dynamic factor model imply that it becomes more precise. In the backcasting months the dynamic factor model is the most precise. Of the other models, random subset regression remains relatively close to the random forest forecasts throughout, whereas MIDAS and random projection remain at the upper level of the RMSFE range.

The forecast sample contains periods of different nature, such as the

Figure 2: RMSFE over varying forecast horizons



Note: The horizontal axis denotes the months at which the forecast is made. 1 to 3 are two-month ahead forecasts, 4 to 6 one-month ahead, 7 to 9 are nowcasts and 10 and 11 backcasts. For the methods see the footnote of Table 2.

Table 3
Forecast precision evaluated over different subperiods

Model	Benchmark		Alternative						
	PM	AR	DFM	MIDAS	LASSO	EN	RS	RP	RF
GM ($N = 64$)									
2Q ahead	*0.56*	1.01	0.96	0.97	0.94	0.95	0.97	0.98	0.97
1Q ahead	*0.55*	1.00	0.94	1.03	0.98	0.96	0.96	0.97	0.93
Nowcast	*0.55*	0.99	0.93	0.99	0.96	0.96	0.92	0.97	0.89
Backcast	*0.55*	0.99	0.86	0.97	0.97	0.94	0.94	1.02	0.87
FC ($N = 15$)									
2Q ahead	*1.27*	1.00	1.02	0.97	0.99	0.98	0.99	1.02	0.93
1Q ahead	*1.27*	1.00	1.01	0.99	0.97	0.97	0.98	1.01	0.92
Nowcast	*1.26*	0.96	0.87	1.01	0.81	0.82	0.89	0.87	0.88
Backcast	*1.26*	0.97	0.66	0.83	0.65	0.67	0.81	0.82	0.82
PFC ($N = 29$)									
2Q ahead	*0.51*	1.00	1.04	1.15	1.06	1.07	1.03	1.14	0.95
1Q ahead	*0.51*	0.98	1.01	1.11	1.00	1.00	0.99	1.08	0.92
Nowcast	*0.51*	0.92	1.03	1.03	0.96	0.96	0.93	1.03	0.89
Backcast	*0.51*	0.94	0.99	0.97	0.88	0.95	0.91	0.96	0.80

GM refers to the Great Moderation period, FC to the financial crisis, and PFC to the Post Financial Crisis period. For further details see the footnote of Table 2.

Table 4
Forecast precision over various economic states

Model	Benchmark			Alternative					
Period	PM	AR	DFM	MIDAS	LASSO	EN	RS	RP	RF
Crisis ($N = 17$)									
2Q ahead	*1.35*	1.00	0.96	0.95	0.95	0.95	0.94	0.99	0.92
1Q ahead	*1.35*	1.00	0.90	0.95	0.86	0.85	0.90	0.93	0.89
Nowcast	*1.34*	0.95	0.79	0.93	0.76	0.77	0.83	0.81	0.85
Backcast	*1.33*	0.96	0.58	0.78	0.67	0.68	0.80	0.83	0.78
Moderate ($N = 31$)									
2Q ahead	*0.55*	1.01	1.06	1.08	1.05	1.05	1.04	1.02	0.96
1Q ahead	*0.55*	1.00	1.10	1.19	1.16	1.14	1.08	1.10	0.96
Nowcast	*0.55*	0.99	1.10	1.17	1.05	1.04	1.03	1.07	0.92
Backcast	*0.54*	0.99	1.04	1.11	0.99	0.98	0.99	1.04	0.91
Expansion ($N = 60$)									
2Q ahead	*0.44*	1.01	1.03	1.03	0.98	1.00	0.99	1.06	0.99
1Q ahead	*0.44*	1.00	1.02	0.99	1.06	1.04	1.00	1.03	0.96
Nowcast	*0.43*	0.98	0.99	0.98	1.01	1.02	0.96	1.03	0.92
Backcast	*0.43*	0.98	0.95	0.96	1.02	1.03	0.97	1.06	0.90

See footnote of Table 2.

Great Moderation and the Financial Crisis. In order to evaluate the robustness of our results, Table 3 reports the (relative) RMSFE over three subperiods: the Great Moderation, Financial Crisis, and the Post-Financial Crisis period. During the Great Moderation, the random forest is the most precise method for the nowcast and the one quarter ahead forecast. For the two period ahead forecast, however, the LASSO is more precise and also the elastic net is more precise than the random forest. The dynamic factor model produces the most precise backcasts. These methods also tend to beat both, the prevailing mean and the AR model, for all horizons. The MIDAS and random projection fail to beat the prevailing mean benchmark for at least one horizon.

During the Financial Crisis, the prevailing mean forecast is substantially less precise in absolute terms. Still, the dynamic factor model and random projection fail to deliver more precise one and two quarter ahead forecasts than this benchmark. In contrast, the random forest, LASSO, and the elastic net are more precise than the benchmark over all horizons. Yet, in the forecasts only the random forest can offer substantial improvements. For the now- and backcast all methods improve substantially over the prevailing mean with the LASSO being the most precise.

Since the Financial Crisis, the random forecast has the most precise forecasts, nowcasts, and backcasts by a substantial margin. All other methods are not or only barely more precise than the prevailing mean. While we did not compute Diebold-Mariano statistics due to the relative small number of forecasts in each subsample, the advantage of the random forest since the Financial Crisis is remarkable.

Table 4 reports the RMSFE in different economic states. During economic crises, the prevailing mean tends to have a relatively large RMSFE and all methods improve over it over all horizons. The random forest provides the relatively most precise two-period ahead forecast and the elastic net the relatively most precise one-period ahead forecast. The LASSO provides the most precise nowcast followed by the elastic net and the dynamic factor model. The precision of the backcasts differ more substantially. The dynamic factor model provides the most precise backcast. The LASSO and elastic net are the next most precise methods.

Over periods of moderate growth and periods of expansions, the random forest provides the most precise forecasts at all horizons at often substantial margins. Only at two-quarter ahead forecasts during expansions does the LASSO improve marginally over the random forest. In fact, over periods of moderate growth and expansions, the random forest is the only method that consistently beats the prevailing mean as the other methods tend to be less precise for several forecast horizons.

A question is whether the forecast precision of the different methods is due to an increase or reduction in biases or variances over the benchmark. Table 5 displays (ratios of) the absolute forecast biases and Table 6 (ratios of) the forecast variances. The prevailing mean results show that the variance dominates the bias for this benchmark forecast. The biases of the dynamic factor model, MIDAS, random subset, and random projection methods are comparable to or even larger than that of the prevailing mean forecast. Only LASSO, elastic net and the random forest can reduce the absolute bias with the random forest offering the largest reduction in bias on average. In contrast, all methods tend to reduce the forecast variances compared to the prevailing mean forecast. For the one and two quarters ahead forecasts, the random forest offers the largest reduction in forecast variance by a substantial margin. For the nowcast, all methods except the MIDAS offer substantial variance reductions. For the backcast, finally, the dynamic factor model reduced the forecast variance by the largest amount. Overall, for most methods the improvements in MSFE over the benchmark can be explained by an improvement in the variance component. The random forest is the only method that offers substantial reduction in biases and variances over all horizons.

