TECHNICAL REPORT

DFROG: The nowcasting model for GDP growth of De Nederlandsche Bank*

Dorinth van Dijk
De Nederlandsche Bank

Mick van Rooijen
De Nederlandsche Bank

Jasper de Winter†
De Nederlandsche Bank

This version: October 2, 2024

Abstract

This paper presents DFROG, the "nowcasting" model employed by De Nederlandsche Bank (DNB) to generate near-term forecasts of the quarter-on-quarter growth rate of Gross Domestic Product (GDP). The core concept of DFROG is that the comovement of a potentially large set of monthly economic indicators can be summarized into a few factors, which can then be used to forecast GDP growth. We compare the forecast accuracy of DFROG with several benchmark models and professional analysts, and conduct a thorough review of the model's optimal specification.

* We gratefully acknowledge comments from Peter van Els. The views expressed herein are those of the authors and do not necessarily reflect the position of De Nederlandsche Bank. All remaining errors are our own.

† Corresponding author: j.m.de.winter@dnb.nl.

1 Introduction

Economists face considerable challenges in accurately assessing the economy because the key indicator, Gross Domestic Product (GDP), is released with a significant delay. In the Netherlands, for instance, Statistics Netherlands publishes an initial growth rate estimate 45 days after the end of a quarter. This initial figure is often subject to revisions, adding to the complexity. Additionally, GDP data is only available on a quarterly basis. Unlike weather forecasters who predict tomorrow's weather based on today's data, economists must forecast the future, the present, and even the recent past simultaneously. Fortunately, there is a stream of economic indicator releases, financial market information and survey data releases that can be used to form a view of the current economic stance. De Nederlandsche Bank (DNB) –like many central banks– has been using a so so-called "nowcasting" model, to translate this incoming stream of information into a forecast of the upcoming GDP release. This model is called The Dutch forecasting model for real-time output growth (DFROG). Like all nowcasting models, DFROG is a mechanical model while DNB's official view on the economy is also influenced by expert judgment. The DFROG forecasts are routinely used within DNB, but especially round new GDP releases and at the start of DNB's Spring and Autumn projections (see e.g. here for DNB Spring projections June 2024).

We recently revised DFROG. This report describes how we use the current model and our decisions on the precise model-specification. Furthermore, we show how accurate the forecasts of our nowcasting model are in comparison to other often used models. The core idea of DFROG is that the co-movement of a set of carefully selected monthly indicators can be summarized in one or more factors that can be used to forecast GDP growth. This type of model has been shown to produce relatively accurate forecasts of Dutch GDP growth in the recent past (see e.g Jansen et al., 2016, Jansen and de Winter, 2018, Hindrayanto et al., 2016), also compared to recently popularized off-the-shelf machine learning models (see Kant et al., 2022).

Apart from revising our nowcasting model, we also rebuild our indicator database. We searched numerous publicly available databases for relevant indicators of the Dutch economy. We combine these with proprietary indicators from data suppliers and sentiment indicators derived from the leading financial newspaper in the Netherlands, Het Financieele Dagblad. The sentiment indicators were developed in co-operation with Het Financieele Dagblad, the technical background and findings are detailed in Dijk, van and de Winter (2023).

The remainder of the paper is organized as follows. Section 2 provides a detailed description of the DFROG model currently in use. Section 3 describes the dataset used in our analysis. Section 4 describes the setup and the outcome of the model specification horse race. Section 5 compares the forecast accuracy of DFROG to a couple of well known benchmarks and forecasts of professional analysts. Section 6 describes how we use DFROG in practice.

2 Description DFROG

2.1 The nowcasting model

Let $y_t = [y_{1,t}, \ldots, y_{n,t}]'$, $t = 1, \ldots, T$, denote the n-dimensional vector of monthly variables, standardized to mean 0 and unit variance. We assume that y_t is driven by a few unobserved factors, that can be described using the following factor model representation:

$$y_t = \Lambda f_t + \epsilon_t \quad \text{for } i = 1, \ldots, n. \tag{1}$$

Here, f_t is a $r \times 1$ vector of unobserved common factors, Λ is an $n \times r$ matrix of time-invariant factor loadings and $\epsilon_t = [\epsilon_{1,t}, \epsilon_{2,t}, \ldots, \epsilon_{n,t}]'$ is a vector of idiosyncratic components. The idiosyncratic errors $\epsilon_{i,t}$ capture the movements that are specific to the individual series. Regarding their dynamics, we assume the idiosyncratic components follow an autoregressive (AR) process of order 1:

$$\epsilon_{i,t} = \rho_i \epsilon_{i,t-1} + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2_{\varepsilon_i}) \quad \text{for } i = 1, \ldots, n \tag{2}$$

where ρ_i is the autoregressive coefficient for indicator i. The common factors f_t are modeled as a vector autoregressive (VAR) process of order p

$$f_t = A_1 f_{t-1} + \cdots + A_p f_{t-p} + u_t, \quad u_t \overset{i.i.d.}{\sim} \mathcal{N}(0, Q) \tag{3}$$

where $A_1, \ldots, A_p$ are $r \times r$ matrices of autoregressive coefficients for the factors, and Q is a diagonal covariance matrix.

Equations 1–3 can be cast into a state-space representation where the common factors and the idiosyncratic components are the unobserved states. Equation 1 is known as the *measurement equation* and links the data to the unobserved states. Equations 2 and 3 are known as the *transition equations*. The outcomes of the model are illustrated in Section 6. We extensively test for the optimal specification in terms of forecast accuracy, described in Section 4. Bańbura and Modugno (2014), Bok et al. (2018) and Doz et al. (2012) establish the viability of this approach. They show that, if the factor structure is strong, maximum likelihood estimators are consistent when the sample size T and the cross-sectional dimension n are large. The estimator is also robust to cross-sectional misspecification, time-series correlation of the idiosyncratic components, and non-Gaussianity. In practice, the estimates can be conveniently computed iteratively using the Kalman smoother and the Expectation Maximization (EM) algorithm.

The results in Doz et al. (2012) also imply that estimating the AR(1) process in Equation 3 is not strictly necessary. However, the choice for an autoregressive process for the error terms can be justified on several grounds. First, most macroeconomic variables are serially correlated. Therefore, it would be inappropriate to assume uncorrelated errors here (Shapiro et al., 2002). In that case, imposing an assumption of no serial correlation is quite restrictive because it is only valid asymptotically. Secondly, the AR process can improve the forecasting performance of the model as it enables to forecast the idiosyncratic component (Bańbura and Modugno, 2014). Finally, it may lead to more efficient estimates of common factors in the case of missing values at the end of our sample (so called ragged edges).

2.2 State space representation

In what follows, we cast the nowcasting model specified in the preceding Section into a state space representation. In the quasi-maximum likelihood approach, the state space system is used in combination with the Kalman framework to evaluate the likelihood and estimate the parameters.

We follow Bańbura and Modugno (2014) and add the idiosyncratic component to the state vector. More precisely, we assume that $\epsilon_{i,t}$ in Equation 1 can be decomposed as:

$$\begin{aligned} \epsilon_{i,t} &= \tilde{\epsilon}_{i,t} + \xi_{i,t}, \quad \xi_{i,t} \overset{i.i.d.}{\sim} \mathcal{N}(0,\kappa) \\ \tilde{\epsilon}_{i,t} &= \rho_i \tilde{\epsilon}_{i,t-1} + e_{i,t}, \quad e_{i,t} \overset{i.i.d.}{\sim} \mathcal{N}(0,\sigma_i^2) \end{aligned} \tag{4}$$

where both $\xi_t = [\xi_{1,t}, \ldots, \xi_{n,t}]'$ and $\tilde{\epsilon}_t = [\tilde{\epsilon}_{1,t}, \ldots, \tilde{\epsilon}_{n,t}]'$ are cross-sectionally uncorrelated and κ is a very small number. Combining Equations 1, 2 and 4 results in the following state space representation:

$$\begin{aligned} y_t &= \tilde{\Lambda}\tilde{f}_t + \xi_t, \quad \xi_t \sim \mathcal{N}(0,\tilde{R}) \\ \tilde{f}_t &= \tilde{A}\tilde{f}_{t-1} + \tilde{u}_t, \quad \tilde{u}_t \sim \mathcal{N}(0,\tilde{Q}) \end{aligned} \tag{5}$$

where

$$\tilde{f}_t = \begin{bmatrix} f_t \\ \tilde{\epsilon}_t \end{bmatrix}, \tilde{u}_t = \begin{bmatrix} u_t \\ \epsilon_t \end{bmatrix}, \tilde{\Lambda} = \begin{bmatrix} \Lambda & I \end{bmatrix}, \tilde{A} = \begin{bmatrix} A & 0 \\ 0 & \text{diag}(\rho_1, \ldots, \rho_n) \end{bmatrix}, \tilde{Q} = \begin{bmatrix} Q & 0 \\ 0 & \text{diag}(\sigma_1^2, \ldots, \sigma_n^2) \end{bmatrix},$$

$e_t = [e_{1,t}, \ldots, e_{n,t}]'$ and $\tilde{R}$ is a fixed diagonal matrix with κ on the diagonal.

A notable challenge for nowcasting models is the discrepancy between quarterly GDP growth figures and the monthly frequency of macroeconomic time series. Restricting our dataset to quarterly variables would render the model inadequate for real-time forecasting. Real-time forecasting requires regularly updated information, much of which is received monthly. Fortunately, mixed frequency datasets can be integrated into the factor model by treating lower frequency series as high-frequency indicators with missing data. As a result, information from indicators collected at a lower frequency (mainly GDP) can still be used to estimate the factors. The model can also be employed to forecast the lower frequency series or enhance their interpolation. To achieve this, we represent the quarterly variables in our model as partially observed monthly variables. Initially, we use the quarterly GDP in volume terms to define its monthly counterparts as follows:

$$GDP_t^Q = GDP_t^M + GDP_{t-1}^M + GDP_{t-2}^M \quad \text{for } t = 3, 6, 9, \ldots \tag{6}$$

Additionally, we specify the transformations:

$$Y_t^Q = 100 \times \log\left(GDP_t^Q\right) \tag{7}$$
$$Y_t^M = 100 \times \log\left(GDP_t^M\right) \tag{8}$$

Using Equation 6 and 8, we can define the monthly and quarterly GDP growth rates as follows:

$$y_t^Q = Y_t^Q - Y_{t-3}^Q \tag{9}$$
$$y_t^M = Y_t^M - Y_{t-1}^Q \tag{10}$$

Next, we define the monthly representation of the quarterly growth rate ($\bar{y}_t^Q$) as:

$$\bar{y}_t^Q = \begin{cases} Y_t^Q - Y_{t-3}^Q & \text{if } t = 3, 6, 9, \ldots \\ \text{NA} & \text{otherwise} \end{cases} \tag{11}$$

In $\bar{y}_t^Q$ the quarterly growth rates are in fact are 'assigned' to the third month of each quarter, following the usual convention in the nowcasting literature. To bridge $\bar{y}_t^Q$ to y_t^Q we follow the approximation developed by Mariano and Murasawa (2003), i.e.:

$$\bar{y}_t^Q = Y_t^Q - Y_{t-3}^Q = (Y_t^M + Y_{t-1}^M + Y_{t-2}^M) - (Y_{t-3}^M + Y_{t-4}^M + Y_{t-5}^M) \tag{12}$$
$$= y_t^M + 2y_{t-1}^M + 3y_{t-2}^M + 2y_{t-3}^M + y_{t-4}^M \tag{13}$$

where y_t^M and $\bar{y}_t^Q$ denote the $n_M \times 1$ and $n_Q \times 1$ vectors of monthly and quarterly data, respectively. Further, let Λ_M and Λ_Q denote the corresponding factor loadings for the monthly y_t^M, and for the unobserved monthly growth rates of the quarterly y_t^Q, respectively. We can then cast the model in the following state space representation. The measurement equation is defined as:

$$
\begin{bmatrix} y_t^M \\ \bar{y}_t^Q \end{bmatrix} = \begin{bmatrix} \Lambda_M & 0 & 0 & 0 & 0 & I_n & 0 & 0 & 0 & 0 & 0 \\ \Lambda_Q & 2\Lambda_Q & 3\Lambda_Q & 2\Lambda_Q & \Lambda_Q & 0 & I_{n_Q} & 2I_{n_Q} & 3I_{n_Q} & 2I_{n_Q} & I_{n_Q} \end{bmatrix} \begin{bmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \\ \varepsilon_t^M \\ \varepsilon_t^Q \\ \varepsilon_{t-1}^Q \\ \varepsilon_{t-2}^Q \\ \varepsilon_{t-3}^Q \\ \varepsilon_{t-4}^Q \end{bmatrix} + \begin{bmatrix} \xi_t^M \\ \xi_t^Q \end{bmatrix} \tag{14}
$$

and the transition equation:

$$
\begin{bmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \\ \tilde{\epsilon}_t^M \\ \tilde{\epsilon}_t^Q \\ \tilde{\epsilon}_{t-1}^Q \\ \tilde{\epsilon}_{t-2}^Q \\ \tilde{\epsilon}_{t-3}^Q \\ \tilde{\epsilon}_{t-4}^Q \end{bmatrix} = \begin{bmatrix} A_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ I_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \rho_M & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \rho_Q & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I_{n_Q} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I_{n_Q} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I_{n_Q} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I_{n_Q} & 0 \end{bmatrix} \begin{bmatrix} f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \\ f_{t-5} \\ \tilde{\epsilon}_{t-1}^M \\ \tilde{\epsilon}_{t-1}^Q \\ \tilde{\epsilon}_{t-2}^Q \\ \tilde{\epsilon}_{t-3}^Q \\ \tilde{\epsilon}_{t-4}^Q \\ \tilde{\epsilon}_{t-5}^Q \end{bmatrix} + \begin{bmatrix} u_t \\ 0 \\ 0 \\ 0 \\ 0 \\ e_t^M \\ e_t^Q \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
$$

Here, $\rho_M = \text{diag}(\rho_{M,1}, \ldots, \rho_{M,n_M})$ and $\rho_Q = \text{diag}(\rho_{Q,1}, \ldots, \rho_{Q,n_Q})$ collect the AR(1) coefficients of the idiosyncratic component of monthly and quarterly data. As described above ξ_t^M and ξ_t^Q have fixed and small variances.[1]

The state-space form allows inference using the Kalman filter and smoother. The state space framework also provides a convenient framework for handling the irregularities of the data in real time (i.e., mixed frequencies and non-synchronicity of the data

[1] Notice that the size of the state-vector quickly expands when the VAR order increases. Notice that this expansion comes at exponentially increasing computational time.

releases) and updating the predictions. The Kalman filter processes incoming data in a clear and intuitive manner. It updates model predictions recursively by weighting the innovation components of new data based on their timeliness and quality. Additionally, because the model generates forecasts for all variables simultaneously, analyzing the data flow doesn't require combining multiple unrelated models.

