Nowcasting GDP using machine learning methods*

Dennis Kant Andreas Pick†
Erasmus University Rotterdam

Jasper de Winter
De Nederlandsche Bank

October 6, 2023

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 period 1992Q1–2018Q4 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 macroeconomic and financial predictors. We find that, on average, the random forest provides the most accurate forecast and nowcasts, whilst the dynamic factor model provides the most accurate backcast.

Keywords: factor models; forecasting competition; machine learning methods; nowcasting.

JEL classification: C32, C53, E37.

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 current quarter, a practice referred to as "nowcasting". Such nowcasting models use dimension reduction techniques to nowcast GDP from a large number of macroeconomic and financial predictors.

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 two and quarters ahead

*The research for this paper was conducted while the first author was at De Nederlandsche Bank. We thank seminar participants at DNB and the ECB, the associate editor and two anonymouns referees for valuable comments. The opinions expressed are those of the authors and do not represent the views of DNB or the Eurosystem.

†Corresponding author: andreas.pick@cantab.net.

forecasts, nowcasts, and backcasts using the example of Dutch GDP with a broad data set of 83 macroeconomic and financial predictors. We consider the performance of several methods: factor models, 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 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 predictors.

Our findings suggest that, on average, the random forest provides the most precise predictions. An exception are the backcasts of the dynamic factor model until the financial crisis. Yet, this advantage of the dynamic factor model has disappeared and the random forest and LASSO model provide slightly more precise backcasts since the financial crisis. We observe that the random forest and random subset selection use the different predictors considerably more evenly over categories as well as over horizons than the dynamic factor model and the LASSO.

An important benchmark for 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. Babii et al. (2022) provide theoretical properties for the LASSO when estimating models with macroeconomic and financial time series data. They compare nowcasts of US GDP from the LASSO, ridge regression, and elastic net to the nowcasts of the New York Fed. They find that only their suggested specification of the LASSO outperforms the dynamic factor model of the FED. This corresponds with our results for the Netherlands, where the LASSO and the elastic net are of comparable precision to the dynamic factor model over the entire forecast period.

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 predictors in the data set in a random manner. The predictions are then averaged to yield the final forecast. Our results show that this is a competitive approach to nowcasting.

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 predictors 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 predictors 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 used to average the forecasts introduce uncertainty, which can negatively impact the forecast precision. In our application, average forecasts are among the best forecasts 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 neighbor 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 predictors 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 predictors in our nowcasts.

Machine learning methods have also been used to predict other macroeconomic predictors. 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 predictors 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 methodology

2.1 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. Since these methods are well established in the literature, we relegate a more formal description to the Online Appendix. We compare the nowcasts from these methods to those from two simple models. The first is the prevailing mean model, which takes the mean in the estimation sample as the forecast. This model serves as the benchmark in that we report the root mean square forecast error (RMSFE) of each method as a ratio to that of the prevailing mean. The second is the autoregressive model with lag length chosen by AIC with a maximum lag number of $p = 4$. All parameters are re-estimated for each nowcast using an expanding window.

Dynamic factor model

The dynamic factor model is widely used in the nowcasting literature. In this paper, we use the model developed by the Dutch central bank, which relies on the specification introduced by Giannone et al. (2008), Bańbura and Rünstler (2011), Jansen et al. (2016), and Bok et al. (2018). The nowcasting dynamic factor model is based on the approximate factor model of Stock and Watson (2002), which uses principal components to estimate the factor loadings in the dynamic factor model. The specification depends on the number of static factors (r), the number of dynamic factors (q) and the number of lags (p) in the VAR. In line with the literature (Kuzin et al., 2013; Jansen et al., 2016), we produce nowcasts for models using all combinations of one to six static factors, one to six lags for the static factors, and one to six dynamic factors (where $q \leq r$). In total we estimate 126 model specification. We then combine these nowcasts using an equal weighted average to obtain 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 of factors 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 estimate the factor-augmented model, using the first factor from each of the 126 specifications, i.e. 1 to 6 static/dynamic factors and 1 to 6 lags in the factor VAR. 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. These nowcasting models relate GDP growth to the full set of macroeconomic and financial predictors and their lags, and we skip sample these predictors to account for their monthly frequency. This leads to a extremely highly parameterized model and the necessary dimension reduction is then achieved through penalization.

Compared to the previous two models, the models using regularization assume a sparse DGP rather than the dense specification of the factor models. The difference between the two models is that the LASSO of Tibshirani (1996) performs both regularization and predictor selection by imposing an ℓ_1 penalty in the estimation of the coefficients, whereas the elastic net of Zou and Hastie (2005) imposes the ℓ_1 norm to select predictors and shrinks the remaining coefficients towards zero through the use of the ℓ_2 norm.

The amount of shrinkage in the LASSO is determined by a scalar parameter, λ, and by two scalar parameters, λ and α in the elastic net. We determine these parameters 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.

Random subspace regression

Random subspace methods encompass methods that reduce the predictor 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 predictors. Boot and Nibbering (2019) show that this approximates the complete subset regression for finite R, such as $R = 1000$.

The full predictor set contains all skip-sampled predictors and their lags. These are then sampled at random and a forecast, nowcast, or backcast is produced from a linear model containing the sampled predictors. The final prediction is the equally weighted average of the R predictions.

A tuning parameter of this method is the size of each predictor 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 predictors is to use weighted averages of the predictors. Boot and Nibbering (2019) discuss this option and call it 'random projection'. The skip-sampled predictors and their lags are combined to a small number k of weighted averages using a random weighting matrix. For Gaussian random projections, the weights are independently drawn from a standard normal distribution. The k weighted averages are then used in a linear regression model to predict GDP growth. R realizations of the weights are drawn and the resulting forecasts are averaged with equal weights. Again, k needs to be determined and, as above, we average nowcasts 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 averages the forecasts of multiple regression trees. To grow a regression tree, the space of predictors, which, similar to above, are skip-sampled, 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 predictor 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 predictors and a subset of observations to grow any given tree. Averaging the forecasts from the trees in the random

Table 1
Hyperparameter choices

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

forest therefore minimizes the variance of the average forecast.

Regression trees, if unchecked, have the tendency to overfit the data. 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 grow 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.

2.2 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 several forecasts from different sources has a long track record in the forecasting literature (Timmermann, 2006). Forecast combinations require combination weights. As optimal weights can be determined only under very specific assumptions, we use a range of practical approaches to determine the weights.

The first, 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.