Table 5
Absolute forecast bias over the period 1992Q1-2018Q4

Model	Benchmark			Alternative					
	PM	AR	DFM	MIDAS	LASSO	EN	RS	RP	RF
2Q ahead	*0.18*	1.01	0.96	1.15	1.00	0.99	1.04	1.30	0.81
1Q ahead	*0.17*	0.96	1.02	1.10	0.96	0.91	1.03	1.30	0.92
Nowcast	*0.17*	0.83	1.07	1.09	0.89	0.86	1.02	1.23	0.82
Backcast	*0.17*	0.94	0.94	1.22	0.89	0.82	0.99	1.23	0.73

Note: See footnote of Table 2.

Table 6

Forecast variance over the period 1992Q1-2018Q4

Model	Benchmark			Alternative					
	PM	AR	DFM	MIDAS	LASSO	EN	RS	RP	RF
2Q ahead	*0.45*	1.01	0.99	0.96	0.95	0.96	0.94	0.98	0.91
1Q ahead	*0.45*	1.00	0.93	1.00	0.94	0.92	0.91	0.93	0.85
Nowcast	*0.44*	0.95	0.79	0.97	0.77	0.78	0.79	0.80	0.78
Backcast	*0.44*	0.95	0.58	0.75	0.67	0.69	0.76	0.81	0.70

Note: See footnote of Table 2.

Table 7

Marginal value coefficients of various nowcast methods

Model	Benchmark			Alternative					
	PM	AR	DFM	MIDAS	LASSO	EN	RS	RP	RF
Backcast									
11	-	-		-	0.27	0.19	0.03	-	0.18
10	-	-		0.67	0.16	0.24	-	-	0.32
Nowcast									
9	-	0.05		0.54	-	0.29	0.26	0.20	0.30
8	0.00	-	0.17	0.08	0.21		0.21	0.08	0.42
7	-	-	-	-	0.12	0.04	0.00	-	
1Q ahead									
6	-	-	-	0.01	0.11	0.15	-	-	
5	-	-	-	0.15	0.12	0.06	-	-	
4	-	0.00	0.16	0.27	-		-	-	0.49
2Q ahead									
3	-	-	0.03	0.22	0.19	0.17	0.23	-	
2	-	-	-	0.30	0.25	0.26	0.00	0.02	
1	-	-	-	0.33	0.39	0.35	0.19	0.18	

Notes: Entries refer to the restricted coefficients of the alternative model when added to the model with the lowest MSFE in the full forecast sample. Grey entries denote the best model. Blank entries indicate corner solutions (i.e. $\delta = 0$).

4.2 Marginal Value and Forecast Combinations

Table 7 reports the results of the encompassing tests. The two benchmark models, the prevailing mean and the AR model, do not bear any additional value in combination with the best models. The dynamic factor model only offers additional information in few cases. The coefficients of the MIDAS model, LASSO, elastic net, and random subset selection are frequently significant, and so are those of the random forest when it is not the best model. This would suggest that forecast combinations may be fruitful way to improve forecast accuracy.

Table 8 reports the results of the forecast combination schemes. The equally weighted forecast combinations are very close to the model with the lowest MSFE for all horizons and is slightly more precise—a relative RMSFE of 0.87 versus 0.88—when it comes to nowcasting. The same is true of the inverse proportional weights irrespective of estimation window.

Table 8
Forecast accuracy of forecast combinations

	PM	EA	IP	IP_{10}	OLS	OLS_{40}
2Q ahead	*0.69*	0.95	0.95	0.95	1.00	1.06
1Q ahead	*0.69*	0.93	0.93	0.93	0.99	1.08
Nowcast	*0.69*	0.87	0.87	0.87	0.95	1.00
Backcast	*0.69*	0.80	0.79	0.79	0.80	0.80

Notes: EA denotes the equally weighted forecast combination, IP forecast combinations with weights inversely proportional to the MSFE, IP_{10} forecast combinations with weights inversely proportional to the MSFE where the weights are calculated using a rolling window of 10 quarters, OLS denotes forecast combinations with weights based on the regression of Granger and Ramanathan (1984), and OLS_{40} denotes forecast combinations with weights based on the regression of Granger and Ramanathan (1984) in a rolling window of 40 quarters. Grey cells indicate the forecast combination with the lowest RMSFE.

Using OLS weights, in contrast, results in less precise forecasts and nowcasts, which suggests that forecast biases should not be a first order concern in this application.

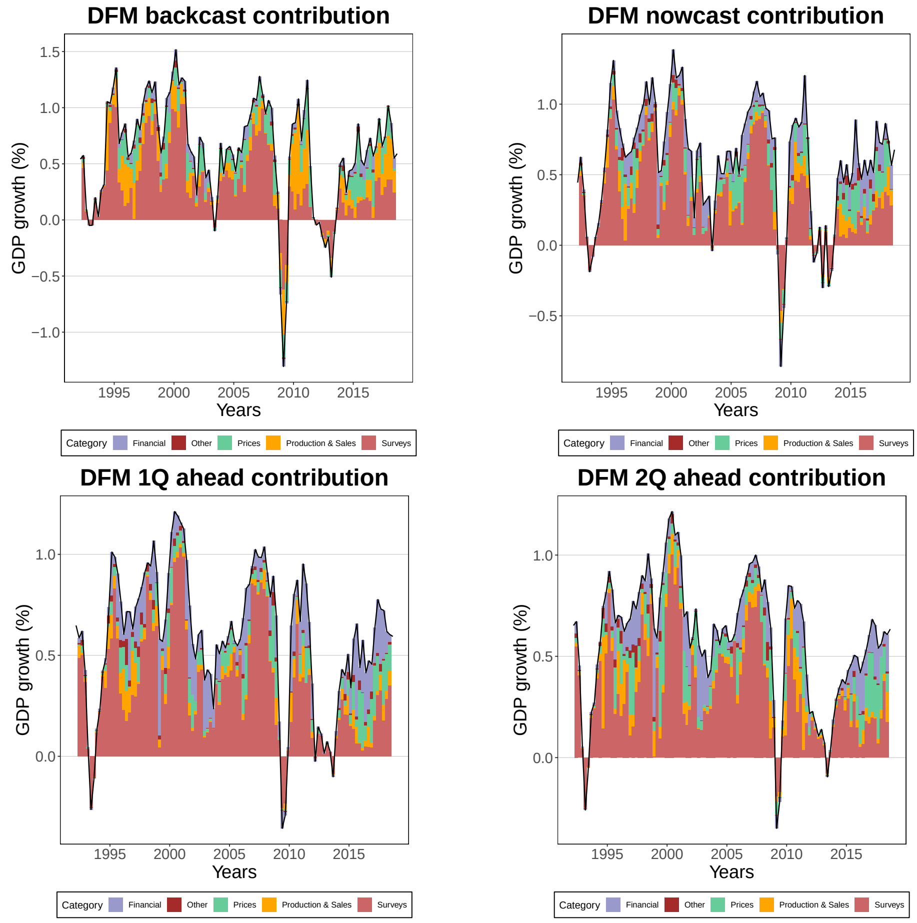
Forecast combinations using equal and inversely proportional weights, therefore, offers a practical way to ensure that forecasts, nowcasts, and backcasts are very close to the best individual model or even provides the most precise predictions. This is particularly useful as our results indicate that no individual method provides the best predictions in all circumstances.