2.3 Model estimation

We estimate the dynamic factor model by quasi-maximum likelihood proposed by Bańbura and Modugno (2014). To summarize, the algorithm starts by computing principal components, and the model parameters are estimated using ordinary least squares (OLS) regression, treating these principal components as the true common factors. This approach is particularly effective for large datasets, as principal components can act as suitable initialization of common factors. This initialization method requires a balanced panel, which we create as follows.

Firstly, we treat quarterly variables as partially observed monthly variables, as described in Equation 11. Next, we remove the last three months of data, based on the time-series with the longest history. We then fill in any remaining missing monthly values using cubic-spline interpolation. For monthly variables with missing values at the beginning or end of the sample, we first fill in the median values and then apply a moving average filter. The filtered series are used to fill in the unobserved gaps. The initial values of the factors are extracted from this balanced panel using principal components.[2] As such, this initialization can greatly reduce computation time, making the method feasible for large panels.

Secondly, the Kalman filter and smoother are used to obtain an updated estimate of the common factors, based on the principal components and OLS estimates from the first step. Stopping at this point provides the two-step estimate of the common factors, as used by Giannone et al. (2008) and studied by Doz et al. (2012). Maximum likelihood estimation is achieved by iterating these two steps until convergence, using the Expectation Maximization (EM) algorithm. The EM algorithm accounts for uncertainty in the factor estimates at each step. The algorithm is described in Bańbura and Mod-

[2] It has been shown that the principal components can provide consistent estimates of factors as both n and T grow large (Shapiro et al., 2002).

ugno (2014). As explained further in Section 4.1, the nowcasting model is re-estimated monthly. Regarding our convergence criterion, we stop the algorithm either after a maximum number of iterations or when the increase in the likelihood between consecutive runs becomes minimal. Specifically, we use the following metric as stopping criterion:

$$c_m = \frac{l(\Omega_w, \theta(m)) - l(\Omega_w, \theta(m-1))}{\frac{1}{2}(|l(\Omega_w, \theta(m))| + |l(\Omega_w, \theta(m-1))|)} \tag{15}$$

Here, $l(\Omega, \theta(m))$ denotes the log-likelihood iteration j of the EM-algorithm, $\theta(m)$ is the estimated coefficients after iteration m and Ω_w is the information set of monthly and quarterly indicators in week w. The algorithm is stopped at M when $c_M < 10^{-5}$ or $M = 500$, whichever condition is met first.

2.4 Forecast contributions of individual variables

The contribution of each variable to the GDP forecast can be derived using the algorithm described in Koopman and Harvey (2003) and applied to dynamic factor models in Bańbura and Rünstler (2011). This Section briefly recapitulates the idea of assigning implicit weights to individual variables in making GDP forecasts.

The vector y_t defined in Equation 1 contains GDP and all other variables from which factors are extracted. In the most generalised state space form, y_t can be expressed as:

$$\begin{aligned} y_t &= W(\theta)\alpha_t + \xi_t, \quad \xi_t \sim \mathcal{N}(0, \Sigma_\xi(\theta)) \\ \alpha_t &= T(\theta)\alpha_{t-1} + \tilde{u}_t, \quad \tilde{u}_t \sim \mathcal{N}(0, \Sigma_{\tilde{u}}(\theta)) \end{aligned} \tag{16}$$

where $\theta = (\Lambda_M, \Lambda_Q, A_1, ..., A_p, \rho_M, \rho_Q, \tilde{R}, \tilde{Q})$ denotes the set of model parameters and errors ξ_t and $\tilde{u}_t$ are defined in Equation 5. The Kalman smoother provides the smoothed estimate of the state vector $a_{t|T} = \mathbb{E}[\alpha_t | y_T]$ conditional on the data. Each individual element of the smoothed state vector $a_{t|T}$ can be decomposed into a weighted sum of all observations through time. The weighted sum is given by:

$$a_{t|T} = \sum_{j=1}^{T} w_j(a_{t|T}) y_j \tag{17}$$

where the weight matrices $w_j(a_{t|T})$ are computed using the algorithm from Koopman and Harvey (2003). Note that the sum in Equation 17 has subscript j, which indicates that the weights are computed after the Kalman smoother has been applied for $t = 1, ..., T$ to get $a_{t|T}$. Hence, for each point in time t we sum all observations in y_j given weight w_j for $j = 1, ..., T$. Naturally, when j gets closer to t, the weight of observations y_j tend to get larger.

Now that we have split up each element of smoothed state vector $a_{t|T}$ into a weighted sum of observables, we can link the weighted sum to the GDP forecast $\bar{y}_t^{GDP} \in \bar{y}_t^Q$ defined in Equation 14. To do so, we plug in the smoothed state vector in Equation 14 and define the forecast contributions as:

$$\bar{y}_t^{GDP} = \lambda a_{t|T} = \lambda \sum_{j=1}^{T} w_j(a_{t|T}) y_j \tag{18}$$

where λ is the loading vector associated to GDP and is part of the loading matrix Λ^Q from Equation 14. Since $\bar{y}_t^{GDP}$ is a normalised value, we multiply the terms in Equation 18 by the standard deviation of GDP and add the mean to revert the GDP forecast to the original units.

Note that the state vector α_t contains both the factors and the idiosyncratic component, as denoted in Equation 14. Hence, the estimated smoothed state vector $a_{t|T}$ (and its equivalent of weighted observables $\sum_{j=1}^{T} w_j(a_{t|T}) y_j$) includes both the factors and the idiosyncratic component. The contribution of a variable to the GDP forecast is the sum of the individual elements in the state vector, scaled by the loading vector for GDP.

2.5 News in forecast revisions

Following Bańbura et al. (2011), we update our GDP forecast on a regular basis and assess the impact of new data releases on the forecast. The state space framework also provides forecasts for all other variables than GDP. Hence, we can extract the unexpected component from the data release and its effect on the GDP forecast. This, we define as news. This framework is used to understand the changes in GDP forecasts when new data come in.

We consider two data vintages $\mathbf{\Omega}_v$ and $\mathbf{\Omega}_{v+1}$ of the same set of variables retrieved at dates v and $v+1$. The data vintage $\mathbf{\Omega}_{v+1}$ contains a set of new observations that are not available in $\mathbf{\Omega}_v$. The new observations are denoted by $\{y_{i_j,t_j}, j = 1, ..., J_{v+1}\}$ and tend to contain at least some new information.[3] We assume that data are not revised and that the incoming new part of the information set is orthogonal to the current information set. Hence, we can write:

$$\mathbb{E}[\bar{y}_t^{GDP}|\boldsymbol{\Omega}_{v+1}] = \mathbb{E}[\bar{y}_t^{GDP}|\boldsymbol{\Omega}_v] + \mathbb{E}[\bar{y}_t^{GDP}|I_{v+1}] \tag{19}$$

where

$$I_{v+1} = [I_{v+1,1}, ..., I_{v+1,J_{v+1}}]', \quad I_{v+1,j} = y_{i_j,t_j} - \mathbb{E}[y_{i_j,t_j}|\boldsymbol{\Omega}_v],\ j = 1, ..., J_{v+1}.$$

The vector I_{v+1} represents the part of the release in each variable y_{i_j,t_j} that is unexpected based on the information in $\boldsymbol{\Omega}_v$. This unexpected part is referred to as the surprise in the data release. Note that if a new observation in $\boldsymbol{\Omega}_{v+1}$ is exactly equal to what is expected based on $\boldsymbol{\Omega}_v$, there is no surprise and the GDP forecast will not change.

The aim is to compute the second term of Equation 19, which is the forecast revision caused by the incoming data. We define a vector $B_{v+1} = [b_{v+1,1}, ..., b_{v+1,J_{v+1}}]$ that links the surprise in the data release of each variable to the new GDP forecast. The forecast revision can be written as:

$$\underbrace{\mathbb{E}[\bar{y}_t^{GDP}|\boldsymbol{\Omega}_{v+1}] - \mathbb{E}[\bar{y}_t^{GDP}|\boldsymbol{\Omega}_v]}_{\text{forecast revision}} = B_{v+1}I_{v+1} = \sum_{j=1}^{J_{v+1}} b_{v+1,j}(\underbrace{y_{i_j,t_j} - \mathbb{E}[y_{i_j,t_j}|\boldsymbol{\Omega}_v]}_{\text{surprise}}). \tag{20}$$

As showed by Bańbura et al. (2011), both matrices B_{v+1} and I_{v+1} can be computed using the Kalman filter and smoother.

Note that the forecast of $\bar{y}_t^{GDP}$ not only depends on a given data vintage $\boldsymbol{\Omega}_v$ or $\boldsymbol{\Omega}_{v+1}$, but also on a parameter vector θ that depends on one of the two data vintages. To identify the news impact of each variable, we should keep the parameter vector θ constant. We estimate the parameter vector θ using $\boldsymbol{\Omega}_{v+1}$ and apply those parameter estimates to compute $\mathbb{E}[\bar{y}_t^{GDP}|\boldsymbol{\Omega}_v]$.

In conclusion, the revision of the GDP forecast is decomposed as a weighted sum of surprises in the most recent data vintage. The news in the GDP forecast depends on the size of the surprises captured in I_{v+1}, as well as on the relevance of the variables in

[3] Note that we use subscript j to indicate that variables might have a different publication days within a month, but can apply to the same month t.

forecasting GDP indicated by B_{v+1}.

3 Data

3.1 Data selection

We construct a small, medium-sized and large dataset for our analysis by meticulously examining all monthly indicators from the publicly available data sources: Statistics Netherlands, Eurostat, the ECB Data portal and the European Commission. In addition to publicly available data, the database also includes information from proprietary sources, i.e: Refinitiv (e.g. PMI manufacturing), Betaalvereniging Nederland for debit card payments, and Het Financieele Dagblad for tone-adjusted news topics based on previous research (see Dijk, van and de Winter, 2023). The small, medium-sized, and large datasets contain 24, 67 and 129 monthly series, respectively. The small, medium-sized, and large datasets contain 4, 9 and 9 quarterly series, respectively.

Following the philosophy of the frequently used FRED-MD macroeconomic database for the USA (McCracken and Ng, 2016), we categorize our data series into seven groups: (1) output & income, (2) labor market, (3) housing, (4) consumption, orders & inventories, (5) money & credit, (6) interest & exchange rates, and (7) stock market. The selection process is guided by two criteria: a direct relationship with macroeconomic developments and a start date no later than March 2000. The model selection Section shows whether the size of the database matters for the forecast accuracy of the model. Following, amongst others, Bańbura and Modugno (2014), we include the headline or market moving indicators in the small dataset. The indicators in the medium-sized and large datasets are in large part guided by the medium-sized and large datasets in Bańbura and Modugno (2014), Alvarez and Perez-Quiros (2016) and Barigozzi and Luciani (2021). The medium-sized dataset adds more disaggregated and sectoral series, and the large dataset includes even more detailed series. We take a practical approach and base our choice of small, medium-sized, and large datasets in line with our in-house expertise and the datasets used in previous research. An alternative to this approach is a data-driven approach that selects the most relevant indicators (see e.g., Bai and Ng, 2008 and Rünstler, 2016). That could be investigated in future work. In our experience, a challenge posed by these data-driven selection methods lies in determining how to handle outcomes that suggest the inclusion of obscure data series while excluding major "market moving" indicators. Our different database sizes in more detail:

- ***Small***, comprises the main market-moving indicators of real activity for the entire economy. These include industrial production, the stock of new orders, retail sales, the unemployment rate, the economic sentiment indicator, the purchasing manager index, and confidence levels reported in newspapers. Additionally, it includes financial series such as stock price indices or raw material prices as well as the headline FD-sentiment index. This dataset encompasses a total of 24 monthly series and 4 quarterly series;
- ***Medium-sized***, in addition to the series covered in the small specification, provides more disaggregated information on industrial production, survey data, and national accounts. It nearly encompasses all the essential real economic indicators for the Netherlands as reported by Statistics Netherlands and Eurostat. Furthermore, it includes four financial newspaper sentiment indices, each representing the sentiment of one of the four main topics that make up the headline index and the European stock market index. This dataset consists of 67 monthly series and 9 quarterly series;
- ***Large***, apart from the indicators in the medium-sized dataset, incorporates series from the large euro area factor model described in works like Bańbura et al. (2011). It offers greater sectoral granularity for industrial production, the services sector, and retail trade. Additionally, it includes sub-indices from sentiment surveys, sixteen more granular newspaper sentiment indices, and PMIs. The large dataset comprises 129 monthly series and 9 quarterly series.