2.3 Nowcast evaluations

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. Define the forecast error as $\hat{e}_{t|t-h,a} = y_t - \hat{y}_{t|t-h,a}$, where subscript a denotes the respective method. Then the ratio of RMSFE is

$$\text{relative RMSFE} = \frac{\sqrt{\frac{1}{T_f}\sum_{t=1}^{T_f} \hat{e}^2_{t|t-h,a}}}{\sqrt{\frac{1}{T_f}\sum_{t=1}^{T_f} \hat{e}^2_{t|t-h,pm}}}$$

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

In order to distinguish the effect of bias and standard deviation on the forecast performance, we also report the ratio of the absolute forecast bias and the ratio of the forecast standard deviations. The ratio of the absolute bias is

$$\text{relative abs.bias} = \frac{\frac{1}{T_f}\sum_{t=1}^{T_f} |\hat{e}_{t|t-h,a}|}{\frac{1}{T_f}\sum_{t=1}^{T_f} |\hat{e}_{t|t-h,pm}|}$$

and the ratio of the standard deviation is

$$\text{relative std.deviation} = \frac{\sqrt{\frac{1}{T_f}\sum_{t=1}^{T_f}\hat{e}^2_{t|t-h,a} - \bar{\hat{e}}^2_{h,a}}}{\sqrt{\frac{1}{T_f}\sum_{t=1}^{T_f}\hat{e}^2_{t|t-h,pm} - \bar{\hat{e}}^2_{h,pm}}}$$

where $\bar{\hat{e}}_{h,a} = \frac{1}{T_f}\sum_{t=1}^{T_f}\hat{e}_{t|t-h,a}$ for a given nowcasting method a, including the prevailing mean model.

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.

2.4 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. Note, however, that the contributions merely provide the importance of predictors in a given model, which need not imply a structural economic interpretation. For the dynamic factor model, the weights that are assigned to predictors in a model are determined using the methods of Koopman and Harvey (2003) and Bańbura and Rünstler (2011).[1]

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 predictors have with GDP. The random subspace methods also fall into the class of linear models conditional on the selection predictors 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 predictor in question to the loss of trees that do not include

[1] We assign the contribution of the constant to the five predictor groups based on the derived average weight per predictor group. We calculate weights for each forecasting horizon and re-scale all weights to lie in the interval $[0.01, +\infty)$. See Koopman and Harvey (2003) for analytical derivation of the weights.

that predictor. Denote the loss of a tree that includes a given predictor i by $L(\mathcal{S} \cup \{i\})$ and that of another tree with the same predictors except predictor i by $L(\mathcal{S})$, where $\mathcal{S}$ is the set of predictors in the tree except predictor i, and $\mathcal{S} \subseteq \mathcal{F}$ with $\mathcal{F}$ denoting the complete set of predictors. As the effect of predictor i likely depends on the other predictors in the tree, the loss differential is computed for all possible predictors in a tree. The contribution of predictor i, ϕ_i, is the weighted average of the loss 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 predictors in trees excluding predictor i. With many predictors, this procedure is computationally burdensome and we therefore use the approximation proposed by Štrumbelj and Kononenko (2014), which uses randomly sampled subsets of predictors to compute the Shapley value for each predictor.

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 and the Dutch RAI association.

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 predictors, 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 predictors. 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 predictors 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. Seasonal adjustment and standardization is done in pseudo-real time so as to mimic the process of a nowcaster over our forecast sample. 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 estimation period starts in 1986M1. 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 models relies on the data of the estimation sample and, after each forecast, the estimation sample is expanded to include the additional observations.

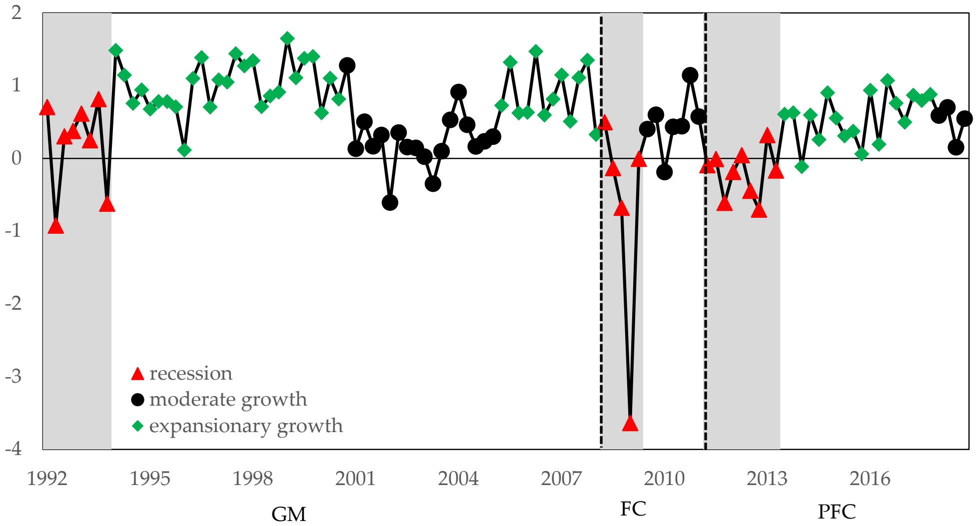
Our forecast period spans different economic episodes. We, therefore, separately analyze the forecasts for the period until the financial crisis, that is, the period of the Great Moderation (1992Q1–208Q1). Next, we consider the Financial Crisis (2008Q2–2011Q1) and, finally, the period Post Financial Crisis (2011Q2–2018Q4). We abbreviate these periods as GM, FC and PFC, respectively. Additionally, we analyze the performance of the model during recessionary periods, expansionary periods and periods of moderate growth, based on the OECD recession indicator for the Netherlands. Based on our expert judgment we remove the 2001 recession and the 2018 recession from the mechanical business cycle dating tool of the OECD because these periods could better be characterized as periods of moderate growth than recessions. Moreover, we characterize the period between the 2008-2009 recession and the 2011-2023 recession as a period of moderate growth as well. During the latter period, growth remained stagnant. Over our entire forecasting period, the Dutch economy was in recession during 22 quarters, expanded during 57 quarters, and during 29 quarters the economy grew moderately. Figure 1 shows the periods graphically.

The ragged edge

The different indicators are published with varying delays. Generally, the hard predictors are accompanied with a significant publication delay, whereas soft predictors (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 accounts for data availability at the moment when the prediction would be made in real time. Our approach is akin to the direct forecasting approach: rather than forecast predictors that are still unreleased and include the predicted values in our nowcasts, we predict GDP using the latest available data. M1, for example, has a publication delay of two month (see the appendix). Predictions for all horizons in May, say, will

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



Note: Three time periods are depicted: GM: Great Moderation (1992Q1-2008Q1), FC: Financial Crisis (2008Q2-2011Q1) and PFC: Post-Financial Crisis (2011Q2-2018Q4). Shaded areas recessions according to OECD recession indicator for the Netherlands.

therefore use the M1 number from March. The estimation of each model equally account for the publication lags such that treatment of publication lags is consistent throughout the modelling. 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 imply 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

Table 2 reports the RMSFE over the entire forecast sample over eleven forecast horizons: three monthly forecasts for each of the two quarters before the target quarter, three monthly nowcasts during the target quarter, and two monthly backcasts in the quarter after the target quarter. It can be seen from the table that for the short-term forecasts and nowcasts, only the random forest beats the prevailing mean at all horizons. It is also the most precise forecast for all the short term forecasts and the first two nowcasts. The Diebold-Mariano test shows statistical significance for these horizons. For the third nowcast, the LASSO is slightly more precise. The dynamic factor model is the most precise methods when backcasting.