4.3 Forecast Contributions

Figure 3 to 6 show the contributions of the different time series in each of the five categories to the forecasts of the dynamic factor model, LASSO, random subset regression, and the random forest. We concentrate on these forecasts as they represent the different classes of models and are the respective models with the best forecast accuracy. The top left graph in each figure shows the contributions to the backcast, the top right figure the contribution to the nowcast, the bottom left the contribution to the one quarter ahead forecast, and the bottom right to the two quarter ahead forecast.

The graphs for the dynamic factor model in Figure 3 show that until the Financial Crisis the time series in the *Surveys* category had the largest contribution to the forecasts across all horizons. Since then, surveys have been substantially less important. This mirrors the findings by Gayer and Marc (2018) who suggest that the relationship between hard and soft variables might have changed before and after the Financial Crisis. For the backcast the variables in the *Prices* category also plays an important role and at times the variables in the *Production & Sales* category. Since the Financial Crisis, the survey data are less important and the other two variables have gained additional importance. For the one and two period ahead forecasts the financial variables play an important role that is not present in the now- and backcast.

Figure 3
Contributions of the time series to dynamic factor model forecasts

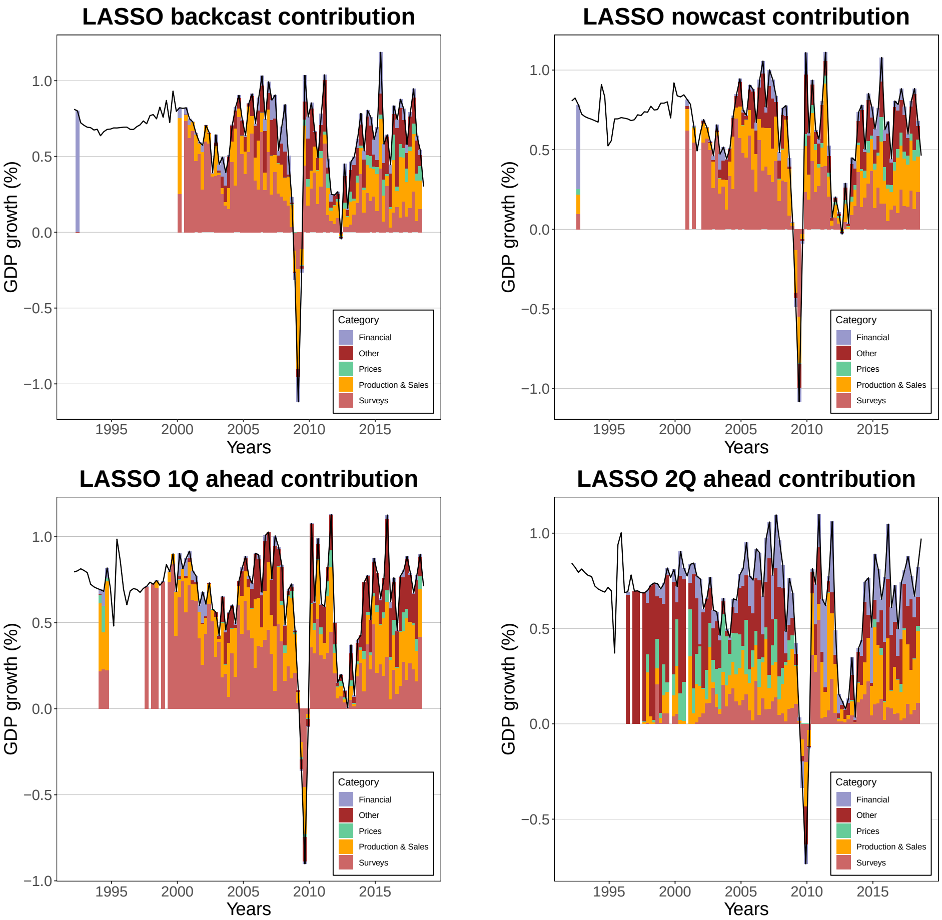


Note: Category bars indicate the relative part of GDP estimate explained by the five time series categories. The black solid line represents the forecast of GDP by the respective method.

The LASSO in Figure 4 uses more varied information already before the Financial Crisis and the decline of the importance of the survey data started earlier. In the predictions in the first half of the Great Moderation most predictions were based only on the intercept, which is likely due to the low variability of GDP growth during this period. Subsequently, LASSO selects a diverse range of variables, except for the *Prices* category. After the Financial Crisis, only little is contributed to the predictions by survey data, which is similar to the dynamic factor model. Compared to the dynamic factor model, however, the series in the *Other* category played a more prominent role. In the two quarter ahead forecasts, survey data play a minor role and financial variables are more important.