Table A.1 shows the series name, group, frequency, transformation, source, start-year and publication lags (in days) of all series in the small dataset. Table A.1 in the appendix shows the mnemonics for all series in the small, medium-sized and large datasets.

4 Model specification horse race

This Section disentangles the influence of modeling choices on the forecast accuracy of DFROG. With this aim, we design an experiment that consists of a pseudo real-time

Table 1: Overview of variables in small dataset

Nr.	Series	Group	Freq.	Trans. Ln.	Trans. Diff.	Source	Start	Publ.lag (days)
1	Consumption households: domestic	Cons., orders & inv.	M	X	X	CBS	jan-'95	38
2	Consumer confidence: headline	Cons., orders & inv.	M	O	X	CBS	apr-'86	-8
3	PMI manufacturing: new orders	Cons., orders & inv.	M	O	X	Refinitiv	mrt-'00	1
4	House price index	Housing	M	X	X	CBS	jan-'95	22
5	Houses sold	Housing	M	X	X	CBS	jan-'95	22
6	10-year government bond yield	Interest & exch. rates	M	O	X	Eurostat	apr-'86	10
7	Nominal effective exchange rate	Interest & exch. rates	M	X	X	Eurostat	jan-'99	30
8	Unemployment rate: 15-75 years	Labor market	M	O	X	CBS	jan-'83	17
9	Hours worked	Labor market	Q	X	X	CBS	mrt-'95	45
10	Loans: households	Money & credit	M	X	X	ECB	sep-'97	26
11	M3	Money & credit	M	X	X	ECB	jan-'70	26
12	Industrial production: manufacturing	Output & income	M	X	X	CBS	jan-'95	39
13	Import of goods	Output & income	M	X	X	CBS	jan-'95	42
14	Export of goods	Output & income	M	X	X	CBS	jan-'95	42
15	Industrial confidence: headline	Output & income	M	O	X	Eurostat	jan-'80	0
16	Economic sentiment: headline	Output & income	M	O	X	Eurostat	jan-'80	0
17	Services confidence: headline	Output & income	M	O	X	Eurostat	jan-'96	0
18	Retail sales: total	Output & income	M	X	X	Eurostat	jan-'00	39
19	PMI manufacturing: headline	Output & income	M	O	X	Refinitiv	mrt-'00	1
20	PMI composite: world (headline)	Output & income	M	O	X	Refinitiv	jul-'98	5
21	FD: headline	Output & income	M	O	O	FD	jan-'85	0
22	Capacity utilization	Output & income	Q	O	X	Eurostat	mrt-'80	-3
23	Gross domestic product: eurozone	Output & income	Q	X	X	Eurostat	mrt-'95	45
24	Gross domestic product	Output & income	Q	X	X	CBS	mrt-'96	45
25	Harmonised index of consumer prices: total	Prices & wages	M	X	X	CBS	jan-'96	1
26	Producer prices: manufacturing	Prices & wages	M	X	X	CBS	jan-'81	30
27	Commodity Price index: non-energy comm.	Prices & wages	M	X	X	ECB	jan-'96	5
28	AEX	Stock market	M	X	X	Refinitiv	jan-'83	1

Freq.: frequency of series (M=monthly, Q= Quarterly), Trans.: transformation of series, Ln: take logarithm (X= year, O= no), Diff.: take first difference of series (X= 0, O= yes), Source: source of series (CBS: Statistics Netherlands, ECB: European Central Bank, Eurostat: Eurostat, FD: Financieele Dagblad, Refinitiv: Refinitiv), Start: start mont/quarter of series, Publ. Lag: publication lag of series in days; negative publ. lag implies serie is released *n* days before the end of the month.

out-of-sample forecasting horse-race between two types of dynamic factor models with different treatments. To create variation around those treatments and generate forecast errors from different models associated to each feature, we follow Coulombe et al. (2022) and Carriero et al. (2019) amongst others.

4.1 Forecasting design

All experiments are conducted in a pseudo real-time setting. The first forecasts are produced based on model estimations with data used up until April 2013, and we re-estimate the model in each consecutive month to produce new forecasts up until the last model estimation in November 2023. We estimate the parameters of all models recursively, using only the information that was available at the time of the forecast. More specifically, starting from April 2013, we reconstruct pseudo-real-time vintages by replicating the data availability pattern as implied by a stylized release schedule. This is done by recursively removing observations from the full dataset according to a fixed schedule. Consequently, a series of forecasts, nowcasts, and backcasts of GDP growth rate is obtained from the Kalman smoother each month, using parameters estimated based solely on the information set available at that time.

Given the unavailability of real-time historic data vintages, only final revised data downloaded on January 26, 2024 is used in the nowcasting exercise. As a result, the role of historic data revisions is ignored and we can only perform a pseudo real-time out-of-sample exercise. However, factor models are known to be robust to data revisions since revision errors are idiosyncratic by nature and may cancel out; see for example Bernanke and Boivin (2003) for the United States and Schumacher and Breitung (2008) for Germany. For similar approaches, see Giannone et al. (2008), Jansen and de Winter (2018) and Kant et al. (2022), among others. Another limitation is that we assume the publication calendar to stay the same throughout the entire sample period. In reality, the release delays might shift a little bit between months, and for some releases the publication delay might have diminished over time due to more efficient data collection.

Table 2: Timing of forecast exercise for third quarter

no.	forecast type	month	forecast on the 1^{st} of
1	one-quarter-ahead	1	April
2		2	May
3		3	June
4	nowcast	1	July
5		2	August
6		3	September
7	backcast	1	October
8		2	November

We construct a sequence of eight forecasts for GDP growth in each quarter, obtained in consecutive months. Table 2 explains the timing of the forecasting exercise in more detail, taking the forecast for the third quarter of 2013 (2013Q3) as an example. We make the first forecast on April 1, 2013, which is called a one-quarter-ahead forecast in month one. Subsequently, we produce monthly forecasts for the next seven months through November. The last forecast is made just two weeks before the first release of GDP in mid-November. Following the conventional terminology, forecasts refer to one or more quarter-ahead forecasts, nowcasts refer to current quarter forecasts, and backcasts refer to forecasts for the preceding quarter, before official GDP figures become available. In the case of our example 2013Q3, we make one-quarter-ahead from April to

June, nowcasts from July to September, and backcasts in October and November.

4.2 Horse race design

The empirical analysis is structured as follows. First, we stack the forecast errors from all treatments in one (potentially very long) vector and tease out the impact of the different model features on the forecast accuracy by defining separate dummy variables for each model specification.[4]

Second, we estimate a series of stacked OLS regressions to tease out the above treatments. More precisely, we estimate each treatment for the two model types, and within these model types for each model specification and dataset size used for estimating the models. In total we estimate and produce forecasts based on 30,060 models, and analyze 167,076 forecast errors.

In each regression, the null hypothesis is that there is no predictive accuracy gain with respect to the base specification, unless otherwise indicated. The base version of DFROG is estimated for each of the dataset sizes small, medium-sized and large, and has the following model features: estimation with data available on the 1st day of the month, mild outlier correction, model coefficients estimated in 10-year rolling windows, 3 static factors and 4 lags in the autoregressive part of the model.

In the main text, we present the outcome over the ten-year period running from 2013Q3 up until 2023Q3. This period holds three sub-periods of economic upswing and just as many downturns, giving a view on the forecast accuracy of the models over the business cycle. Moreover, the evaluation period is long enough to determine the statistical significance reliably.[5]

The analyzed period is also special as it includes the aftermath of the European debt crisis, the COVID-19 pandemic, and a surge in energy prices due to the Russian invasion of Ukraine. The swings in GDP during the COVID crisis have no precedence in terms

[4] More precisely, we include the natural logarithm of the squared forecast errors as dependent variable, and dummies for the model features. In this log-level specification the percentage impact of the model features on the squared forecast error can be calculated as (exp(coefficient) -1) $\times$ 100), where exp is the exponent and coefficient is the coefficient of the feature. The squared forecast error puts a higher weight on large forecast errors, in line with the loss function of a central bank, i.e., making large forecast errors is much more costly than small forecast errors.

[5] Periods of economic upswing and downturn defined on the basis of the DNB business cycle indicator, see here.

of size. To assess the robustness of our regression results for the COVID crisis, we also estimate regression models over the period 2013Q3-2023Q3, excluding the COVID crisis. The latter is defined as the period 2020Q1-2020Q3. We also run separate regressions for all models during the period leading up to the COVID crisis (2013Q3-2019Q4) to assess the forecasting performance of the different model specifications in more tranquil times. The statistical outcomes for the sub-periods are somewhat less reliable compared to those over the full sample period, because the number of observations is smaller. We will analyze the impact of modeling choices on the forecasting performance along four dimensions, i.e.:

- ***Model specification***: We evaluate if it matters how we summarize the information in the DNB indicator dataset: are more factors better or worse for forecasting performance? We also investigate to what extent the dynamics in the model matter: is it better to have models with short memory or long memory, depending on the number of lags in the autoregressive part of DFROG?;
- ***Data choice***: Does size matter? We investigate if it matters how many data series we use, by estimating the models with the small, medium-sized, and large datasets extracted from the DNB indicator database. Besides the size of the dataset, we also investigate the merits of including survey data and quarterly data. We also determine the added value of data that are only available with a subscription (Refinitiv) or that are confidential (payments data or the FD sentiment indicator);
- ***Estimation & transformation***: Does it matter for forecast accuracy how we treat outliers, and does the estimation period matter? We show the impact on forecast accuracy by comparing the performance when we do not treat the data for outliers, do a mild outlier correction, or a more stringent correction. Besides, we test to what extent the forecast accuracy of the model depends on the length of the estimation period. We evaluate three versions: an expanding window and two rolling windows of respectively 10 and 15 years. We also assess how to transform survey data. There is an ongoing debate on this issue in the literature on whether to include surveys in levels or differences. The rolling window estimations have the advantage over the expanding window estimation that they better allow for

changes in the relationship between the monthly indicators and GDP over time.[6];

- ***Timing***: Does forecast accuracy increase when more monthly data are available for the quarter to be forecasted? Intuitively it should, but we test for this empirically in three different variants: the distance to the forecasted quarter in terms of quarters (backcast, nowcast, one quarter ahead forecast), the month within the quarter (month 1, month 2 and month 3), and the day within each month (first day of the month, middle of the month or end of the month).

4.3 Outcome horse race

This Section presents the impact of the modeling choices described in the previous paragraph. The main text presents the figures based on the whole sample period, but we will discuss the impact of the modeling choice separately for the pre-COVID period and the period as of COVID if the outcomes deviate strongly from the whole sample outcomes. The figures for the latter two periods are collected in the Appendix.

Including one lag is sufficient, but adding more factors is advisable

Recently, Miranda et al. (2022) conducted an in-depth review of the optimal model specification of dynamic factor models in the context of the euro area. They concluded that the out-of-sample forecast accuracy is minimized when forecasts are based on models with one factor and parsimonious dynamics in the factors.

Our regression results only partly corroborate this finding for the Netherlands, as shown in Figure 1. The figure presents the gain achieved by adding more factors or more lags in the VAR-part of the dynamic factor model. The footnote in Figure 1 provides additional details on the estimation period and the regression diagnostics. All effects are rounded to whole percentages and presented relative to the base specification.

[6] An alternative way to deal with changes in the relationship between the monthly indicators and quarterly GDP is to incorporate, non-linearities, stochastic volatility and/or time-varying parameters (see e.g. Eraslan and Schröder, 2023). We abstained from these possibilities for now, because to estimate these concepts with a reasonable reliability we need either a (very) long history of the indicators or impose strong priors on these relationships. Moreover, the time needed to re-estimate these models is quite long given our computing power which complicates using these models for frequent and fast updates. Moreover, for some series we lack (very) long history which is necessary to estimate these models. Our current approach tries to hedge against time-variation and non-linearities by re-estimating the model each month. However, it could be fruitful for future research to explore non-linearities, time-variation and stochastic volatility in dynamic factor models in more detail, using recent advances in speeding up EM estimation techniques (e.g. Opschoor and van Dijk, 2023).

For clarity, we express the outcomes visually using figures, focusing solely on the full period in the main text to maintain conciseness. Our evaluation of model specifications involves two approaches:

- ***Economically meaningful***: If a bar falls within the shaded region, the squared errors deviate by more than 5% from the base DFROG model specification.[7] This ±5% threshold serves as a rough, informal gauge of the economic significance of forecast accuracy gains resulting from this model feature, as previously employed by Jansen and de Winter (2018);
- ***Statistical significance***: Coefficient significance is formally tested at the 5%-level, following methodologies by Coulombe et al. (2022) and Carriero et al. (2019). Colored bars indicate statistically significant coefficients:
  - ***Green bars***: Represent an improvement in forecast accuracy;
  - ***Red bars***: Represent a deterioration in forecast accuracy.