The AR and random projection are generally no more precise than the dynamic factor model. The MIDAS model is slightly more precise than the DFM at four horizons but is less precise at four horizons. The LASSO, elastic net, and random subset regression are slightly more precise than the dynamic factor at most forecasting and nowcasting horizons 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 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 levels of RMSFEs, which generally decrease as more information becomes available. The random forest has consistently lower RMSFE then the other methods for the forecasting months. For the first two nowcasting months it is also the most precise. During the last nowcasting month, the regularization methods and the dynamic factor model catch up with the random forest. In the backcasting months the dynamic factor model improves substantially over the other methods. Of the other models, the MIDAS and random projection methods remain at the upper level of the RMSFE range.

The forecast sample contains periods of different nature, such as the Great Moderation and the Financial Crisis. In order to evaluate the robustness of our results, Table 3 reports the (ratio of) RMSFE over three subperiods: the Great Moderation, Financial Crisis, and the Post-Financial Crisis period.

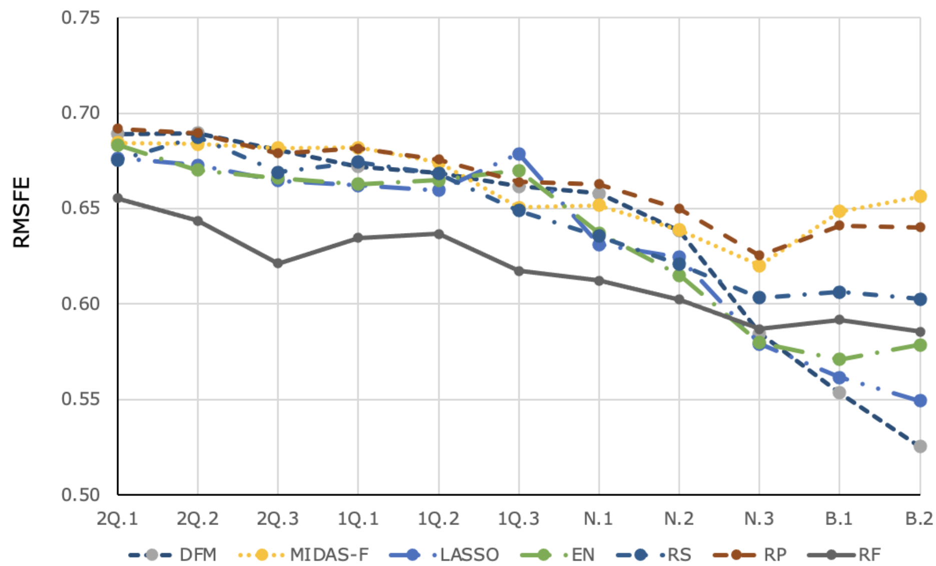
During the Great Moderation, the two quarter ahead forecast from all models, with the exception of the AR model, are rouphly equivalent in precision to that of the benchmark model for the first two months. In the last month of the two-quarter ahead forecast, the random forest is substantially better than the benchmark and significantly better than the DFM. As we get closer to the target quarter, the random forest continues to provide

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

Horizons	Benchmarks			Alternatives					
	PM	AR	DFM	MIDAS	LASSO	EN	RS	RP	RF
2Q.1	*0.68*	1.04	1.02	1.01	1.00	1.01	1.00	1.02	0.97*
2Q.2	*0.68*	1.04	1.02	1.01	1.00	0.99	1.02	1.02	0.95**
2Q.3	*0.67*	1.03	1.01	1.01	0.99	0.99	0.99	1.01	0.92**
1Q.1	*0.67*	1.03	1.00	1.01	0.98	0.99	1.00	1.01	0.94**
1Q.2	*0.67*	1.03	0.99	1.00	0.98	0.99	0.99	1.00	0.95*
1Q.3	*0.67*	1.02	0.99	0.97	1.01	1.00	0.97	0.99	0.92**
N.1	*0.67*	1.01	0.98	0.97	0.94*	0.95*	0.95**	0.99	0.91**
N.2	*0.67*	1.01	0.95	0.95	0.93	0.92	0.93	0.97	0.90**
N.3	*0.67*	0.96	0.88	0.93	0.87	0.87	0.90	0.94	0.88
B.1	*0.67*	1.00	0.83	0.97	0.84	0.86	0.91	0.96	0.89
B.2	*0.67*	1.00	0.79	0.98	0.82	0.87	0.90	0.96	0.88

Note: The table reports the results over the entire forecast sample. The first column reports the periods of origin of the forecast: XQ.Y denotes a X quarters ahead forecast made in the Y-th month in the quarter, N.Y denotes nowcasts made in the Y-th month of the quarter, and B.Y denote backcasts made in the Y-th month of the quarter. 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 one-sided Diebold-Mariano test relative to the DFM (the significance levels should be interpreted with caution due to the use of expanding windows).

Figure 2: RMSFE over varying forecast horizons



Note: The table reports the level of the RMSFE of the different methods. For further details see the footnote of Table 2.

the most precise predictions one quarter ahead and also the most precise nowcasts by a substantial margin. The most precise backcasts, however, come from the dynamic factor model. The AR and MIDAS models fail to beat the prevailing mean benchmark at any horizon.

During the Financial Crisis, the prevailing mean forecast is substantially less precise in absolute terms. Still, most models fail to deliver more precise one and two quarter ahead forecasts than this benchmark. The exception is, the random forest, which is more precise than the prevailing mean over all forecasting horizons and offers improvements of up to 18% aginst the prevailing mean. For the now- and backcast all methods improve substantially over the prevailing mean with the LASSO being the most precise, with the dynamic factor model and elastic net trailing by a small amount. The gain in relative forecasting accuracy can amount to 35% against the prevailing mean

Since the Financial Crisis, the random forest has the most precise forecasts, nowcasts, and first backcast with improvements of up to 13%. The DFM also improves over the benchmark when nowcasting and backcasting albeit to a smaller degree than the random forecast. The MIDAS and regularization methods fail to considently beat the benchmark over this period. The random subspace methods, in particular the random subset regression, provide subsantial improvements when nowcasting and backcasting, even if the forecasts are not consistently better than the benchmark.

Table 4 reports the RMSFE in different economic states. During recessions, the prevailing mean has a relatively large RMSFE and all methods, with the exception of the AR model, improve over it over all horizons. This holds especially for the late nowcasts and backcast, when the forecasting gains amount to 37% for the best model. The random forest provides the relatively most precise two-quarter ahead forecasts. The LASSO model is the most precise for the one quarter ahead forecasts and nowcast for most months. The dynamic factor model provides the most precise backcast.

Over periods of moderate growth, the random forest provides the most precise forecasts at all horizons. It is the only methods that consistently beats the prevailing mean benchmark. During expansions, the prevailing mean benchmark is considerably more precise than during the recessions and moderate growth. Most other methods provide similarly precise forecasts but cannot consistently beat the prevailing mean. Compared to the dynamic factor model, the random forecast provides more precise nowcasts, whereas the regularized methods provide considerably less precise nowcasts.