The random subset regression forecasts in Figure 5 also show a more balanced and steady use of the variables and the decline in the importance

Figure 4
Contributions of the time series to LASSO forecasts

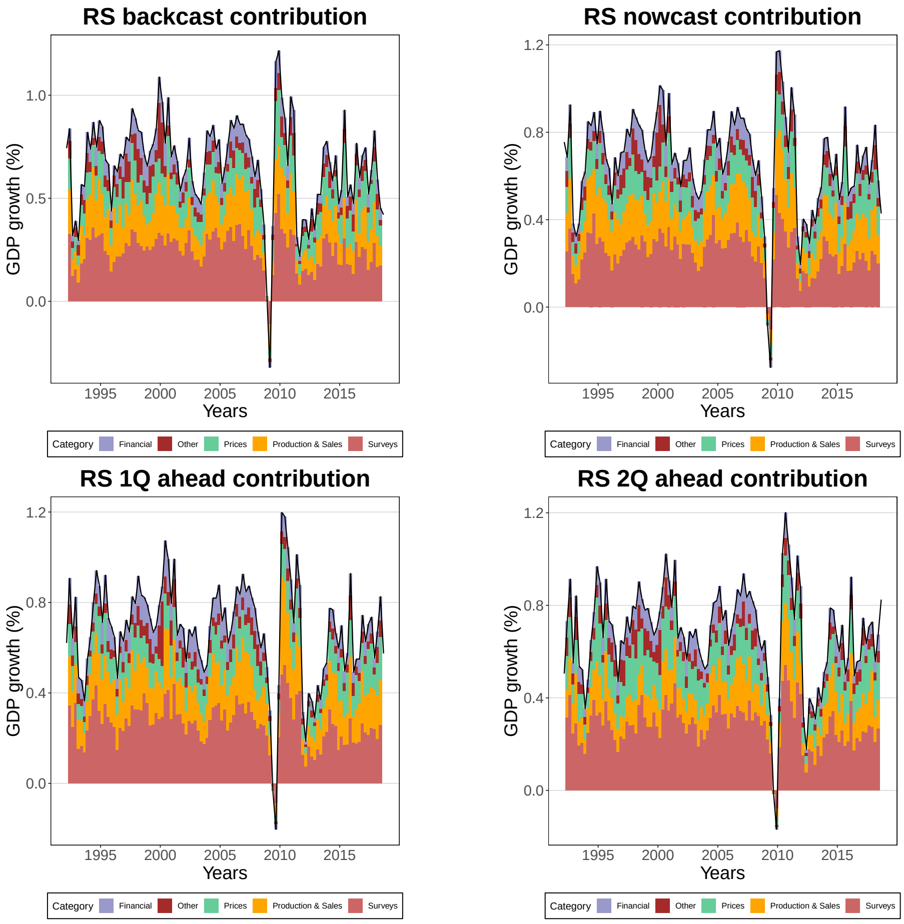


Note: See footnote of Figure 3.

of survey data is much smaller. Interestingly, the use of variables shows no major difference between the forecast horizons.

Similar to random subset regression, the random forest in Figure 6 displays a more balanced and steady use of the variables in the different categories. The importance of survey data, which had a sizable but non-dominant role before the crisis, remained strong. Variables in the *Production & Sales* category play an evenly large role and the remaining variables also contribute throughout. Again, the difference between the forecast horizons is minimal. As time proceeds, more is contributed by the category *Other*. This category comprises only four time series but these include unemployment and wages, and much information seems to be extracted by the random forest, in particular when compared to the dynamic factor model. Given that the random forest produces more precise forecasts than the dynamic factor model for all horizons since the financial crisis, the shift in

Figure 5
Contributions of the time series to random subset regression forecasts



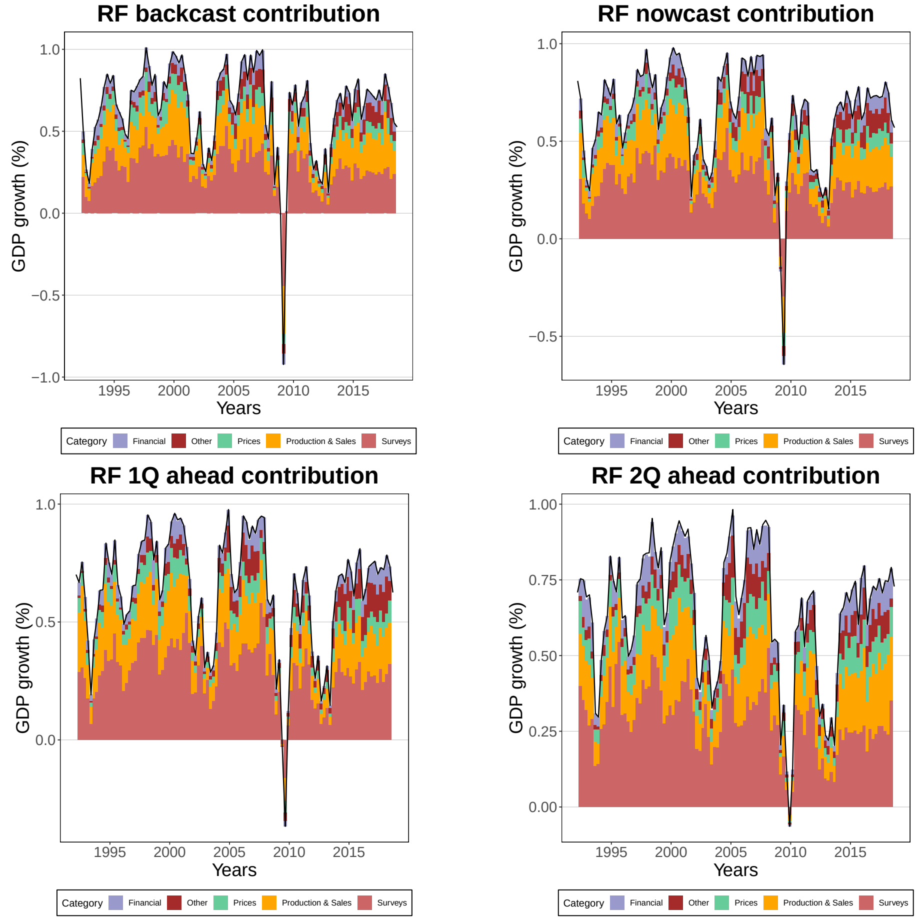
Note: RS: random subset regression. See footnote of Figure 3.

importance of the variables away from surveys may have contributed to the relative deterioration of the forecast quality of the dynamic factor model.

5 Conclusion

In this paper, we investigate whether a range of statistical methods often attributed to the machine learning literature can deliver accurate short term forecasts, nowcasts, and backcasts of Dutch GDP. Our findings suggest that since the Financial Crisis, the relative forecast performance of the dynamic factor model, which is used in many central banks, has deteriorated relative to other models with the random forest delivering the most accurate predictions. Closer inspection reveals that the contributions of the variables in the different categories to the nowcasts has changed for the dynamic factor model but not the random forest. This, together with the fact that the

Figure 6
Contributions of the time series to the random forest forecasts



Notes: RF: random forest. See footnote of Figure 3.

random forest may be better at catching potential nonlinearities, may be an explanation for this shift in relative forecast performance.

References

Bańbura, M. and Rünstler, G. (2011). A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP. *International Journal of Forecasting*, 27(2):333–346.

Bernanke, B. S. and Boivin, J. (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics*, 50(3):525–546.