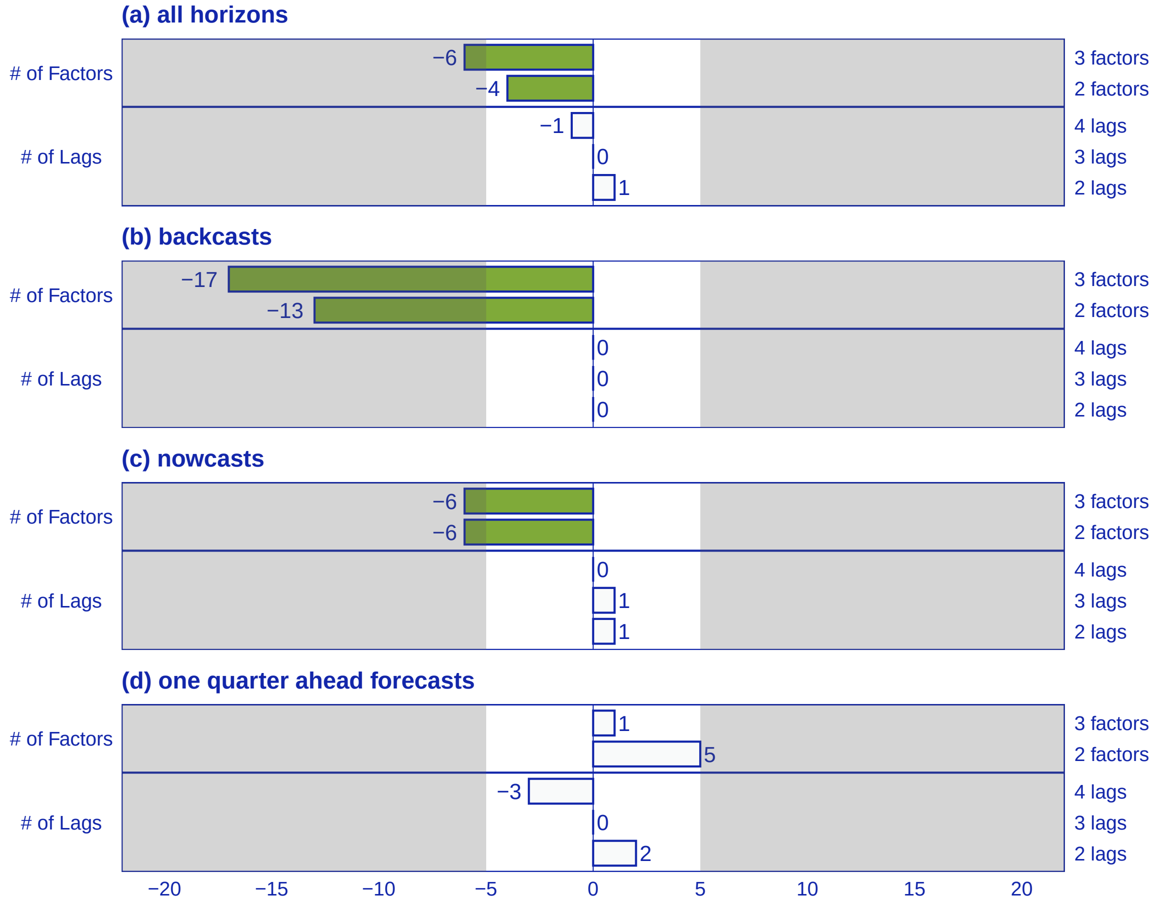
One of the messages conveyed by Figure 1 is that one lag in the VAR-part of the model appears to be sufficient. However, to better capture the variance in the monthly dataset, the model requires additional factors. The number of factors included in the regression significantly impacts the outcome. It could potentially increase the forecast accuracy of the competitor to DFROG up to 17% for backcasting. The gains in forecast accuracy are largest for backcasts, followed by the nowcasts.

Overall, the best-performing specifications have three factors.[8] This outcome is strongly driven by the period post-COVID. Pre-COVID, the only statistically and economically significant impact on the forecast accuracy stems from the third factor, when backcasting (see Figure A.1 in the Appendix). This indicates that more model features are required to capture the more complicated dynamics since the COVID-crisis.

[7] The base DFROG model specification includes 1 factor and 2 lags, and all coefficients are estimated using a 10 year rolling window. The model is estimated using a small dataset including both public & non-public data, excluding series with a quarterly frequency. All series are corrected for extreme outliers and all survey data are included in first difference. The ragged edges in the data are constructed as if the data where downloaded on the first day of the month.

[8] This outcome might raise the question if there is any increment from going from three to four factors. This increment is (very) small and is in line with the statistical tests conducted to determine the number of factors, following Bai and Ng (2002), and scree plot tests Catell (1966). Moreover, including too many factors can yield non-negligible estimation errors, or overfitting of the model (see e.g., Miranda et al. (2022) and Barigozzi and Cho (2020).

Figure 1: Impact of model-specification on forecast accuracy, full sample
Impact of features on out-of sample mean squared forecast, compared to base model, in percent



Notes: Squared forecast error regressed on all model features via stacked regression. A colored bar indicates the coefficient is statistically significant at the 5% level. If a bar is inside the shaded the absolute errors differs by more than ±5%. All effects against a model with the following specifications: small dataset, 1 factor, 2 lags, estimation with a rolling window of 10 years, using both public & non−public data, no quarterly variables included, only extreme outliers deleted, surveys in first difference, vintages downloaded on the first of the month.

Estimation period: 2013Q3−2023Q3. Regression diagnostics: Panel a: number of observations = 123,192, adjusted R−squared = 0.349, p−value = 0.000. Panel b: number of observations = 28,584, adjusted R−squared =0.303, p−value =0.000. Panel c: number of observations = 47,304, adjusted R−squared = 0.335, p−value = 0.000. Panel d: number of observations = 47,304, adjusted R−squared = 0.423, p−value = 0.000.

Source: Own calculations.

For simplicity, we could opt for a specification with only 1 lag. However, to safeguard against occasional forecast inaccuracies, we decide the opt for averaging across specifications with 1 to 3 lags. For the same reason, we decide to average over 1 to 3 factors.

Size does not matter, but including surveys increases accuracy

The main conclusion from analyzing the impact of dataset size in the dynamic factor model reads as follows: while the sectoral information in the medium-sized dataset may be useful for interpretation, it is not necessarily essential for accurate GDP forecasts. This is depicted in Figure 2. Interestingly, the small-sized dataset specification, which includes series measuring only total economy concepts, performs comparably.[9] This may be attributed to the challenges of extracting a relevant signal when dealing with indicators of varying quality, as highlighted by researchers such as Boivin and Ng (2006) and Bańbura and Modugno (2014). Interestingly, our findings align with the existing academic literature (e.g., Caggiano et al., 2011 and Havrlant et al., 2016) which suggests that medium-sized datasets (typically containing 10-30 variables) perform just as well as models with larger datasets (containing over 100 variables). The results remain consistent across different time periods, including the tranquil pre-COVID period and the period without the COVID crisis.

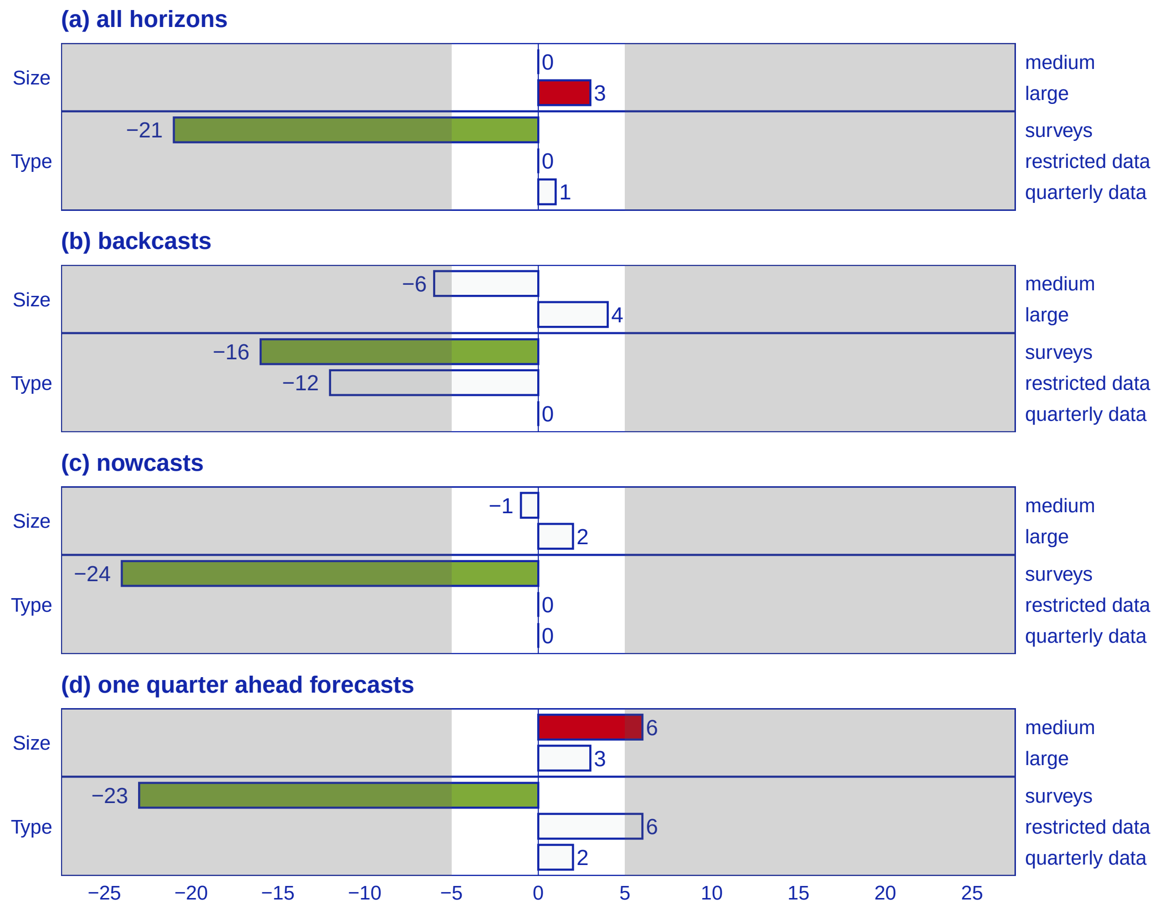
When considering the type of data to include, Figure 2 strongly indicates that incorporating surveys strongly and significantly improves forecast accuracy. This results is apparent in both the pre-COVID period as well as the period without COVID. Adding quarterly data to the forecasting model does not increase the forecast accuracy in a meaningful way. Furthermore, adding quarterly data –such as production capacity and sub-components of GDP– does not lead to improved forecast accuracy in our sample. Measured over the total sample period, adding series from restricted data series (Refinitiv, FD) does not impact the forecast accuracy, and this results is quite persistent.[10] There is one exception: Adding the restricted series lowers the average RMSFE of the backcasts during the no-COVID period by 14%. This implies that including restricted series increases the forecasts accuracy of backcast after the COVID-crisis.

Based on these findings, we conclude that expanding the dataset size and incorporating quarterly data do not necessarily enhance forecast accuracy. However, policymakers might still prefer a medium-sized or large dataset over a small one, for the purpose of interpreting the information conveyed by their releases. Notably, the inclusion of survey indicators significantly improves forecast accuracy and should be considered. Although adding restricted data to the dataset does not increase the forecast accuracy, it also does not hurt. Some indicators, such as the PMI and financial market data are strongly favored by policy makers and we therefore include them in our model. We decide to opt

[9] The gains from using a large dataset are comparatively small. Measured over the total sample the deterioration is statistically significant, but economically, the deterioration is not sizable (3%).

[10] The FD-indicators significantly improves the forecast accuracy in a nowcasting model over a longer sample. This is also one of the reasons why we opt to include FD-indicators, see Dijk, van and de Winter (2023).

Figure 2: Impact of choice of data on forecast accuracy
Impact of features on out-of-sample mean squared forecast, compared to base model, in percent



Notes: Squared forecast error regressed on all model features via stacked regression. A colored bar indicates the coefficient is statistically significant at the 5% level. If a bar is inside the shaded the absolute errors differs by more than ±5%. All effects against a model with the following specifications: small dataset, 1 factor, 2 lags, estimation with a rolling window of 10 years, using both public & non−public data, no quarterly variables included, only extreme outliers deleted, surveys in first difference, vintages downloaded on the first of the month.

Estimation period: 2013Q3−2023Q3. Regression diagnostics: Panel a: number of observations = 123,192, adjusted R−squared = 0.349, p−value = 0.000. Panel b: number of observations = 28,584, adjusted R−squared =0.303, p−value =0.000. Panel c: number of observations = 47,304, adjusted R−squared = 0.335, p−value = 0.000. Panel d: number of observations = 47,304, adjusted R−squared = 0.423, p−value = 0.000.

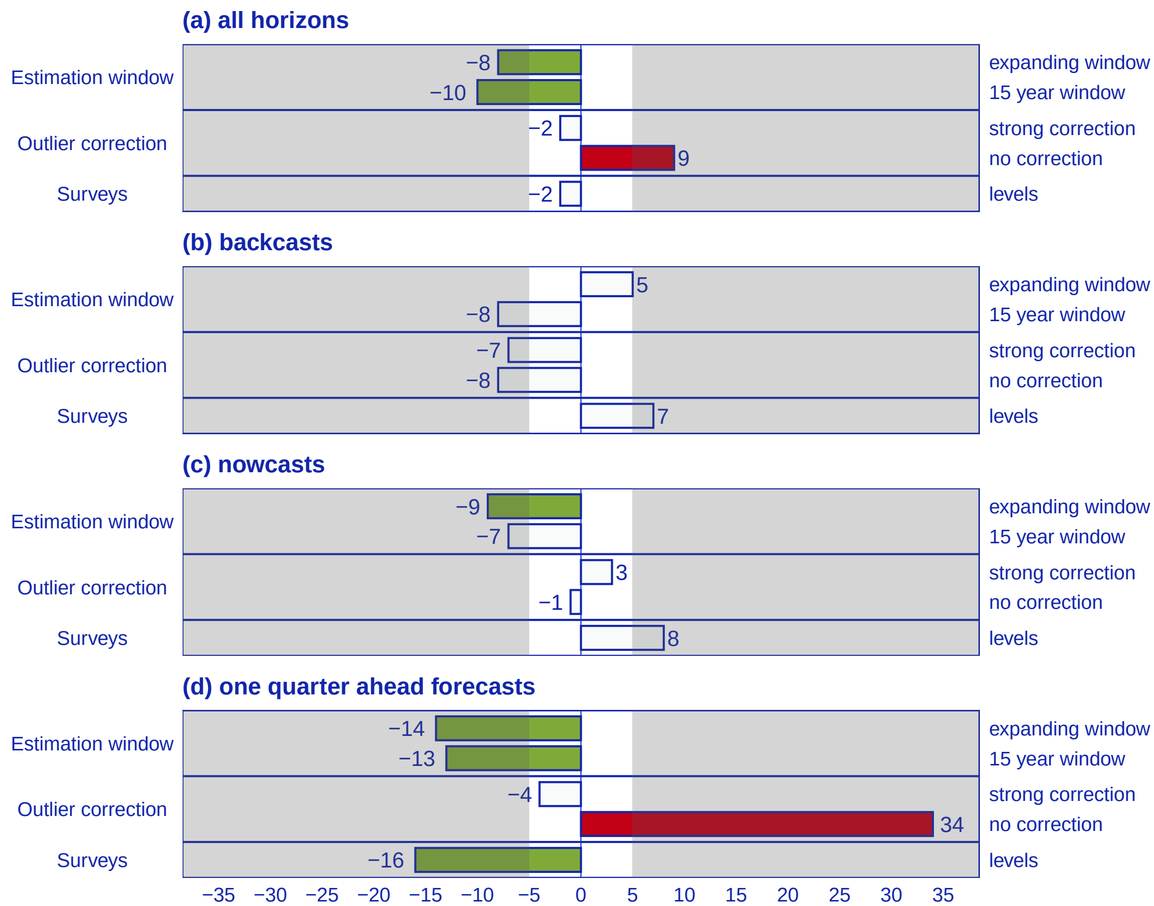
Source: Own calculations.

for a medium-sized dataset including survey indicators and series from restricted data sources. We do not include quarterly data.