A question is whether the forecast precision of the different methods is due to an increase or reduction in biases or standard deviations over the benchmark. Table 5 displays (ratios of) the absolute forecast biases. The prevailing mean results show that the bias is the relatively smaller component for this benchmark forecast. The biases of the dynamic factor model, MIDAS, random subset, and random projection methods are comparable to

Table 3
Forecast precision evaluated over different subperiods

Horizons	Benchmarks			Alternatives					
	PM	AR	DFM	MIDAS	LASSO	EN	RS	RP	RF
Great Moderation (N= 65)									
2Q.1	*0.53*	1.06	1.00	1.03	1.00	1.03	1.00	1.05	1.03
2Q.2	*0.53*	1.06	1.00	1.02	1.01	0.99	1.05	1.02	1.00
2Q.2	*0.53*	1.05	0.99	1.02	0.95	0.95	1.00	0.98	0.92*
1Q.1	*0.53*	1.05	0.97	1.03	1.03	1.02	1.02	1.02	0.96
1Q.2	*0.53*	1.05	0.98	1.02	1.00	1.01	1.00	1.00	0.95
1Q.3	*0.53*	1.04	0.98	1.01	1.04	0.99	0.98	0.98	0.92*
N.1	*0.53*	1.04	0.98	1.02	1.00	1.01	0.97	1.03	0.94*
N.2	*0.53*	1.04	0.98	1.02	1.01	0.99	0.95	1.02	0.92**
N.3	*0.53*	1.02	0.95	1.01	0.99	0.99	0.97	1.00	0.91*
B.1	*0.53*	1.03	0.92	1.02	0.99	0.98	0.99	1.03	0.97
B.2	*0.53*	1.03	0.88	1.00	0.98	0.98	0.97	1.02	0.94
Financial Crisis (N= 12)									
2Q.1	*1.36*	1.02	1.02	1.00	1.00	0.99	0.99	0.99*	0.92*
2Q.2	*1.36*	1.02	1.02	1.00*	0.97	0.97	0.99	1.00	0.90*
2Q.3	*1.36*	1.03	1.02	1.00	0.98	1.00	0.99	1.01	0.91*
1Q.1	*1.36*	1.02	1.02	1.01	0.95	0.97	0.99	1.00	0.94*
1Q.2	*1.36*	1.02	1.01	0.99	0.95	0.95	0.99	1.00	0.95
1Q.3	*1.35*	1.02	0.99	0.94*	0.98	0.97	0.96*	0.99	0.92*
N.1	*1.35*	1.01	0.99	0.93**	0.90**	0.92**	0.95*	0.96	0.90*
N.2	*1.35*	1.01	0.93	0.89	0.78*	0.79	0.91	0.93	0.88
N.3	*1.35*	0.93	0.78	0.84	0.69*	0.71*	0.85	0.87	0.85
B.2	*1.35*	0.99	0.71	0.92	0.65*	0.72	0.83	0.90	0.82
B.1	*1.35*	0.99	0.66	0.96	0.65	0.69	0.83	0.91	0.83
Post Financial Crisis (N= 31)									
2Q.1	*0.52*	1.03	1.04	1.00**	1.02	1.02	1.03	1.05	0.95*
2Q.2	*0.52*	1.03	1.05	1.02	1.03	1.04	1.02	1.08	0.97
2Q.3	*0.52*	1.01	1.03	1.03	1.08	1.05	1.02	1.08	0.97
1Q.1	*0.52*	1.02	1.01	1.01	0.97	0.96	0.98	1.02	0.92*
1Q.2	*0.52*	1.02	0.97	0.98	1.00	1.04	0.99	1.02	0.92
1Q.3	*0.52*	0.98	0.99	0.97	1.04	1.11	0.97	1.02	0.92
N.1	*0.52*	0.96	0.98	0.98	0.93	0.92	0.92	0.99	0.88
N.2	*0.52*	0.96	0.96	0.97	1.12	1.05	0.92	0.98	0.90
N.3	*0.51*	0.89	0.93	0.97	1.00	0.98	0.90	0.96	0.88*
B.1	*0.51*	0.96	0.90	1.01	0.92	0.90	0.92	0.94	0.87
B.2	*0.51*	0.96	0.88	1.00	0.86	1.02	0.93	0.95	0.87

For further details see the footnote of Table 2.

Table 4
Forecast precision over various economic states

Horizons	Benchmarks			Alternatives					
	PM	AR	DFM	MIDAS	LASSO	EN	RS	RP	RF
Recessions (N= 22)									
2Q.1	*1.21*	1.03	1.00	0.97**	0.99	0.99	0.97	1.00	0.92*
2Q.2	*1.21*	1.03	0.99	0.97*	0.96	0.97	0.99	1.01	0.92*
2Q.3	*1.21*	1.03	0.97	0.96	0.97	0.97	0.97	1.00	0.91*
1Q.1	*1.21*	1.03	0.95	0.98	0.90	0.92	0.97	0.99	0.91**
1Q.2	*1.21*	1.03	0.94	0.96	0.88	0.90	0.94	0.98	0.92
1Q.3	*1.20*	1.02	0.95	0.92	0.92	0.91	0.93	0.97	0.89**
N.1	*1.20*	1.00	0.93	0.94	0.88*	0.89	0.91	0.95	0.89*
N.2	*1.20*	1.00	0.87	0.91	0.82	0.83	0.88	0.93	0.88
N.3	*1.20*	0.92	0.75	0.87	0.76	0.76	0.83	0.88	0.85
B.1	*1.20*	0.99	0.68	0.94	0.73	0.77	0.84	0.91	0.84
B.2	*1.20*	0.99	0.63	0.96	0.66	0.69	0.84	0.91	0.83
Moderate Growth (N= 29)									
2Q.1	*0.53*	1.06	1.07	1.09	1.06	1.07	1.05	1.07	0.96*
2Q.2	*0.53*	1.06	1.09	1.10	1.11	1.03	1.06	1.06	0.89**
2Q.3	*0.52*	1.06	1.12	1.12	0.99	1.03	1.05	1.05	0.84**
1Q.1	*0.52*	1.04	1.12	1.11	1.21	1.14	1.09	1.08	0.97
1Q.2	*0.52*	1.04	1.12	1.11	1.17	1.14	1.11	1.07	0.94
1Q.3	*0.52*	1.05	1.07	1.09	1.14	1.10	1.07	1.05	0.89*
N.1	*0.52*	1.02	1.11	1.06	1.06	1.08	1.04*	1.08	0.94
N.2	*0.52*	1.02	1.14	1.06	1.00	0.97*	1.04*	1.08	0.89**
N.3	*0.52*	0.97	1.09	1.06	0.97	0.91**	1.06	1.06	0.84**
B.1	*0.52*	1.02	1.07	1.08	0.96	0.95	1.02	1.07	0.90*
B.2	*0.52*	1.02	1.05	1.09	1.02	1.03	1.03	1.08	0.89**
Expansionary Growth (N= 57)									
2Q.1	*0.39***	1.03	1.06	1.08	0.99*	1.02	1.05	1.07	1.14
2Q.2	*0.39**	1.03	1.05	1.08	1.01	1.03	1.06	1.04	1.10
2Q.2	*0.39**	1.02	1.07	1.08	1.06	1.03	1.04	1.02*	1.04
1Q.1	*0.39*	1.03	1.06	1.05	1.06	1.07	1.06	1.04	1.04
1Q.2	*0.39*	1.03	1.06	1.06	1.15	1.16	1.08	1.02	1.05
1Q.3	*0.39*	1.03	1.06	1.04	1.21	1.22	1.01	1.02	1.06
N.1	*0.39*	1.03	1.05	1.03	1.04	1.03	1.00	1.05	0.99
N.2	*0.39*	1.03	1.06	1.03	1.22	1.16	1.00*	1.02	0.98**
N.3	*0.39*	1.08	1.07	1.02	1.11*	1.16	0.99	1.01	1.02
B.1	*0.39*	1.03	1.06	1.00	1.07	1.05	1.03	1.04	1.04
B.2	*0.39*	1.03	1.01	0.95	1.11	1.22	1.01	1.01	1.01

For further details see the footnote of Table 2.

or larger than that of the prevailing mean forecast. Only the LASSO, elastic net and the random forest can reduce the absolute bias with the random forest offering the largest reduction in bias on average.