Bok, B., Caratelli, D., Giannone, D., Sbordone, A., and Tambalotti, A. (2018). Macroeconomic nowcasting and forecasting with big data. *Annual Review of Economics*, 10:615–643.

Boot, T. and Nibbering, D. (2019). Forecasting using random subspace methods. *Journal of Econometrics*, 209(2):391–406.

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

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):134–144.

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

Elliott, G. and Timmermann, A. (2016). *Economic Forecasting*. Princeton University Press.

Foroni, C. and Marcellino, M. (2014). A comparison of mixed frequency approaches for nowcasting euro area macroeconomic aggregates. *International Journal of Forecasting*, 30(3):554–568.

Gayer, C. and Marc, B. (2018). A 'new modesty'? Level shifts in survey data and the decreasing trend of 'normal' growth. *Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.*

Ghysels, E., Sinko, A., and Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews*, 26(1):53–90.

Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4):665–676.

Gogas, P., Papadimitriou, T., and Sofianos, E. (2021). Forecasting unemployment with machine learning. *Journal of Forecasting*, 41(1):551–566.

Granger, C. and Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, 3(2):197–204.

Guhaniyogi, R. and Dunson, D. B. (2015). Bayesian compressed regression. *Journal of the American Statistical Association*, 110(512):1500–1514.

Hindrayanto, I., Koopman, S., and de Winter, J. (2016). Forecasting and nowcasting economic growth in the euro area using factor models. *International Journal of Forecasting*, 32(4):1284–1305.

Jansen, W., Jin, X., and de Winter, J. M. (2016). Forecasting and nowcasting real GDP: Comparing statistical models and subjective forecasts. *International Journal of Forecasting*, 32(2):411–436.

Jardet, C. and Meunier, B. (2022). Nowcasting world GDP growth with high-frequency data. *Journal of Forecasting*, (forthcoming).

Jönsson, K. (2020). Machine learning and nowcasts of Swedish GDP. *Journal of Business Cycle Research*, 16:123—134.

Koopman, S. and Harvey, A. (2003). Computing observation weights for signal extraction and filtering. *Journal of Economic Dynamics & Control*, 27(7):1317–1333.

Kuzin, V., Marcellino, M., and Schumacher, C. (2011). MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. *International Journal of Forecasting*, 27(2):529–542.

Kuzin, V., Marcellino, M., and Schumacher, C. (2013). Pooling versus model selection for nowcasting GDP with many predictors: Empirical evidence for six industrialized countries. *Journal of Applied Econometrics*, 28(3):392–411.

Lundberg, S. and Lee, S. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, pages 4765–4774.

Maehashi, K. and Shintani, M. (2020). Macroeconomic forecasting using factor models and machine learning: an application to Japan. *Journal of the Japanese and International Economies*, 58(101104).

Marcellino, M. and Schumacher, C. (2010). Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP. *Oxford Bulletin of Economics & Statistics*, 72(4):518–550.

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

Pick, A. and Carpay, M. (2022). Multi-step forecasting with large vector autoregressions. *Advances in Econometrics*, 43A:73–98.

Richardson, A., van Florenstein Mulder, T., and Vehbi, T. (2018). Nowcasting New Zealand GDP using machine learning algorithms. *CAMA Discussion Paper*.

Schumacher, C. and Breitung, J. (2008). Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. *International Journal of Forecasting*, 24(3):386—-398.

Shapley, L. (1953). A value for n-person games. *Contributions to the Theory of Games*, 2(28):307–317.

Štrumbelj, E. and Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, 41(3):647–665.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society B*, 58(1):267–288.

Timmermann, A. (2006). Forecast combinations. In Elliott, G., Granger, C., and Timmermann, A., editors, *Handbook of Economic Forecasting, Volume 1*, chapter 4, pages 135–196. Elsevier.

Wallis, K. F. (1986). Forecasting with an econometric model: The 'ragged edge' problem. *Journal of Forecasting*, 5(1):1–13.

Yoon, J. (2022). Forecasting of real GDP growth using machine learning models: Gradient boosting and random forest approach. *Journal of Forecasting*, 57:247—-265.

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

A Appendix

Table A.1
Macroeconomic time series of various economic indicators transformed into growth rates

No.	Variable	transformation				source	start	end	publ. delay
		ln.	dif.	fil.	sa.				
I. Production & Sales (N = 21)									
1	Av. daily prod. - prod. industries	1	1	3	NSA	CBS	jan 1965	jan 2019	2
2	Av. daily prod. - industry	1	1	3	NSA	CBS	jan 1965	jan 2019	2
3	Ind. prod. - cap. goods industry	1	1	3	NSA	ECB	jan 1970	jan 2019	2
4	Cons. exp. - households, dom. cons.	1	1	3	NSA	CBS	feb 1977	jan 2019	2
5	Ind. prod. - Manufacture tobacco	1	1	3	NSA	ECB	jan 1965	jan 2019	2
6	Ind. prod. - Manufacture wearing apparel	1	1	3	NSA	ECB	jan 1970	jan 2019	2
7	Ind. prod. - Manufacture motor vehicles/(semi-)trailers	1	1	3	NSA	ECB	jan 1985	jan 2019	2
8	Ind. prod. - Manufacture other transport equipment	1	1	3	NSA	ECB	jan 1970	jan 2019	2
9	Ind. prod. - Manufacturing	1	1	3	SA	ECB	dec 1979	jan 2019	2
10	Ind. prod. - Manufacture of textiles	1	1	3	SA	ECB	jan 1980	jan 2019	2
11	Ind. prod. - Printing/reproduction of recorded media	1	1	3	SA	ECB	jan 1980	jan 2019	2
12	Ind. prod. - Constr.	1	1	3	SA	ECB	jan 1985	jan 2019	2
13	Ind. prod. - MIG capital goods ind.	1	1	3	SA	ECB	jan 1970	jan 2019	2
14	Belgium, Retail trade excl. fuel, motor vehicles/cycles	1	1	3	SA	ECB	jan 1970	jan 2019	2
15	Germany, Total ind. (excl. constr.)	1	1	3	SA	ECB	jan 1965	jan 2019	2
16	Germany, Retail trade excl. fuel, motor vehicles/cycles	1	1	3	SA	ECB	jan 1968	jan 2019	2
17	Spain, Total ind. (excl. constr.)	1	1	3	SA	ECB	jan 1965	jan 2019	2
18	France, Total ind. (excl. constr.)	1	1	3	SA	ECB	jan 1965	jan 2019	2
19	France, Retail trade excl. fuel, motor vehicles/cycles	1	1	3	SA	ECB	jan 1970	jan 2019	2
20	Italy, Total ind. (excl. constr.)	1	1	3	SA	ECB	jan 1965	jan 2019	2
21	Germany, Total ind.	1	1	3	SA	ECB	mrt 1978	jan 2019	2