15-year window and mild outlier correction increase forecast accuracy

Figure 3 illustrates the impact of estimation and transformation options in the model, revealing some interesting insights. Over the entire period, the 15-year moving estimation window is the preferred option for estimation. Using a 15-year estimation window results in a 10% increase in forecast accuracy, averaged over all forecasting horizons. This result is strongly driven by the COVID period and afterwards, as we do not find any significant effect of the size of the estimation window on the forecast accuracy pre-COVID.

Figure 3: Impact of estimation & transformation on forecast accuracy
Impact of features on out-of sample mean squared forecast, compared to base model, in percent



Notes: Squared forecast error regressed on all model features via stacked regression. A colored bar indicates the coefficient is statistically significant at the 5% level. If a bar is inside the shaded the absolute errors differs by more than ±5%. All effects against a model with the following specifications: small dataset, 1 factor, 2 lags, estimation with a rolling window of 10 years, using both public & non−public data, no quarterly variables included, only extreme outliers deleted, surveys in first difference, vintages downloaded on the first of the month.

Estimation period: 2013Q3−2023Q3. Regression diagnostics: Panel a: number of observations = 123,192, adjusted R−squared = 0.349, p−value = 0.000. Panel b: number of observations = 28,584, adjusted R−squared =0.303, p−value =0.000. Panel c: number of observations = 47,304, adjusted R−squared = 0.335, p−value = 0.000. Panel d: number of observations = 47,304, adjusted R−squared = 0.423, p−value = 0.000.

Source: Own calculations.

Second, a mild correction for outliers in the dependent variable appears sufficient. The forecast accuracy is not adversely affected by using a more stringent outlier correction. Using no outlier correction strongly worsens accuracy, especially when forecasting. It can lead to a 37% lower accuracy for forecasts in the period excluding the COVID crisis and 27% lower forecast accuracy in the pre-COVID period.[11]

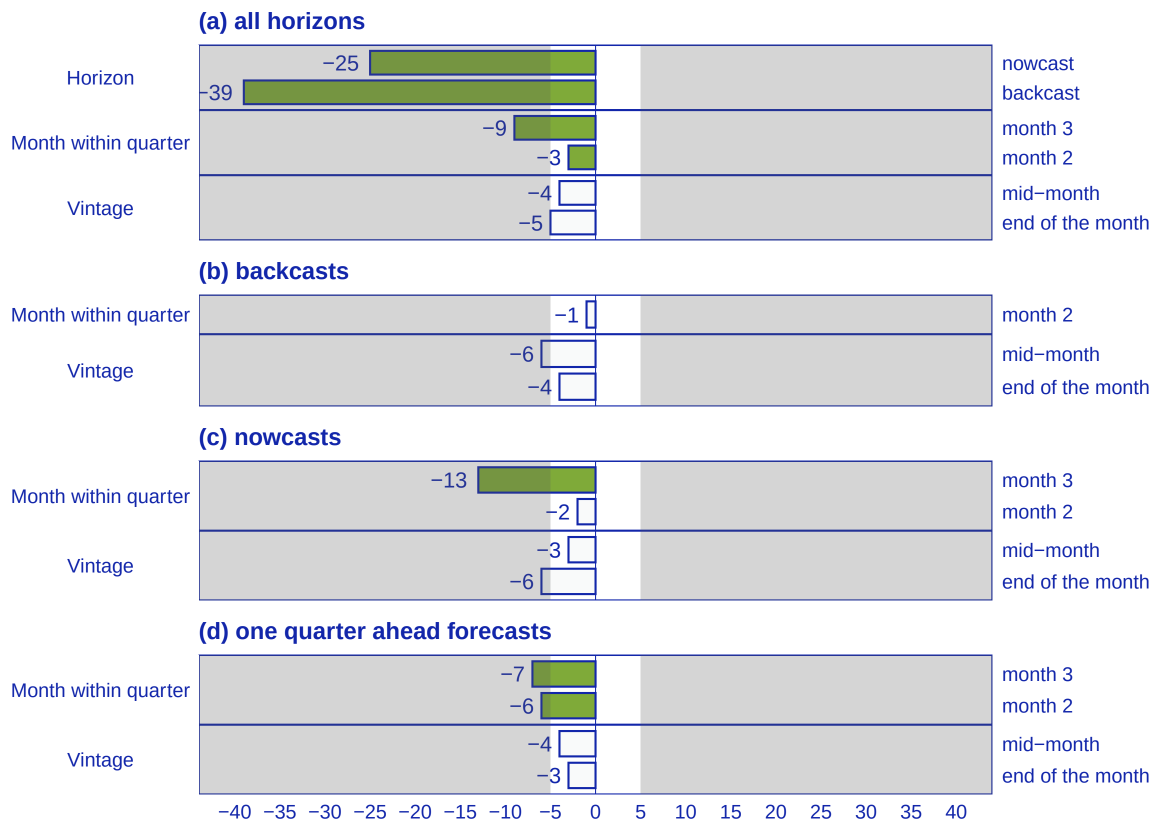
[11] Outlier correction is performed for each indicator, based on the observed statistical distribution of that indicator. For strong outlier correction, all indicator values that fall outside the interval defined by the median of the indicator ±3 times the interquartile range of the indicator distribution are removed. For

Based on these results, we choose a 15-year moving estimation window. Apart from the increase in forecast accuracy, this choice also helps to safeguard against changes in statistical relationships among the variables. Additionally, a mild outlier correction should be sufficient.

Forecast accuracy increases with more recent information

Figure 4 presents an intuitive result. When there is more information available for the quarter being forecasted, the forecast accuracy is (much) higher.

Figure 4: Impact of timing on forecast accuracy
Impact of features on out-of sample mean squared forecast, compared to base model, in percent



Notes: Squared forecast error regressed on all model features via stacked regression. A colored bar indicates the coefficient is statistically significant at the 5% level. If a bar is inside the shaded the absolute errors differs by more than ±5%. All effects against a model with the following specifications: small dataset, 1 factor, 2 lags, estimation with a rolling window of 10 years, using both public & non−public data, no quarterly variables included, only extreme outliers deleted, surveys in first difference, vintages downloaded on the first of the month.

Estimation period: 2013Q3−2023Q3. Regression diagnostics: Panel a: number of observations = 123,192, adjusted R−squared = 0.349, p−value = 0.000. Panel b: number of observations = 28,584, adjusted R−squared =0.303, p−value =0.000. Panel c: number of observations = 47,304, adjusted R−squared = 0.335, p−value = 0.000. Panel d: number of observations = 47,304, adjusted R−squared = 0.423, p−value = 0.000.

Source: Own calculations.

Backcasts have the significant advantage of being able to access all the monthly data for

strong outlier correction, the interval is defined by the median of the indicator ± 2 times the interquartile range of the indicator distribution. Note that the outlier-correction is conducted over the estimation window. We calculate the forecast accuracy using the true (non-outlier corrected) GDP growth.

the quarter being forecasted, which enhances its accuracy. Measured over the full period, the nowcasts and backcasts are 25% and 49% more accurate than the one-quarter-ahead forecasts, respectively. The same logic applies when forecasting later in the month: forecasting in the third month of a quarter improves the forecast accuracy by 9% compared to the first month. Measured over the complete sample, this effect is much more pronounced for the nowcasts than the one quarter ahead forecast. In the pre-COVID period, the impact of the increase in forecast accuracy of forecasting in the second and third month of a quarter is only economically meaningful and statistically significant when forecasting. The forecast accuracy for backcasting in the first or second month is almost the same and does not significantly differ. This indicates that most of the important information for backcasting a quarter is available by the end of the first month.

The differences in forecasting at the beginning, middle, or end of the month are neither economically meaningful nor statistically significant. This intuitive result implies that it is much more important whether you are backcasting, nowcasting, or forecasting a quarter, and which specific month you are in within a quarter, rather than the specific day of the month.

Summing up: DFROG forecast is average of 18 specifications

Based on our empirical results our DFROG model is the average of the forecasts over 18 different specifications of the model. These 18 vary in the number of lags, the number of factors and the transformation of the survey data. We included all combinations of 1 to 3 lags in the VAR, 1 to 3 factors. In total this results in 9 forecasts. We estimate these 9 models for two variants: one where all the survey indicators are in levels, and one where all survey indicators are in first difference, resulting in a total of 18 models. All 18 models are estimated using a 15-year rolling window, mild outlier-correction, and a medium-sized dataset which includes surveys and restricted data.

5 Comparison to benchmark models

This Section compares the forecast accuracy of DFROG with an alternative dynamic factor model specification with three benchmark models: a random walk with drift, an autoregressive model and a popular alternative dynamic factor model specification

(Bańbura et al., 2011). Additionally, we report the relative forecast accuracy of DFROG against the forecasts of professional analysts. For the latter, we use the quarterly forecasts for Dutch GDP growth, supplied by Consensus Forecasts. The new Consensus forecasts are released each last month of a quarter and constitute the average quarterly forecast of the surveyed professional forecasters. The panelists supply GDP forecasts for six consecutive forecasts, starting from the first unpublished quarter. The well-known annual Consensus forecasts have been analyzed in several papers. However, to the best of our knowledge, the quarterly GDP forecasts have not been used before, except in a recent case study for the Netherlands (Jansen and de Winter, 2018).

5.1 Comparing forecast accuracy

Table 3 presents the main outcomes of our analysis. Panel (a) shows the out-of-sample relative Root Mean Squared Forecast Error (rRMSFE) of the benchmark models and the Consensus Forecasts over the total evaluation period, i.e. the period 2013Q3–2023Q3. All RMSFEs are in relative terms (rRMSFE) against DFROG, i.e.: $\text{rRMSFE}_{\text{ALT}} = \frac{\text{RMSFE}_{\text{ALT}}}{\text{RMSFE}_{\text{DFROG}}}$. Here, ALT denotes the forecast of the alternative, i.e. one of the benchmark models or the Consensus forecasts. A value higher/lower than 1 indicates that the RMSFE of the alternative forecast is higher/lower than the RMSFE of DFROG. Bold cells indicate the cases where the alternative model RMSFE is at least 5% higher than the RMSFE of DFROG. Starred entries (*, **, ***) indicate that the one-sided Diebold and Mariano (1995) test indicate the difference is statistically significant at the 10%, 5%, and 1% levels, respectively.[12] Panel (b) shows the outcomes for the same key figures for the Pre-COVID period, whilst Panel (c) shows the outcomes for the entire period excluding COVID. The outcomes in Table 3 point to several interesting results.

First, DFROG backcasts and nowcasts are economically and statistically more accurate than the random walk and autoregressive models, further the "naive benchmark models". This conclusion holds both for the entire period, the pre-COVID period as well as the entire period without the very volatile COVID quarters. When forecasting one-quarter ahead DFROG does not always beat the naive benchmark models. This is a well-known phenomenon in the literature: DFMs perform well when information on

[12] The RMSFEs are based on the most recent realization of GDP growth, the outcomes are robust to using the "first estimate" of GDP growth of CBS at the time of release.