Table 6 shows the (ratios of) the forecast standard deviations. In contrast to the results for the bias, all methods, except the AR model, tend to reduce the forecast standard deviation compared to the prevailing mean forecast. For the first eight horizons, the random forest offers the largest reduction by a substantial margin. For the nowcast, all methods except the AR model, offer substantial reductions. For the backcast, finally, the dynamic factor model reduced the forecast standard deviation by the largest amount. Overall, for most methods the improvements in MSFE over the benchmark can be explained by an improvement in this component. The random forest is the only method that offers substantial reduction in biases and standard deviation 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.1	*0.17*	1.07	1.03	1.12	0.97	0.96	1.08	1.22	0.85
2Q.2	*0.17*	1.07	1.04	1.14	1.00	1.05	1.11	1.33	0.92
2Q.3	*0.17*	1.05	1.05	1.15	1.02	0.97	1.11	1.33	0.88
1Q.1	*0.17*	1.03	1.06	1.10	0.89	0.92	1.05	1.20	0.84
1Q.2	*0.17*	1.03	1.10	1.13	0.90	0.95	1.11	1.31	0.92
1Q.3	*0.16*	0.96	1.14	1.12	1.05	1.03	1.11	1.30	0.82
N.1	*0.16*	0.95	1.15	1.08	0.93	0.93	1.02	1.15	0.79
N.2	*0.16*	0.95	1.17	1.10	0.85	0.83	1.08	1.27	0.86
N.3	*0.16*	0.72	1.13	1.11	1.04	0.93	1.12	1.24	0.81
B.1	*0.16*	0.98	1.05	1.09	0.83	0.86	1.00	1.15	0.77
B.2	*0.16*	0.98	0.98	1.18	0.91	0.90	1.06	1.27	0.86

Note: See footnote of Table 2.

Table 6

Forecast standard deviation over the period 1992Q1-2018Q4

Model	Benchmark			Alternative					
	PM	AR	DFM	MIDAS	LASSO	EN	RS	RP	RF
2Q.1	*0.65*	1.03	1.02	1.01	1.00	1.01	0.99	1.01	0.98
2Q.2	*0.65*	1.03	1.02	1.00	1.00	0.99	1.01	1.00	0.95
2Q.3	*0.65*	1.03	1.01	1.00	0.99	0.99	0.99	0.99	0.93
1Q.1	*0.65*	1.03	1.00	1.01	0.99	0.99	1.00	1.00	0.95
1Q.2	*0.65*	1.03	0.99	0.99	0.99	0.99	0.99	0.98	0.95
1Q.3	*0.65*	1.03	0.98	0.96	1.01	1.00	0.96	0.97	0.93
N.1	*0.65*	1.01	0.97	0.97	0.94	0.95	0.95	0.98	0.92
N.2	*0.65*	1.01	0.94	0.94	0.94	0.92	0.92	0.95	0.90
N.3	*0.65*	0.97	0.86	0.92	0.86	0.86	0.89	0.91	0.88
B.1	*0.65*	1.00	0.81	0.96	0.84	0.86	0.90	0.95	0.89
B.2	*0.65*	1.00	0.77	0.97	0.82	0.86	0.89	0.94	0.88

Note: See footnote of Table 2.

Table 7
Forecast accuracy of forecast combinations

	PM	EA	IP	IP_{10}	OLS	OLS_{40}
2Q.1	*0.68*	0.99**	0.99**	0.99**	1.04	1.09
2Q.2	*0.68*	0.98**	0.98**	0.98**	1.04	1.13
2Q.3	*0.67*	0.97**	0.96**	0.97**	0.97	1.05
1Q.1	*0.67*	0.98*	0.98	0.98	1.13	1.24
1Q.2	*0.67*	0.97*	0.97*	0.97*	1.12	1.17
1Q.3	*0.67*	0.95**	0.95**	0.95**	0.99	1.07
N.1	*0.67*	0.94***	0.94***	0.94***	1.02	1.11
N.2	*0.67*	0.91***	0.91***	0.91***	1.04	1.08
N.3	*0.67*	0.87	0.86	0.87	0.93	0.95
B.1	*0.67*	0.87	0.86	0.87	0.88	0.92
B.2	*0.67*	0.85	0.84	0.85	0.81	0.83

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.

4.2 Forecast Combinations

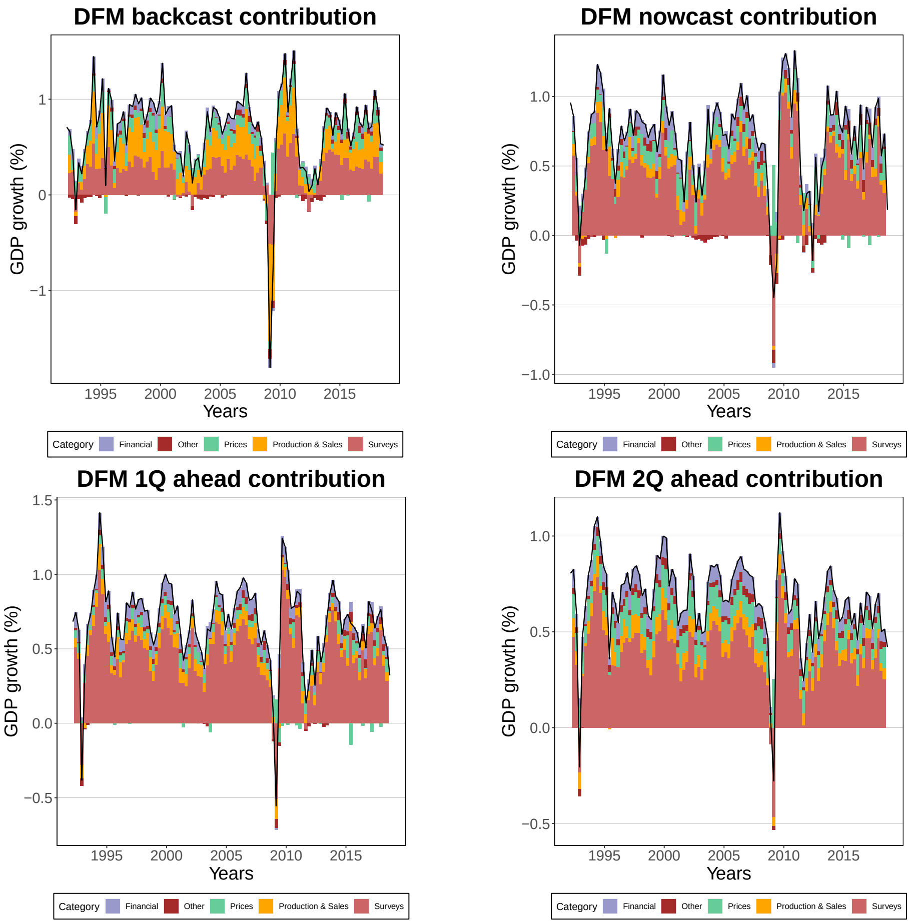
While the random forest tends to provide the most precise forecasts over many economics states, this is not uniformly so. It may therefore be worth to consider forecast combinations, where we combine the seven models using dimension reduction. Table 7 reports the results of the forecast combination schemes. The equally weighted forecast combinations and the combinations with weights inversely proportional to past MSFE are extremely similar. The combination scheme with weights based on an expanding window of MSFEs are marginally better than the other two. Compared to the forecasts in Table 2, the combinations provide forecasts that are very close to the model with the lowest MSFE for all horizons. In fact, they are the second most precise forecasts for nine horizons, for the third month nowcast, they are the most precise forecast and for the second month backcast the third most precise forecast. Using OLS weights, in contrast, results in less precise forecasts and nowcasts, which reiterates that forecast biases are not 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 contributions are averaged over the months in each quarter to keep the number of plots manageable.