Table continued on next page

Macroeconomic time series of various economic indicators transformed into growth rates (*continued)*

No.	Variable	transformation				source	start	end	publ. delay
		ln.	dif.	fil.	sa.				
II. Surveys (N = 36)									
22	Prod. conf. - Headline	0	0	3	SA	ES	jan 1985	feb 2019	1
23	Constr. conf. - Headline	0	0	3	SA	ES	jan 1985	feb 2019	1
24	Constr. conf. - Building development past 3 months	0	0	3	SA	ES	jan 1985	feb 2019	1
25	Constr. conf. - Evolution current overall order books	0	0	3	SA	ES	jan 1985	feb 2019	1
26	Constr. conf. - Employment expect. next 3 months	0	0	3	SA	ES	jan 1985	feb 2019	1
27	Ind. conf. - Headline	0	0	3	SA	ES	jan 1985	feb 2019	1
28	Ind. Confidence - Production trend observed in recent months	0	0	3	SA	ES	jan 1985	feb 2019	1
29	Ind. Confidence - Assessment of order-book levels	0	0	3	SA	ES	jan 1985	feb 2019	1
30	Ind. Confidence - Assessment of stocks of finished products	0	0	3	SA	ES	jan 1985	feb 2019	1
31	Ind Confidence - Production expectations for the months ahead	0	0	3	SA	ES	jan 1985	feb 2019	1
32	Ind. Confidence - Employment expectations for the months ahead	0	0	3	SA	ES	jan 1985	feb 2019	1
33	Cons. conf. - Headline	0	0	3	SA	ES	jan 1985	feb 2019	1
34	Cons. conf. - Financial situation over last 12 months	0	0	3	SA	ES	jan 1985	feb 2019	1
35	Cons. conf. - Financial situation over next 12 months	0	0	3	SA	ES	jan 1985	feb 2019	1
36	Cons. conf. - General economic situation over last 12 months	0	0	3	SA	ES	jan 1985	feb 2019	1
37	Cons. conf. - General economic situation over next 12 months	0	0	3	SA	ES	jan 1985	feb 2019	1
38	Cons. conf. - Unemployment expectations over next 12 months	0	0	3	SA	ES	jan 1985	feb 2019	1
39	Cons. conf. - Major purchases at present	0	0	3	SA	ES	jan 1985	feb 2019	1
40	Cons. conf. - Major purchases over next 12 months	0	0	3	SA	ES	jan 1985	feb 2019	1
41	Cons. conf. - Savings at present	0	0	3	SA	ES	jan 1985	feb 2019	1
42	Cons. conf. - Savings over next 12 months	0	0	3	SA	ES	jan 1985	feb 2019	1
43	Cons. conf. - Statement on financial situation of household	0	0	3	SA	ES	jan 1985	feb 2019	1
44	BNB-indicator, gross-index	0	0	3	SA	BNB	jan 1985	feb 2019	1
45	Belgium, Cons. confidence	0	0	3	SA	ES	jan 1985	feb 2019	1
46	Germany, Cons. confidence	0	0	3	SA	ES	jan 1985	feb 2019	1

Table continued on next page

Macroeconomic time series of various economic indicators transformed into growth rates (*continued)*

No.	Variable	transformation				source	start	end	publ. delay
		ln.	dif.	fil.	sa.				
47	France, Cons. confidence	0	0	3	SA	ES	jan 1985	feb 2019	1
48	Italy, Cons. confidence	0	0	3	SA	ES	jan 1985	feb 2019	1
49	Belgium, Ind. confidence	0	0	3	SA	ES	jan 1985	feb 2019	1
50	Germany, Ind. confidence	0	0	3	SA	ES	jan 1985	feb 2019	1
51	Italy, Ind. confidence	0	0	3	SA	ES	jan 1985	feb 2019	1
52	United Kingdom, Ind. confidence	0	0	3	SA	ES	jan 1985	feb 2019	1
53	United Kingdom, Cons. confidence	0	0	3	SA	ES	jan 1985	feb 2019	1
54	Ind. Confidence - (CBS definition)	0	0	3	SA	CBS	jan 1985	feb 2019	1
55	Ind. Confidence - Prod. expect, months ahead (CBS definition)	0	0	3	SA	CBS	jan 1985	feb 2019	1
56	Ind. Confidence - Ass. of order-book levels (CBS definition)	0	0	3	SA	CBS	jan 1985	feb 2019	1
57	Ind. Confidence - Ass. of stocks of fin. products (CBS definition)	0	0	3	SA	CBS	jan 1985	feb 2019	1
III. Financial (N = 8)									
58	Loans to the private sector	1	1	3	NSA	ECB	dec 1982	jan 2019	2
59	M1	1	2	3	NSA	ECB	jan 1980	jan 2019	2
60	M3 (money in circulation inclusive)	1	2	3	NSA	ECB	jan 1970	jan 2019	2
61	Interest rate (short term) - euro	0	1	3	NSA	DNB	nov 1984	feb 2019	1
62	Loans on mortgage (nominal rate 5 to 10 years mortgage)	0	1	3	NSA	ECB	jan 1980	jan 2019	2
63	Interest rate (long term)	0	1	3	NSA	DS	jan 1965	mrt 2019	0
64	Share index, AEX	1	1	3	NSA	DS	jan 1983	mrt 2019	0
65	Share index, Amsterdam Midkap-index	1	1	3	NSA	DS	jan 1983	mrt 2019	0
IV. Prices (N = 14)									
66	Exchange rate, US-Dollar per Euro	0	1	3	NSA	ECB	jan 1965	feb 2019	1
67	Housing price	1	2	3	NSA	CBS	jan 1976	feb 2019	1
68	Consumerprice index, total CPI, all households	1	2	3	NSA	CBS	jan 1965	feb 2019	1
69	Consumerprice index, underlying inflation	1	2	3	NSA	CBS	jan 1976	feb 2019	1
70	World market commodity prices, overall	1	2	3	NSA	HWWI	sep 1978	feb 2019	1

Table continued on next page

Macroeconomic time series of various economic indicators transformed into growth rates (*continued)*