Table 3: Relative RMSFE of benchmark models and Consensus forecasts versus DFROG

	Backcast		Nowcast			1Q Ahead forecast		
	M2	M1	M3	M2	M1	M3	M2	M1
(a) Total evaluation period: 2013Q3 - 2023Q3 (*N*= 41)								
Random walk	**1.19***	**1.20***	**1.19****	**1.22****	**1.14*****	**1.08****	0.95	1.02
Autoregressive model	**1.19****	**1.19****	**1.18****	**1.12***	**1.05****	0.98	0.93	0.99**
Dynamic factor model B&R	0.98	1.02*	**1.05***	1.03	**1.06**	1.02	0.95	0.98
Consensus forecast	0.67	0.68	0.67	0.90	0.84	0.79	0.90	0.97
(b) Pre-COVID: 2013Q3 - 2019Q4 (*N*= 26)								
Random walk	**1.30****	**1.28***	**1.24***	**1.22***	**1.26****	**1.11**	**1.34****	**1.36**
Autoregressive model	**1.12***	**1.09***	**1.07**	1.03	**1.07***	0.94	1.01	1.04
Dynamic factor model B&R	**1.07**	**1.07**	**1.07**	**1.11***	**1.16****	1.03	1.04	**1.06**
Consensus forecast	**1.35****	**1.33****	**1.25****	1.03	**1.07**	0.95	0.87	0.87
(c) No COVID: 2013Q3 - 2023Q3, excluding 2020Q1 -2020Q3 (*N* = 38)								
Random walk	**1.38***	**1.20***	**1.19***	**1.22*****	**1.14****	**1.07****	0.94	1.01
Autoregressive model	**1.25***	**1.25***	**1.20***	**1.21***	**1.15***	1.04*	1.01	1.00
Dynamic factor model B&R	1.03	1.03	1.00	0.99	1.00	0.91	0.91	0.93
Consensus forecast	**1.46****	**1.46****	**1.41****	0.95	0.91	0.82	0.84	0.82

Note: Bold cells indicate the RMSFE is at least 5% worse than DFROG. Starred entries (*, **, ***) indicate that the one-sided Diebold-Mariano test (alternative is worse than the baseline) is significant at the 10%, 5%, and 1% levels, respectively.
Source: Own calculations.

the quarter to forecast is partly known, but does not necessarily have the competitive edge when the forecast horizon is longer and no information on the quarters is available, see e.g. Giannone et al. (2008) and Jansen and de Winter (2018). The added value of DFROG increases when more data for the forecasted quarter arrive and can increase up to 38% for the period excluding the COVID period.

Second, comparing the outcomes in panel (a) of Table 3 with the outcomes in panel (b) and (c) of Table 3 it is evident that the competitive edge of DFROG over the naive benchmark models diminished during the COVID crisis. This is not surprising, as the onset and severity of the crisis were largely unapparent in the monthly data releases, partly because all series were adjusted for extreme outliers."

Third, DFROG outperforms the popular dynamic factor model of Bańbura and Rünstler (2011) in the Pre-COVID period, but only by a small margin. When backcasting and nowcasting the advantage is no more than 7% on average. The nowcasts of the Bańbura and Rünstler (2011) for the first and second month where 11% and 16% worse, respectively. However, measured over the whole sample as well as excluding the COVID crisis, DFROG does not systemically beat the Bańbura and Rünstler (2011) model. In these periods the RMSFE of DFROG is never both economically and statistically different from the model of Bańbura et al. (2011).

Fourth, DFROG outperforms the Consensus forecasts when nowcasting and backcasting in normal times, but the latter are more accurate in times of large distress. This result has been documented before Jansen and de Winter (2018) and Lundquist and Stekler (2012). It reflects the inability of mechanical statistical model to incorporate expert knowledge. Professional forecasters are very responsive to the latest information about the state of the economy that is not captures in the monthly indicators and adjust their predictions quickly. Strikingly, professional analysts fail to produce accurate back- and nowcasts during tranquil times, and the forecasting performance is the mirror image of the good forecasting performance during the COVID crisis. Over the whole evaluation period excluding the three COVID quarters the professional analysts are beaten by a 46% margin by DFROG when backcasting.

Fifth, the RMSFE of the one quarter ahead forecasts of professional forecasters is smaller than the RMSFE of DFROG, regardless of the period considered. However, the differences are not statistically significant indicating that professional forecasters are not structurally beating DFROG on this forecasting horizon.

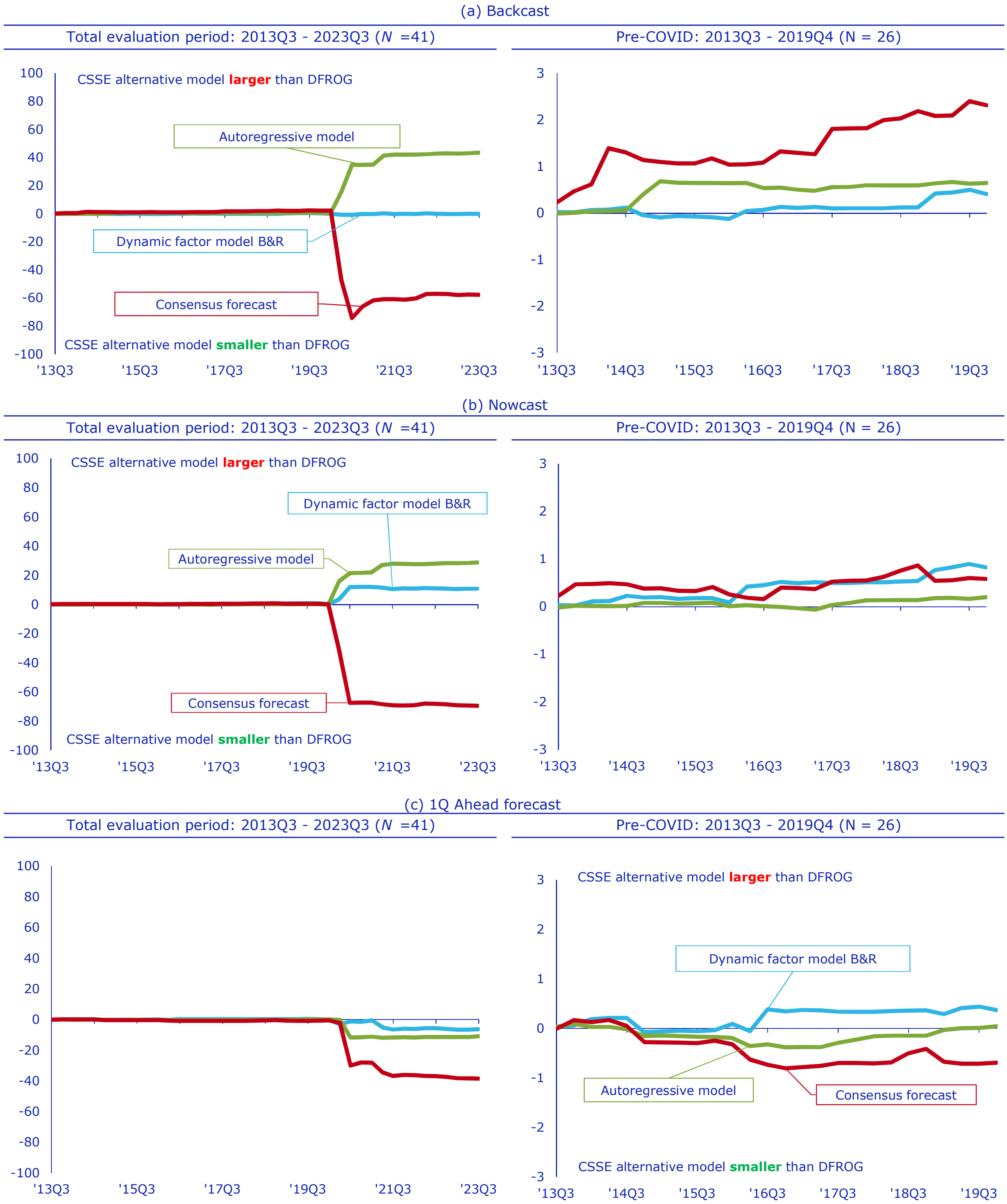
5.2 Comparing the evolution of forecast accuracy in time

Figure 5 casts the outcomes in Table 3 into the time dimension, showing the *cumulative* mean squared forecast error difference (CSSED) moving forward in time, calculated as the cumulative sum of squared errors of the alternative model *minus* the cumulative squared error of DFROG. A CSSED *below* zero indicates that the alternative model's forecasts have a *lower* CSSE up until that that point in time, and are therefore more accurate than DFROG. If the CSSED is *above* zero this indicates the reverse and the alternative model has a a lower forecast accuracy at that point in time. Furthermore, an increase in the CSSED indicates that the model performance of the alternative model is decreasing vis-a-vis DFROG. A decline indicates the opposite.

The left-hand graphs in panel a, b and c of Figure 5 describe the evolution of the CSSED's over the entire evaluation period for the backcasts, nowcasts and one quarter

Figure 5: Cumulative sum of squared error difference (CSSED)

CSSE alternative model - CSSE DFROG



Note: Cumulative Sum of Squared Errors Difference (CSSED) calculated for the average errors in the backcast (M1 and M2), nowcast (M1, M2 and M3) and 1 quarter ahead forecast (M1, M2 and M3). Calculated as CSSE alternative model - CSSE DFROG.

Source: Own calculations.

ahead forecasts, respectively. The left-hand graphs of Figure 5 obscure the development prior to the COVID crisis due to significant shifts in the relative forecast accuracy of the models in COVID. To gain a clearer understanding of the pre-COVID CSSED evolution, the right-hand graphs display the CSSED up to the COVID period for the backcast, nowcast, and one-quarter-ahead forecast, respectively.

Panel a of Figure 5 shows that when backcasting, the forecasting advantage of the Consensus forecast solely stems from the COVID period. There is a huge and sudden decline in the CSSE of the Consensus forecast vis-a-vis DFROG. After COVID, there is an upward trend in the red line, indicating that DFROG regained some of its forecasting advantage. The COVID period caused a significant decline in the forecast accuracy of the autoregressive model, as evidenced by the large increase in the green line. The alternative dynamic factor model (Bańbura and Rünstler, 2011) showed no noticeable change, suggesting that its forecast accuracy during backcasting remained roughly the same as DFROG.

The evolution of the CSSED for the nowcasts (Panel b of Figure 5) is similar to that for the backcasts, though by a lesser extent. A notable difference is the significant deterioration in the forecast accuracy of the dynamic factor model of Bańbura and Rünstler (2011). When forecasting one quarter ahead (panel c of Figure 5) the models are slightly more accurate than DFROG measured over the entire period. In line with the outcomes in Table 3 the differences are quite small, and mainly caused by deviating forecast performance during the COVID crisis.

Interestingly, the forecast accuracy of the Consensus forecasts was declining compared to DFROG in the period leading up to the crisis, for both the backcasts (panel a) and the nowcasts (panel b). Note that the scale of this decline is significantly smaller than that shown in the left-hand graphs. Additionally, there was a noticeable deterioration in the forecast accuracy of the dynamic factor model by Bańbura and Rünstler (2011) compared to DFROG, starting in the second half of 2015 and continuing right up to the COVID crisis. This trend is observed in backcasts, nowcasts, as well as one-quarter-ahead forecasts.

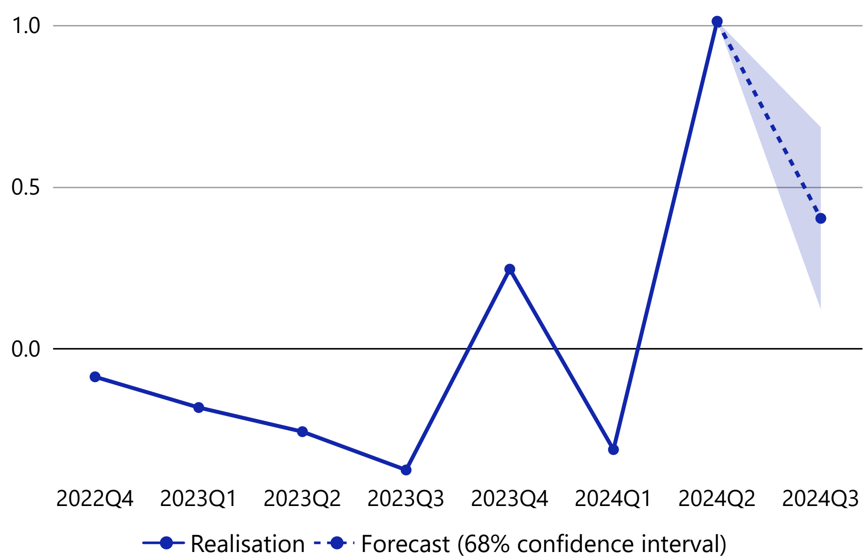
6 Using DFROG in practice

DFROG can be quickly updated through automated data updates and model estimation. This Section shows how we interpret and use the main model outputs of DFROG. We use the backcast for 2024Q3, made on October 1, 2024, as an example.

6.1 DFROG backcast for 2024Q3 GDP growth

Figure 6 shows the backcast for 2024Q3 as of October 1, 2024.[13] On October 1st, the DFROG backcast for 2024Q3 GDp growth was 0.4% quarter-on-quarter (q-o-q). It's important to acknowledge the significant uncertainty associated with this backcasts: the 68%-confidence interval for 2024Q3 ranges from 0.1% q-o-q to 0.7% q-o-q. A GDP growth figure lower than 0.1% q-o-q or higher than 0.7% q-o-q is possible, it would however be somewhat surprising given the historical uncertainty bands around the DFROG forecast.

Figure 6: DFROG backcast for 2024Q3 on October 1, 2024
Quarter-on-quarter growth of GDP



Source: Own calculations.