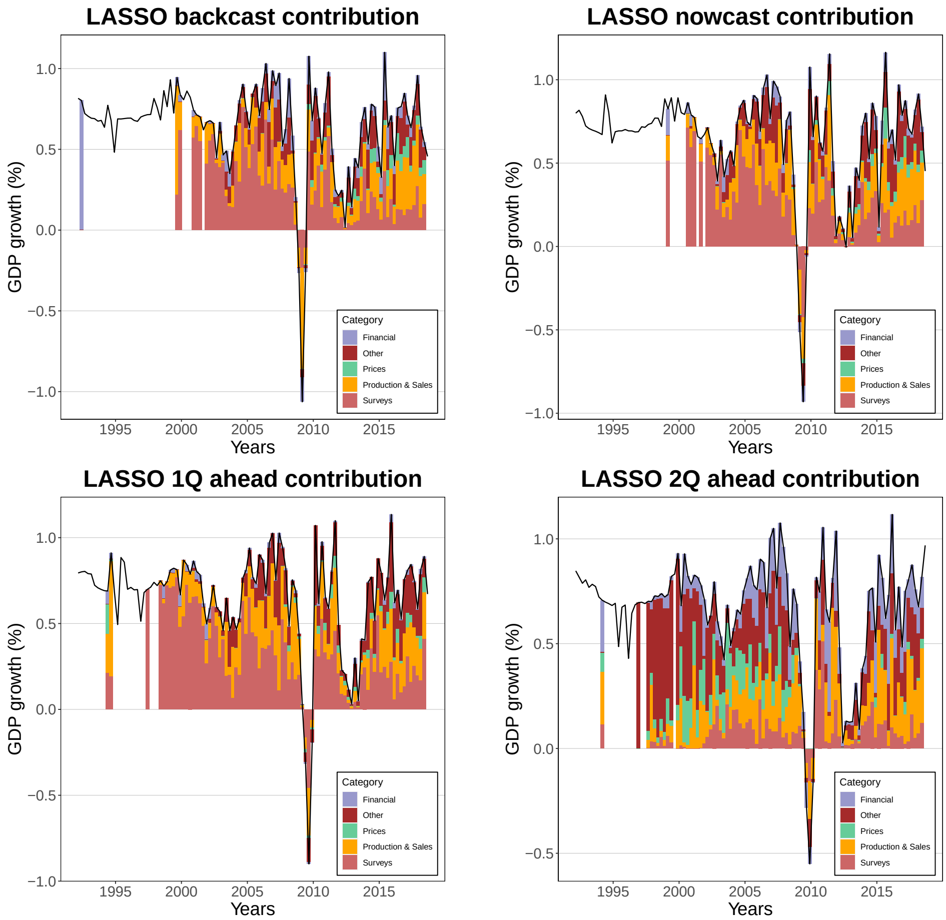
Figure 3
Contributions of the time series to dynamic factor model forecasts



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 graphs for the dynamic factor model in Figure 3 show that the time series in the *Surveys* category had the largest contribution to the forecasts with the relatively largest contribution in the one quarter ahead forecast as relatively more timely survey data arrive. For the nowcasts, the surveys still have the largest, if somewhat smaller, importance and for backcasts the arrival of hard data for the target quarter means that they now contribute more to the backcast. The importance of surveys for nowcasts is in line with the findings of the literature, for example, Giannone et al. (2008) and Bańbura and Rünstler (2011). Gayer and Marc (2018) suggest that the relationship between hard and soft predictors might have changed before and after the Financial Crisis but our results show no such change for the Netherlands.

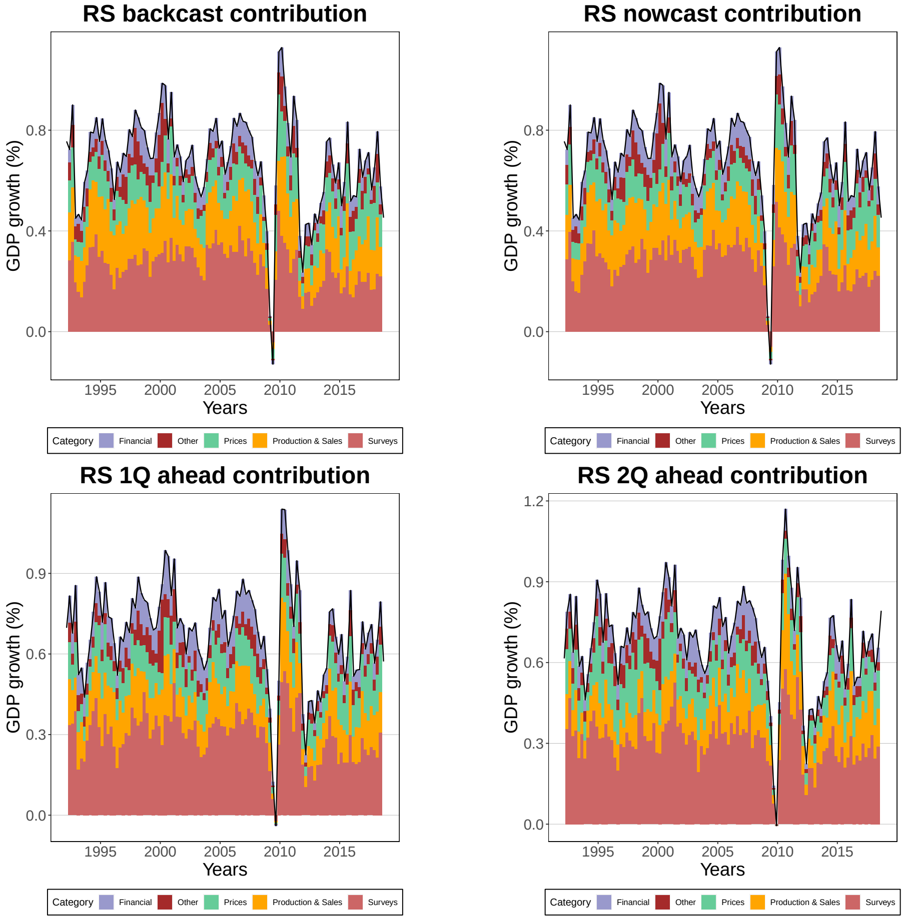
Figure 4
Contributions of the time series to LASSO forecasts



Note: See footnote of Figure 3.