No.	Variable	transformation				source	start	end	publ. delay
		ln.	dif.	fil.	sa.				
71	World market commodity prices, industrial materials	1	2	3	NSA	HWWI	sep 1978	feb 2019	1
72	World market commodity prices, agric. & ind. materials	1	2	3	NSA	HWWI	sep 1978	feb 2019	1
73	World market commodity prices, metals	1	2	3	NSA	HWWI	sep 1978	feb 2019	1
74	World market commodity prices, energy-components	1	2	3	NSA	HWWI	sep 1978	feb 2019	1
75	Producer prices, total intermed. & fi. products (dom. market)	1	2	3	NSA	CBS	jan 1981	jan 2019	2
76	Producer prices, consumer goods (dom. market)	1	2	3	NSA	ECB	jan 1976	jan 2019	2
77	Producer prices, intermediate goods (dom. market)	1	2	3	NSA	ECB	jan 1976	jan 2019	2
78	Producer prices, intermediate & final products (for. market)	1	2	3	NSA	CBS	jan 1981	jan 2019	2
79	Producer prices, energy (dom. market)	1	2	3	NSA	ECB	jan 1980	jan 2019	2
V. Other (N = 4)									
80	Unemployment	0	1	3	SA	ES	jan 1983	feb 2019	1
81	Issued vehicle registration certificates	1	1	3	NSA	CBS	jan 1965	feb 2019	1
82	Bankruptcies	1	1	3	NSA	CBS	jan 1965	feb 2019	1
83	Hourly wages (collective labour agreement), industry	1	1	3	NSA	CBS	jan 1972	feb 2019	1

Note: The table presents the transformations of the monthly series that are used for estimation of forecasting models. Transformation: ln.: 0 = no logarithm, 1 = logarithm; dif.: degree of differencing 1 = first difference, 2 = second difference; fil.: moving average filter of degree n; sa: SA = seasonally adjusted at the source, NSA = not seasonally adjusted, adjusted with X12-ARIMA; source: CBS = Statistics Netherlands, BNB = National Bank of Belgium, DNB = National Bank of the Netherlands, DS: Datastream, ECB: European Central Bank, ES = Eurostat, HWWI = Hamburg Institute of International Economics; start: Starting year and month of the series, end: Final year and month of the series; delay: publication delay of the series.

Online Appendix: Details of the nowcasting models

Dynamic factor model

Consider a vector of n stationary monthly series $\boldsymbol{x}_m = (x_{1,m}, \ldots, x_{n,m})'$, with monthly time index $m = 1, 2, \ldots, T_m$, which have been standardized to have zero mean and unit variance. The dynamic factor model is

$$\boldsymbol{x}_m = \boldsymbol{\Lambda} \boldsymbol{f}_m + \boldsymbol{\xi}_m, \quad \boldsymbol{\xi}_m \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_\xi)$$

$$\boldsymbol{f}_m = \sum_{i=1}^{p} \boldsymbol{A}_i \boldsymbol{f}_{m-i} + \boldsymbol{B} \boldsymbol{\eta}_m, \quad \boldsymbol{\eta}_m \sim N(\boldsymbol{0}, \boldsymbol{I}_q)$$

where $\boldsymbol{f}_m$ is a $q \times 1$ vector of factors, $\boldsymbol{\Lambda}$ is a $n \times q$ matrix of factor loadings, $\boldsymbol{A}_i$ is a $q \times q$ matrix of coefficients, and $\boldsymbol{\xi}_m$ and $\boldsymbol{\eta}_m$ are $n \times 1$ and $q \times 1$ vectors of disturbances.

The latent monthly GDP growth, y_m^*, is related to the common factors through

$$y_m^* = \boldsymbol{\lambda}' \boldsymbol{f}_m + \varepsilon_m, \quad \varepsilon_m \sim \mathcal{N}(0, \sigma_\varepsilon^2) \tag{1}$$

The observed quarterly GDP growth series, y_t, with quarterly time index $t = 1, 2, \ldots, T_q$, is then

$$y_t = (y_{3t}^* + y_{3,t-1}^* + y_{3,t-2}^*)/3$$

The aggregation for the quarterly GDP growth implies that y_m is in terms of 3-month growth rates. The state space form contains the monthly quarterly GDP growth in the third month of the respective quarter with the remaining observations treated as missing

$$y_m = \begin{cases} y_{3t}, & t = 1, 2, \ldots, T \\ \text{unobserved}, & \text{otherwise} \end{cases}$$

The literature estimates the matrix of factor loadings, $\boldsymbol{\Lambda}$, a static principal components analysis applied to a balanced sub-sample of the data, where observations in periods with missing data are discarded. However, only the last few rows are discarded, as missing data are the result of publication lags. The static principal component analysis also gives sample estimates of the common factors.

The number of common factors and the number of lags in the vector autoregressive process need to be specified. We take the equally weighted average of forecasts over a range of values with the maximum value of q and p set to six, see Kuzin et al. (2013) and Jansen et al. (2016) for similar choices.

Mixed-data sampling factor-augmented model

The MIDAS model of Ghysels et al. (2007) has been adapted for nowcasting by Marcellino and Schumacher (2010). The purpose of MIDAS is to jointly model variables of different frequency, here quarterly GDP and monthly economic indicators. In the factor-augmented MIDAS model, factors are extracted at the monthly frequency and then linked to lower frequency GDP growth. The model for h-period ahead GDP growth, y_{t+h}, in this model is

$$y_{t+h} = \alpha + \boldsymbol{\beta}' C(L_M;\boldsymbol{\theta})\boldsymbol{f}_t^{(3)} + \varepsilon_{t+h}$$
$$C(L_M;\boldsymbol{\theta}) = \sum_{k=0}^{K} c(k,\boldsymbol{\theta}) L_M^k$$

where L_M^k denotes the monthly lag operator for skip-sampled lag k of the regressor at time t, α is a scalar, $\boldsymbol{f}_t^{(3)}$ the skip-sampled factors extracted from the monthly indicators , where the superscript three indicates the skip sampling of monthly indicators to quarterly frequency. Various specifications of nonlinear weighting schemes $C(L_M;\boldsymbol{\theta})$ can be employed to parsimoniously parameterize the coefficients (Ghysels et al., 2007).

The mixed-data sampling model is estimated with ordinary or nonlinear least squares for the unrestricted or restricted model. The restricted model uses the exponential Almon lag and the unrestricted model uses skip sampling.

We assessed multiple methods to extract factors at first, but used the method described in the description of the MIDAS model in Section 2 for the remainder of the paper. The methods we used to extract factors were, first, principal components on the set of explanatory variables, where we adjusted for missing observations by deleting periods with missing observations. Next, we use principal components but use skip sampling. Finally, we use principal components on a skip sampled data set including the lags of the variables.

Regularization techniques

We use the least absolute shrinkage and selection operator (LASSO) and the elastic net in this paper. We also obtained results for ridge regression and adaptive LASSO. However, the results were strictly dominated by the LASSO and elastic net and for brevity we therefore omit these results.

The LASSO of Tibshirani (1996) performs both regularization and variable selection by imposing an ℓ_1 penalty in the estimation of the coefficients. As the response variable and regressors are of different frequencies, the mixed-data sampling approach of Section 2 is employed using skip sampling for the monthly regressors.