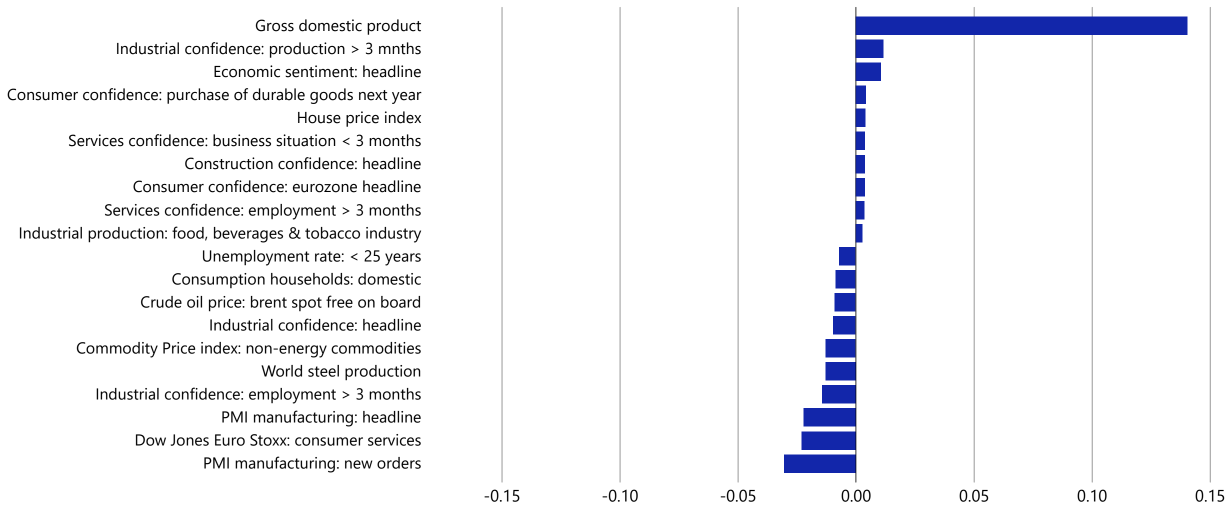
Figure 7 shows the contributions of the DFROG series to the backcast for 2024Q3. The method to compute these contributions is described in Section 2.4. Panel (a) of Figure 7a displays the top and bottom ten contributions to the 2024Q3 forecast. Panel (b) of of

[13] The confidence interval is constructed using the smoothed uncertainty of backcasts in the past, following from the Kalman smoother.

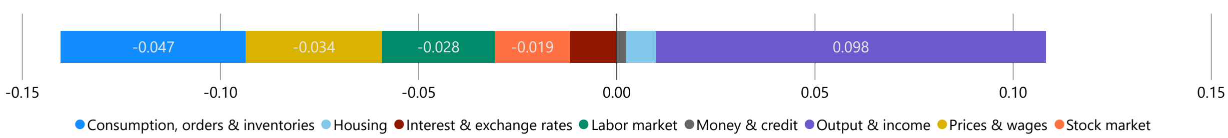
Figure 7 shows the contributions of the variables grouped into the eight categories described in Section 3, i.e.: (1) consumption, orders & inventories, (2) housing, (3) interest & exchange rates, (4) labor market, (5) money & credit, (6) output & income, (7) prices & wages and (8) stock market. A positive or negative contribution indicates a value that is higher or lower than the average over the estimation period, respectively.

Figure 7: Contributions to DFROG backcast for 2024Q3 on October 1, 2024
Quarter-on-quarter growth of GDP, in percentage points

(a) Top and bottom 10 variable contributions



(b) Contributions per group



The sum of the contributions + average GDP growth (0.434) over the estimation sample equals the DFROG forecast for 2024Q3.
Source: Own calculations.

Panel (b) of Figure 7 highlights that, as of October 1, 2024 the"output & income" group provides the highest positive contributions to our backcast. The "housing" and "money & credit" groups also contribute positively, albeit by a much lesser extent. The large positive contribution of the "output & income" group can be mainly traced back to GDP in Panel (a) of Figure 7. As can be seen from that panel GDP growth is the largest contributor to the backast: The relatively high GDP growth in 2024Q2 (1.0% quarter-on-quarter), has has had a substantial positive impact on the bakcast for 2024Q3. Besides the increase in economic sentiment, as well as the production expectation in industry have had a positive impact.

The negative contributions are more evenly spread across groups. Although he largest negative contributions come from the "consumption, orders & inventories" group, the negative contributions from the "prices & wages", "labor market" and "stock market" are also clearly visible. This can also be seen in Panel (a) of Figure 7a. Both the less optimistic new order in industry ("consumption, orders & inventories") and employment expectations in industry ('labor market") contribute negatively to the DFROG backcast. Moreover, the consumer services index of the Dow Jones Euro Stoxx ("stock market") and crude oil and non-energy commodity prices (both "prices & wages") contributed negatively to the DFROG backcast.

Table 4: Change in DFROG backcast in period 15 September – 1 October
Impact of new releases, top and bottom 10, in percentage points.

Last month	Series description	Group	Surprise	Impact forecast
September 2024	Economic sentiment: headline	Output & income	▲	0.012
September 2024	Services confidence: employment > 3 months	Labor market	▲	0.006
September 2024	Dow Jones Euro Stoxx: 50	Stock market	▲	0.004
September 2024	Industrial confidence: production > 3 mnths	Output & income	▲	0.004
September 2024	Eurozone high yield bond spread	Stock market	▼	0.003
September 2024	Construction confidence: headline	Housing	▲	0.003
September 2024	Consumer confidence: unemployment > 12 months.	Labor market	▼	0.003
August 2024	Retail sales: total	Output & income	▲	0.002
September 2024	Consumer confidence: purchase of durable goods next year	Consumption, orders & inventories	▲	0.002
September 2024	Amsterdam exchange index	Stock market	▲	0.002
August 2024	Producer prices: manufacturing	Prices & wages	▼	0.000
September 2024	Retail confidence: headline	Output & income	▼	0.000
September 2024	Services conf.: ev. of de. < 3 m.: empl. activities	Output & income	▼	-0.001
September 2024	Services conf.: ev. of de. < 3 m.: conslt. & head off.	Output & income	▼	-0.001
September 2024	Harmonised index of consumer prices (HICP): total	Prices & wages	▼	-0.002
September 2024	Crude oil price: brent spot free on board	Prices & wages	▼	-0.004
August 2024	Passenger car registration	Consumption, orders & inventories	▼	-0.004
August 2024	Unemployment rate: 15-75 years	Labor market	▲	-0.004
August 2024	Unemployment rate: < 25 years	Labor market	▲	-0.006
August 2024	World steel production	Output & income	▼	-0.006

Source: Own calculations.

Table 4 shows the impact on the DFROG backcast of new data releases in the two week between September 15 and October 1. The method to compute these contributions is described in Section 2.5. This approach helps policy makers assess unexpected changes in the estimates of current economic activity over time and evaluate the significance of each data release. Moreover, it formalizes the essential aspects of how policy makers have traditionally made forecasts. This process involves monitoring numerous data releases, forming expectations, and then revising the assessment of the current state of the economy whenever actual data deviates from those expectations. For clarity, we only show the 10 new releases with the largest positive and negative impact on the backcast. The "impact forecast" column at the end of the table indicates the impact of the data release on our backcast (in percentage points). The "surprise"'columns indicates the difference between the actual value and the expected value for that indicator following from the DFROG estimation on September 15. Only the sign of the surprise is shown. An upward pointing green triangle represents a positive data surprise, and a downward pointing triangle indicates a negative surprise. The other columns show the release of the last data-point of each series ("last month"), a description of the series ("series description") and the series group ("group").

In the two weeks between September 15 and October 1 the data release with the largest positive impact on the DFROG backcast was the economic sentiment indicator. The most negative impact on the backcast came from the higher than expected incoming unemployment figures, lower than expected world steel production and lower than expected passenger car registration.

References

Alvarez, R. and G. Perez-Quiros (2016). Aggregate versus disaggregate information in dynamic factor models. *International Journal of Forecasting 32*, 680–694. link.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221. link.

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

Bańbura, M., D. Giannone, and L. Reichlin (2011). Nowcasting. In M. P. Clements and D. F. Hendry (Eds.), *Oxford Handbook on Economic Forecasting*, pp. 63–90. Oxford University Press. link.

Bańbura, M. and M. Modugno (2014). Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data. *Journal of Applied Econometrics 29*(1), 133–160. link.

Bańbura, M. and G. Rünstler (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. link.

Barigozzi, M. and H. Cho (2020). Consistent estimation of high-dimensional factor models when the factor number is over-estimated. *Electronic Journal of Statistics 14*(2), 2892–2921. link.

Barigozzi, M. and M. Luciani (2021). Quasi maximum likelihood estimation and inference of large approximate dynamic factor models via the EM algorithm. Papers 1910.03821, arXiv. link.

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

Boivin, J. and S. Ng (2006). Are more data always better for factor analysis? *Journal of Econometrics 1*, 169–194. link.

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

Caggiano, C., G. Kapetanios, and V. Labhard (2011). Are more data always better for factor analysis? results for the euro area, the six largest euro area countries and the uk. *Journal of Forecasting 30*, 736–753. link.

Carriero, A., A. B. Galvão, and G. Kapetanios (2019). A comprehensive evaluation of macroeconomic forecasting methods. *International Journal of Forecasting 35*, 1226–1239. link.

Catell, R. B. (1966). The scree test for the number of factors. *Multivariate behavioral research 1*(2), 245–276. link.

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

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

Dijk, van, D. W. and J. M. de Winter (2023). Nowcasting GDP using tone-adjusted time varying news topics: Evidence from the financial press. Working Paper 766, De Nederlandsche Bank. link.

Doz, C., D. Giannone, and L. Reichlin (2012). A quasi-maximum likelihood approach for large, approximate dynamic factor models. *The Review of Economics and Statistics 94*(4), 1014–1024. link.

Eraslan, S. and M. Schröder (2023). Nowcasting GDP with a pool of factor models and a fast estimation algorithm. *International Journal of Forecasting 39*(3), 1460–1476. link.

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

Havrlant, D., P. Tóth, and J. Wörz (2016). On the optimal number of indicators – nowcasting GDP growth in CESEE. Focus on European Economic Integration 4, Oesterreichische Nationalbank. link.

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

Jansen, W. J. and J. M. de Winter (2018). Combining model-based near-term GDP forecasts and judgmental forecasts: A real-time exercise for the G7 countries. *Oxford Bulletin of Economics and Statistics 80*(6), 1213–1242. link.

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

Kant, D., A. Pick, and J. M. de Winter (2022). Nowcasting GDP using machine learning methods. Working Paper 754, De Nederlandsche Bank. link.

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

Lundquist, K. and H. Stekler (2012). Interpreting the performance of business economists during the great recession. *Business Economics 47*, 148–154. link.

Mariano, R. and Y. Murasawa (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of applied Econometrics 18*(4), 427–443. link.

McCracken, M. and S. Ng (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics 34*(3), 574–589. link.

Miranda, K., P. Poncela, and E. Ruiz (2022). Dynamic factor models: Does the specification matter? *SERIEs: Journal of the Spanish Economic Association 13*(1), 397–428. link.

Opschoor, D. and D. van Dijk (2023). Slow expectation-maximization convergence in low-noise dynamic factor models. Discussion Papers 23-018/III, Tinbergen Institute. link.

Rünstler, G. (2016). On the design of data sets for forecasting with dynamic factor models. In E. Hillebrand and S. J. Koopman (Eds.), *Dynamic Factor Models (Advances in Econetrics, Vol. 35)*, pp. 629–666. Emerald Group Publishing Limited. link.

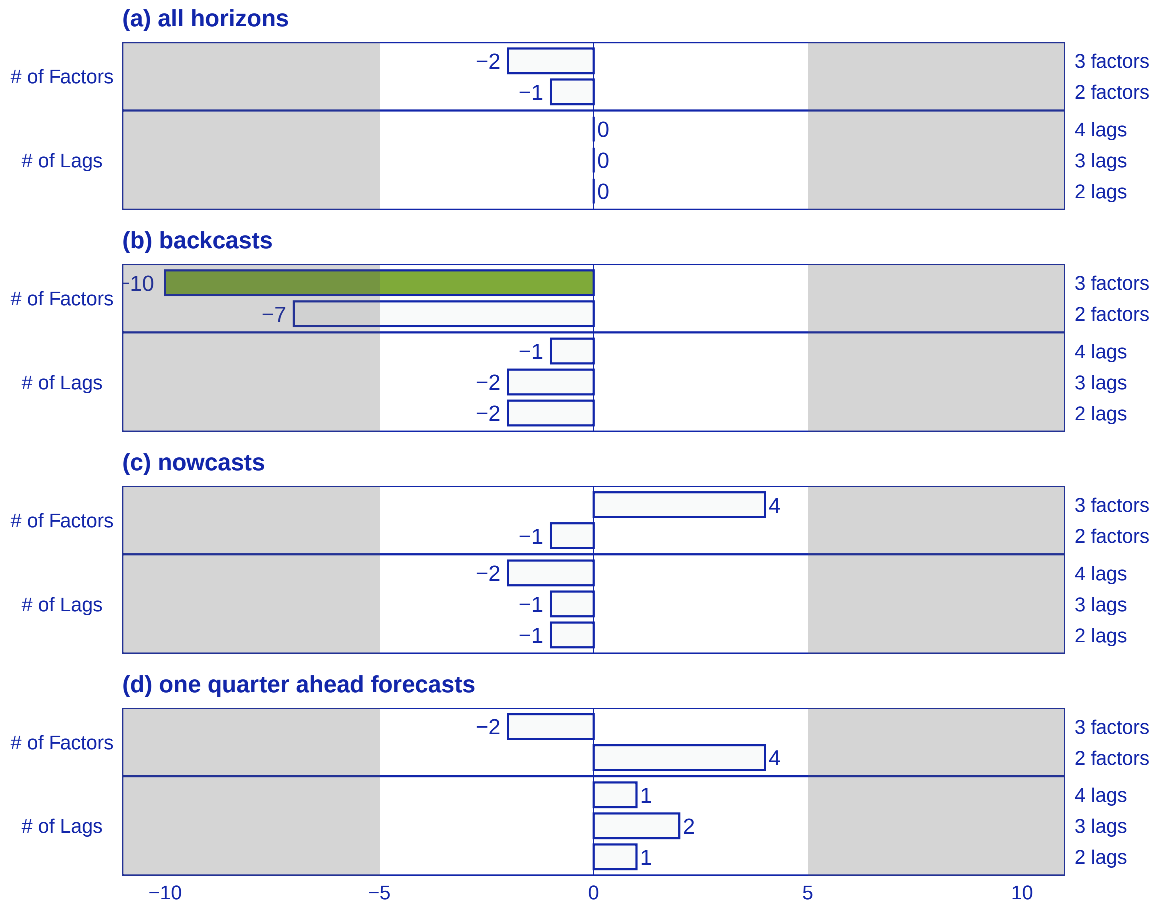
Schumacher, C. and J. Breitung (2008). Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. *International Journal of Forecasting 24*(3), 386–98. link.