Figure 4 reports the contributions of the predictor categories when forecasting with the LASSO. Three observations stand out. First, the contributions for 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. This also explains the results in Table 3, where the LASSO produced results with very similar precision to that of the prevailing mean forecast over the great moderation period. Second, in the subsequent periods, the LASSO makes more varied use of the information that also fluctuated substantially over time. The two-quarter ahead forecasts makes little use of surveys and relies mainly on hard data. Surveys are more important in one-quarter ahead forecasts, nowcasts and backcasts. However, their importance seemed to decline already before the financial crisis.

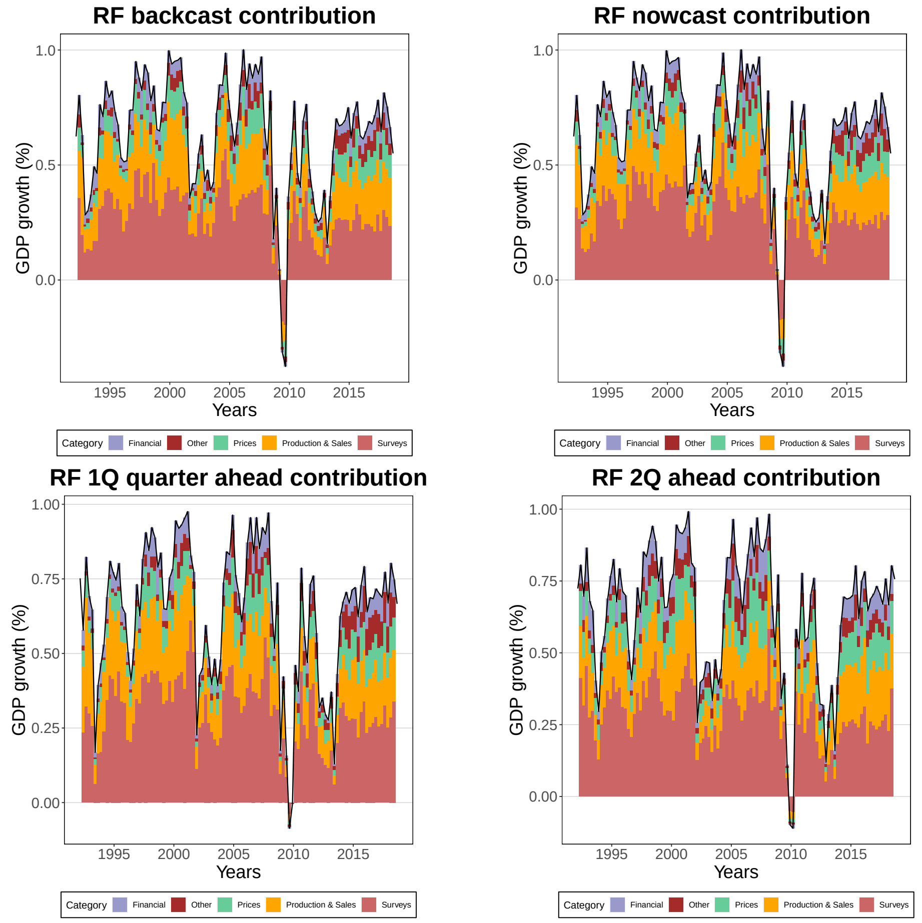
Figure 5
Contributions of the time series to random subset regression forecasts



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

The contributions when using random subset regression shown in Figure 5 are considerably more equal for the different categories and relative constant throughout time and forecast horizon. *Surveys* are the most important category, followed by *Production & Sales* and *Prices*.

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



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

A similar picture emerges for the random forest. Figure 5 displays a more balanced and steady use of the predictors in the different categories. The importance of survey data, which had a sizable but non-dominant role before the crisis, remained strong. predictors in the *Production & Sales* category play an evenly large role and the remaining predictors also contribute throughout. Again, the difference between the forecast horizons is minimal.

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, over the entire forecast period, the random forest delivers the most accurate forecast and nowcasting, whilst the dynamic factor model has the highest forecasting accuracy for backcasting, especially during recessions. Regularization methods perform very well, during the financial crisis in particular. Since the financial crisis, the random forest has provided the most precise forecast and nowcasts by, for some horizons, substantial margins.

References

Babii, A., Ghysels, E., and Stiaukas, J. (2022). Machine learning times series regression with an application to nowcasting. *Journal of Business and Economic Statistics*, 40(3):1094–1106.

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*, 41(6):1181–1200.

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.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):147–162.

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.	predictor	transformation				source	start	end	link	publ. delay
		ln.	dif.	fil.	sa.					
1	Av. daily prod. - prod. industries	1	1	3	NSA	CBS	jan 1965	jan 2019	link	2
2	Av. daily prod. - industry	1	1	3	NSA	CBS	jan 1965	jan 2019	link	2
3	Ind. prod. - cap. goods industry	1	1	3	NSA	ECB	jan 1970	jan 2019	restricted	2
4	Cons. exp. - households, dom. cons.	1	1	3	NSA	CBS	feb 1977	jan 2019	link	2
5	Ind. prod. - Manufacture tobacco	1	1	3	NSA	ECB	jan 1965	jan 2019	restricted	2
6	Ind. prod. - Manufacture wearing apparel	1	1	3	NSA	ECB	jan 1970	jan 2019	restricted	2
7	Ind. prod. - Manufacture motor vehicles/(semi-)trailers	1	1	3	NSA	ECB	jan 1985	jan 2019	restricted	2
8	Ind. prod. - Manufacture other transport equipment	1	1	3	NSA	ECB	jan 1970	jan 2019	restricted	2
9	Ind. prod. - Manufacturing	1	1	3	SA	ECB	dec 1979	jan 2019	restricted	2
10	Ind. prod. - Manufacture of textiles	1	1	3	SA	ECB	jan 1980	jan 2019	restricted	2
11	Ind. prod. - Printing/reproduction of recorded media	1	1	3	SA	ECB	jan 1980	jan 2019	restricted	2
12	Ind. prod. - Constr.	1	1	3	SA	ECB	jan 1985	jan 2019	restricted	2
13	Ind. prod. - MIG capital goods ind.	1	1	3	SA	ECB	jan 1970	jan 2019	restricted	2
14	Belgium, Retail trade excl. fuel, motor vehicles/cycles	1	1	3	SA	ECB	jan 1970	jan 2019	restricted	2
15	Germany, Total ind. (excl. constr.)	1	1	3	SA	ECB	jan 1965	jan 2019	restricted	2
16	Germany, Retail trade excl. fuel, motor vehicles/cycles	1	1	3	SA	ECB	jan 1968	jan 2019	restricted	2
17	Spain, Total ind. (excl. constr.)	1	1	3	SA	ECB	jan 1965	jan 2019	restricted	2
18	France, Total ind. (excl. constr.)	1	1	3	SA	ECB	jan 1965	jan 2019	restricted	2
19	France, Retail trade excl. fuel, motor vehicles/cycles	1	1	3	SA	ECB	jan 1970	jan 2019	restricted	2
20	Italy, Total ind. (excl. constr.)	1	1	3	SA	ECB	jan 1965	jan 2019	restricted	2
21	Germany, Total ind.	1	1	3	SA	ECB	mrt 1978	jan 2019	restricted	2
II. Surveys (N = 36)										