Given a sample of length N consisting of n covariates $x_m := (x_{1,m}, x_{2,m}, \ldots, x_{n,m})$, $\forall m \in \{1, 2, \ldots, T_m\}$, one obtains the parameter estimates optimizing the penalized loss function

$$\min_{\beta_0,\boldsymbol{\beta}} \|\boldsymbol{y}_h - \beta_0 \iota_{T-h} - \boldsymbol{x}^{(3)}\boldsymbol{\beta}\|_2^2 \quad \text{subject to } \|\boldsymbol{\beta}\|_1^2 \leq \lambda$$

where $\|\boldsymbol{\beta}\|_p$ denotes the ℓ_p norm of $\boldsymbol{\beta}$, and ι_{T-h} an $T-h \times 1$ vector of ones and λ determines the extent of regularization. The optimal regularization parameter, λ, is determined via cross-validation.

We assessed the performance of LASSO in the case of various regularization parameters $\hat{\lambda}_{min}$ and $\hat{\lambda}_{\sigma}$. $\hat{\lambda}_{min}$ and $\hat{\lambda}_{\sigma}$ refer to the regularization parameter estimate that results in a minimal cross-validated estimated error and the parameter estimate that results in a cross-validated estimated error that is one standard deviation away from the minimum, where the latter implies greater regularization. Due to better results of the LASSO using $\hat{\lambda}_{min}$ instead of $\hat{\lambda}_{\sigma}$ as the regularization parameter, we have used $\hat{\lambda}_{min}$ throughout the paper.

The elastic net of Zou and Hastie (2005) imposes a combination of ℓ_1 and ℓ_2 penalties. Similar to the LASSO, the ℓ_1 norm selects parameters by shrinking some to zero but it also shrinks the remaining coefficients towards zero through the use of the ℓ_2 norm. The elastic net regression is

$$\min_{\beta_0,\boldsymbol{\beta}} \|\boldsymbol{y}_h - \beta_0 \iota_{T-h} - \boldsymbol{x}^{(3)}\boldsymbol{\beta}\|_2^2 \quad \text{subject to } \alpha\|\boldsymbol{\beta}\|_1^2 + (1-\alpha)\|\boldsymbol{\beta}\|_2^2 \leq \lambda$$

$\mathbf{y_h}$ the relevant vector of GDP realizations for forecasting horizon h, ι_N an $N \times 1$ vector of ones, $\mathbf{x^{(3)}}$ a matrix of the skip-sampled versions of $x_{i,m}$ λ a pre-specified parameter determining the extent of regularization and α determining the relative extent of regularization performed by both norms. Again, the parameters α and λ are determined via cross-validation.

Random subspace regression

Model averaging has been shown to reduce the MSFE. Based on this observation, Elliott et al. (2013) introduce complete subset regression, where forecasts are constructed from all combinations of k regressors out of the regressor pool. The forecasts are then averaged. If, however, the regressor pool is large the number of combinations of k regressors is prohibitively large. A solution is to take R randomly chosen subsets of regressors. Boot and Nibbering (2019) show that this approximates the complete subset regression for mildly large R, such as $R = 1000$.

In the nowcasting context, the regression is

$$y_{t+h} = \boldsymbol{x}_t^{(3)}\boldsymbol{R}\boldsymbol{\beta}_{\boldsymbol{R}} + \varepsilon_{\boldsymbol{R},t+h}$$

where $\boldsymbol{R}$ is an $K \times k$ random selection matrix, $\boldsymbol{\beta}_R$ the associated $k \times 1$ vector of coefficients. More specifically, $\boldsymbol{R}$ is a random selection matrix that selects random sets of k predictors out of the total available n predictors, that is, it is a matrix of zeros except for k elements: the j, l-th element, which is unity if the l-th regressor in the random subset regression is the j-th variable.

A tuning parameter of this method is the size of each regressor subset, k. Theoretical results by Boot and Nibbering (2019) suggest that k should be chosen relatively large at about 30. The experience of Pick and Carpay (2022) suggests that smaller k can deliver more precise forecasts. We initially experimented with different choices of k up to 30 and our experience confirms that smaller choices of k deliver better nowcasts. As a result we average nowcasts over those obtained using $k = 2, 3, 4, 5$.

An alternative to selecting regressors would be to combine the regressors with random weights. Boot and Nibbering (2019) discuss this option and name it random projection. In place of a selection matrix, random projection uses a random weighting matrix, that calculates k regressors that are weighted averages of the n regressors. For Gaussian random projections, the weights are drawn from a normal distribution and each entry of the matrix $\boldsymbol{R}$ is independently and identically distributed as

$$\left[R\right]_{i,j} \sim \mathcal{N}(0,1), \quad 1 \leq i \leq n, \quad 1 \leq j \leq k$$

Multiple realizations of the random matrix R are drawn and the resulting forecasts are averaged. Again, the choice variable k needs to be determined. Again, we average nowcasts over those obtained using $k = 2, 3, 4, 5$.

Other weighting schemes are possible, for example, random compression used by Guhaniyogi and Dunson (2015). The analysis of Pick and Carpay (2022), however, suggests that random subset selection and random projection deliver superior forecast performances and for brevity we therefore limit our attention to these two methods.

Random forest

The random forest forecast averages the forecasts of multiple decision trees. To grow a decision tree, the space of predictor values is partitioned with the aim of minimizing the in-sample squared error. At each partition, the algorithm chooses a split based on one of the predictors that realizes the largest decrease in squared error. Hence, the split of a predictor that is minimizing the cost function is chosen at each node, i.e.

$$C = \sum_{R_g} \sum_{t_j \in R_g} (\bar{y}_{R_g} - y_j)^2$$

where C is the cost to be minimized, R_g for $g \in [1, \ldots, G]$ is the set of partitioned responses, $\bar{y}_{R_g}$ is the average GDP realization within cluster R_g and y_j is the j^{th} element of partition R_g. A tree is therefore a nonlinear combination of the predictors and allows for a nonlinear underlying GDP.

Trees are designed to have a high degree of independence of each other by randomly drawing a subset of predictor variables and a subset of observations to grow any given tree. Averaging the forecasts from the trees in the random forest therefore minimizes the variance of the average forecast.

In order to reduce overfitting, each estimation sample is divided in a training and a validation set. The share of the training set in all estimation samples is varied such that ω of the estimation sample is assigned to the training set, with $\omega \in \{0.6, 0.7, 0.8, 0.9\}$ and choose $\kappa \in [1, 2, \ldots, 249]$ skip-sampled predictors to split the tree. Subsequently, a validation set is used to measure the performance for each κ. We use the prediction of the 400 trees that resulted in the lowest prediction error in the training set.