Table continued on next page

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

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

Table continued on next page

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

No.	predictor	transformation				source	start	end	link	publ. delay
		ln.	dif.	fil.	sa.					
48	Italy, Cons. confidence	0	1	3	SA	ES	jan 1985	feb 2019	link	1
49	Belgium, Ind. confidence	0	1	3	SA	ES	jan 1985	feb 2019	link	1
50	Germany, Ind. confidence	0	1	3	SA	ES	jan 1985	feb 2019	link	1
51	Italy, Ind. confidence	0	1	3	SA	ES	jan 1985	feb 2019	link	1
52	United Kingdom, Ind. confidence	0	1	3	SA	ES	jan 1985	feb 2019	link	1
53	United Kingdom, Cons. confidence	0	1	3	SA	ES	jan 1985	feb 2019	link	1
54	Ind. Confidence - (CBS definition)	0	1	3	SA	CBS	jan 1985	feb 2019	link	1
55	Ind. Confidence - Prod. expect, months ahead (CBS definition)	0	1	3	SA	CBS	jan 1985	feb 2019	link	1
56	Ind. Confidence - Ass. of order-book levels (CBS definition)	0	1	3	SA	CBS	jan 1985	feb 2019	link	1
57	Ind. Confidence - Ass. of stocks of fin. products (CBS definition)	0	1	3	SA	CBS	jan 1985	feb 2019	link	1
III. Financial (N = 8)										
58	Loans to the private sector	1	1	3	NSA	ECB	dec 1982	jan 2019	restricted	2
59	M1	1	2	3	NSA	ECB	jan 1980	jan 2019	restricted	2
60	M3 (money in circulation inclusive)	1	2	3	NSA	ECB	jan 1970	jan 2019	restricted	2
61	Interest rate (short term) - euro	0	1	3	NSA	DNB	nov 1984	feb 2019	restricted	1
62	Loans on mortgage (nominal rate 5 to 10 years mortgage)	0	1	3	NSA	ECB	jan 1980	jan 2019	link	2
63	Interest rate (long term)	0	1	3	NSA	DS	jan 1965	mrt 2019	NLGBD10	0
64	Share index, AEX	1	1	3	NSA	DS	jan 1983	mrt 2019	AMSTEOE	0
65	Share index, Amsterdam Midkap-index	1	1	3	NSA	DS	jan 1983	mrt 2019	AMSMKAP	0
IV. Prices (N = 14)										
66	Exchange rate, US-Dollar per Euro	0	1	3	NSA	ECB	jan 1965	feb 2019	restricted	1
67	Housing price	1	2	3	NSA	CBS	jan 1976	feb 2019	link	1
68	Consumerprice index, total CPI, all households	1	2	3	NSA	CBS	jan 1965	feb 2019	link	1
69	Consumerprice index, underlying inflation	1	2	3	NSA	CBS	jan 1976	feb 2019	link	1
70	World market commodity prices, overall	1	2	3	NSA	HWWI	sep 1978	feb 2019	link	1
71	World market commodity prices, industrial materials	1	2	3	NSA	HWWI	sep 1978	feb 2019	link	1

Table continued on next page

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

No.	predictor	transformation				source	start	end	link	publ. delay
		ln.	dif.	fil.	sa.					
72	World market commodity prices, agric. & ind. materials	1	2	3	NSA	HWWI	sep 1978	feb 2019	link	1
73	World market commodity prices, metals	1	2	3	NSA	HWWI	sep 1978	feb 2019	link	1
74	World market commodity prices, energy-components	1	2	3	NSA	HWWI	sep 1978	feb 2019	link	1
75	Producer prices, total intermed. & fi. products (dom. market)	1	2	3	NSA	CBS	jan 1981	jan 2019	link	2
76	Producer prices, consumer goods (dom. market)	1	2	3	NSA	ECB	jan 1976	jan 2019	restricted	2
77	Producer prices, intermediate goods (dom. market)	1	2	3	NSA	ECB	jan 1976	jan 2019	restricted	2
78	Producer prices, intermediate & final products (for. market)	1	2	3	NSA	CBS	jan 1981	jan 2019	link	2
79	Producer prices, energy (dom. market)	1	2	3	NSA	ECB	jan 1980	jan 2019	restricted	2
V. Other ($N = 4$)										
80	Unemployment	0	1	3	SA	ES	jan 1983	feb 2019	link	1
81	Issued vehicle registration certificates	1	1	3	NSA	RAI	jan 1965	feb 2019	link	1
82	Bankruptcies	1	1	3	NSA	CBS	jan 1965	feb 2019	link	1
83	Hourly wages (collective labour agreement), industry	1	1	3	NSA	CBS	jan 1972	feb 2019	link	1
Quarterly variables ($N = 1$)										
84	Gross domestic product (GDP)	1	1	3	SA	CBS	1970Q1	2018Q4	link	3

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, RAI = RAI Association; start: Starting year and month of the series, end: Final year and month of the series; code: link = link to the data, restricted = series not publicly available, code = Refinitiv code (only available for subscribed users); publ. delay: publication delay of the series in months.

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

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

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

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

where $\boldsymbol{\lambda}$ is the vector of loading coefficients of the factors on latent GDP growth. 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, $\boldsymbol{\Lambda}$, via a static principal components analysis applied to a balanced sub-sample of the data, where observations in periods with missing data are discarded. In our data set, however, only the rows of the first few observations are discarded, which are missing as the result of vertical alignment due to 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 predictors 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 predictor 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 obtain estimates of the factors via principal components on a skip sampled data set including the lags of the predictors.

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 predictor selection by imposing an ℓ_1 penalty in the estimation of the coefficients. As the response predictor and predictors are of different frequencies, the mixed-data sampling approach of Section 2.1 is employed using skip sampling for the monthly predictors.

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 $\mathbf{y_h}$ is GDP growth for forecasting horizon h, ι_N an $N \times 1$ vector of ones, $\mathbf{x^{(3)}}$ a matrix of the skip-sampled versions of $x_{i,m}$, $\boldsymbol{\beta}$ a vector of coefficients, $\|\cdot\|_p$ denotes the ℓ_p, 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.

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

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

α determines the relative extent of regularization performed by both norms and is 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 predictors out of the varible pool. The forecasts are then averaged. If, however, the predictor pool is large the number of combinations of k predictors is prohibitively large. A solution is to take R randomly chosen subsets of predictors. 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 predictor in the random subset regression is the j-th predictor.

A tuning parameter of this method is the size of each predictor 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 predictors would be to combine the predictors 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 predictors that are weighted averages of the n predictors. 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 predictor k needs to be determined. Again, we average nowcasts over those obtained using $k = 2, 3, 4, 5$.

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 skip-sampled predictor that is minimizing the cost function is chosen at each node, i.e.

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

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 predictors 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.