Shapiro, A. H., M. Sudhof, and D. J. Wilson (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association 97*, 1167–1179. link.

A Additional results

A.1 Results pre-COVID period: 2013Q3 – 2019Q4

Figure A.1: Impact of model-specification on forecast accuracy: pre-COVID period
Impact of features on out-of sample mean squared forecast, in percent

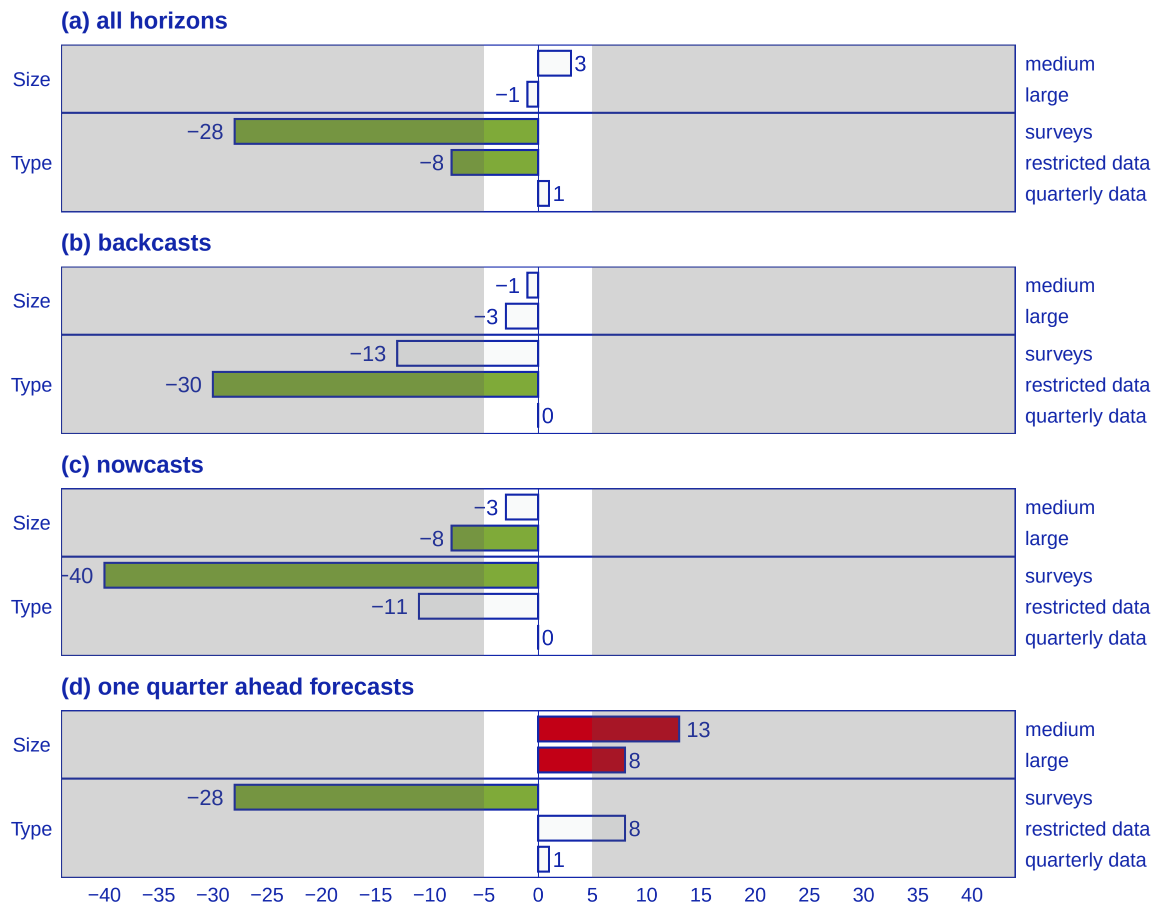


Notes: Squared forecast error regressed on all model features via stacked regression. A colored bar indicates the coefficient is statistically significant at the 5% level. If a bar is inside the shaded the absolute errors differs by more than ±5%. All effects against a model with the following specifications: small dataset, 1 factor, 2 lags, estimation with a rolling window of 10 years, using both public & non−public data, no quarterly variables included, only extreme outliers deleted, surveys in first difference, vintages downloaded on the first of the month.

Estimation period: 2013Q3−2019Q4. Regression diagnostics: Panel a: number of observations = 76,752, adjusted R−squared = 0.081, p−value = 0.000. Panel b: number of observations = 17,784, adjusted R−squared =0.141, p−value =0.000. Panel c: number of observations = 29,484, adjusted R−squared = 0.073, p−value = 0.000. Panel d: number of observations = 29,484, adjusted R−squared = 0.092, p−value = 0.000.

Source: Own calculations.

Figure A.2: Impact of choice of data on forecast accuracy: pre-COVID period
Impact of features on out-of-sample mean squared forecast, in percent



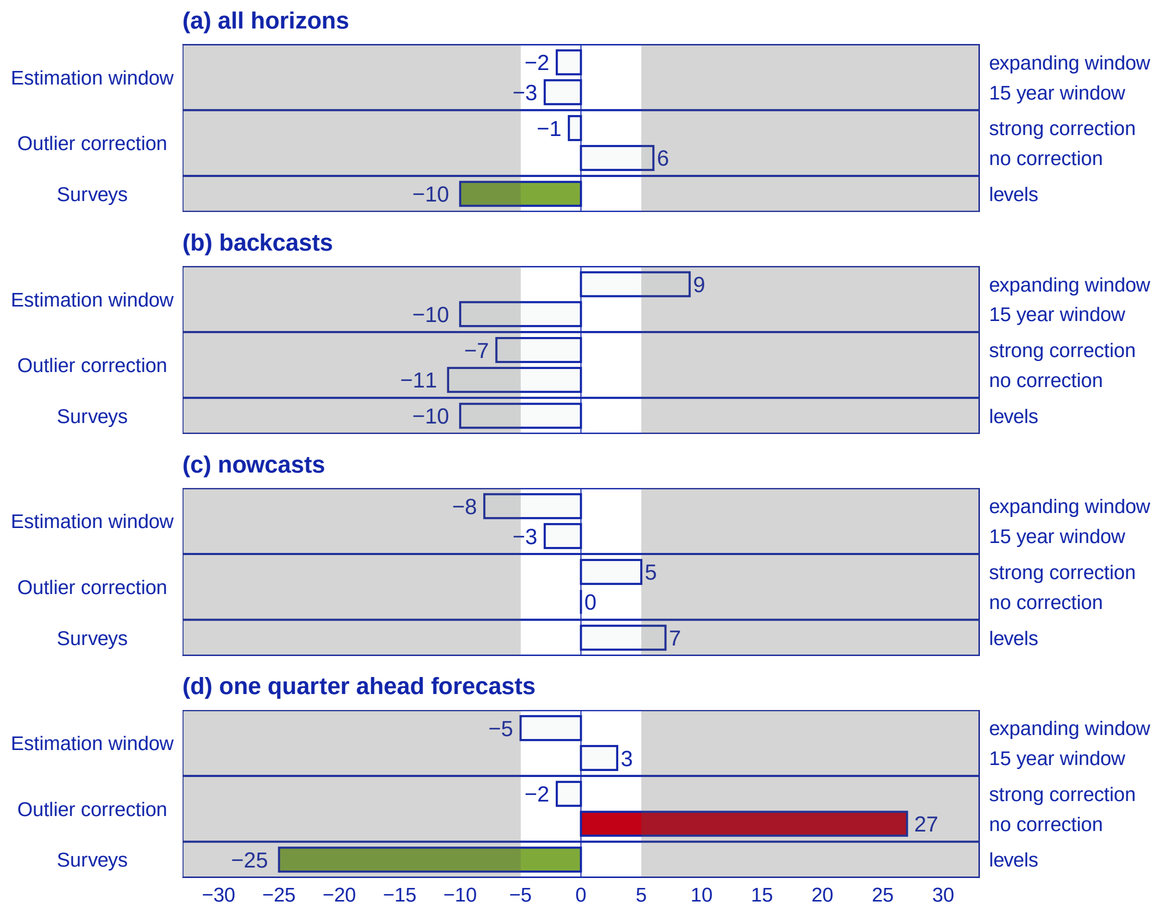
Notes: Squared forecast error regressed on all model features via stacked regression. A colored bar indicates the coefficient is statistically significant at the 5% level. If a bar is inside the shaded the absolute errors differs by more than ±5%. All effects against a model with the following specifications: small dataset, 1 factor, 2 lags, estimation with a rolling window of 10 years, using both public & non−public data, no quarterly variables included, only extreme outliers deleted, surveys in first difference, vintages downloaded on the first of the month.

Estimation period: 2013Q3−2019Q4. Regression diagnostics: Panel a: number of observations = 76,752, adjusted R−squared = 0.081, p−value = 0.000. Panel b: number of observations = 17,784, adjusted R−squared =0.141, p−value =0.000. Panel c: number of observations = 29,484, adjusted R−squared = 0.073, p−value = 0.000. Panel d: number of observations = 29,484, adjusted R−squared = 0.092, p−value = 0.000.

Source: Own calculations.

Figure A.3: Impact of estimation & transformation on forecast accuracy: pre-COVID period

Impact of features on out-of sample mean squared forecast, in percent



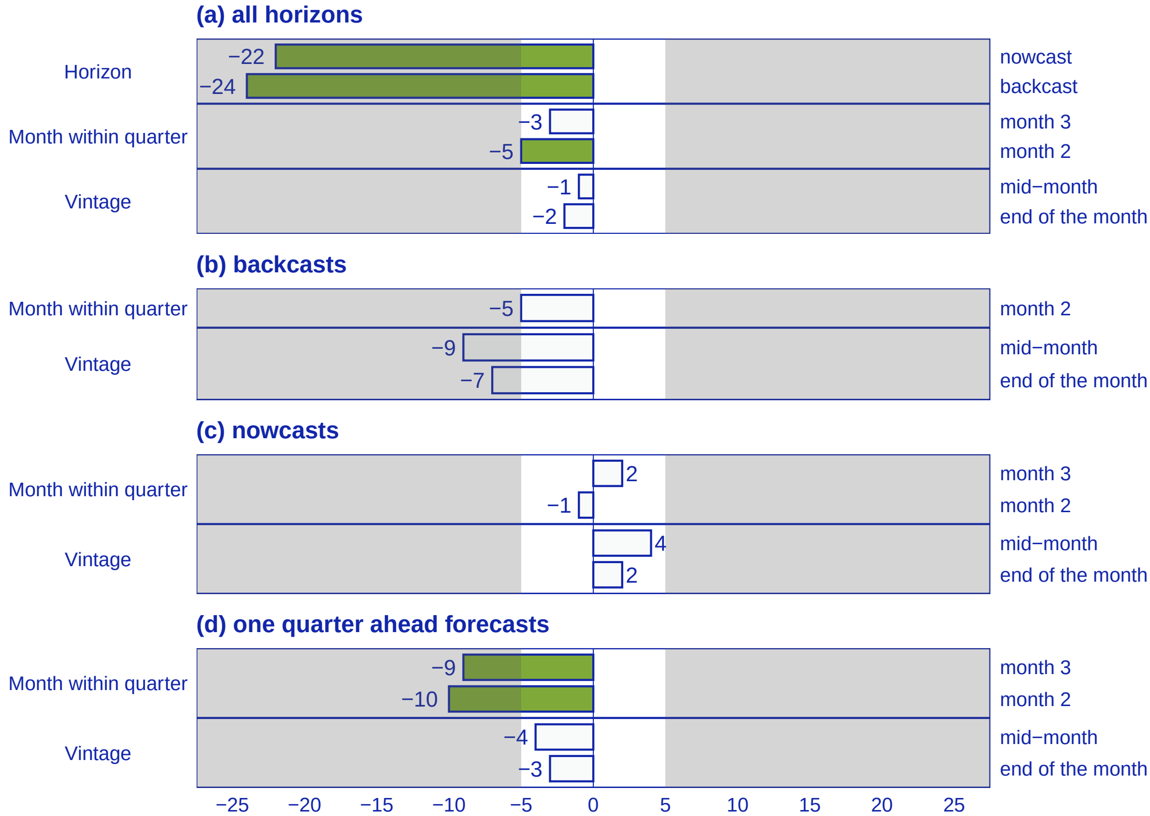
Notes: Squared forecast error regressed on all model features via stacked regression. A colored bar indicates the coefficient is statistically significant at the 5% level. If a bar is inside the shaded the absolute errors differs by more than ±5%. All effects against a model with the following specifications: small dataset, 1 factor, 2 lags, estimation with a rolling window of 10 years, using both public & non−public data, no quarterly variables included, only extreme outliers deleted, surveys in first difference, vintages downloaded on the first of the month.

Estimation period: 2013Q3−2019Q4. Regression diagnostics: Panel a: number of observations = 76,752, adjusted R−squared = 0.081, p−value = 0.000. Panel b: number of observations = 17,784, adjusted R−squared =0.141, p−value =0.000. Panel c: number of observations = 29,484, adjusted R−squared = 0.073, p−value = 0.000. Panel d: number of observations = 29,484, adjusted R−squared = 0.092, p−value = 0.000.

Source: Own calculations.

Figure A.4: Impact of timing on forecast accuracy: pre-COVID period

Impact of features on out-of sample mean squared forecast, in percent



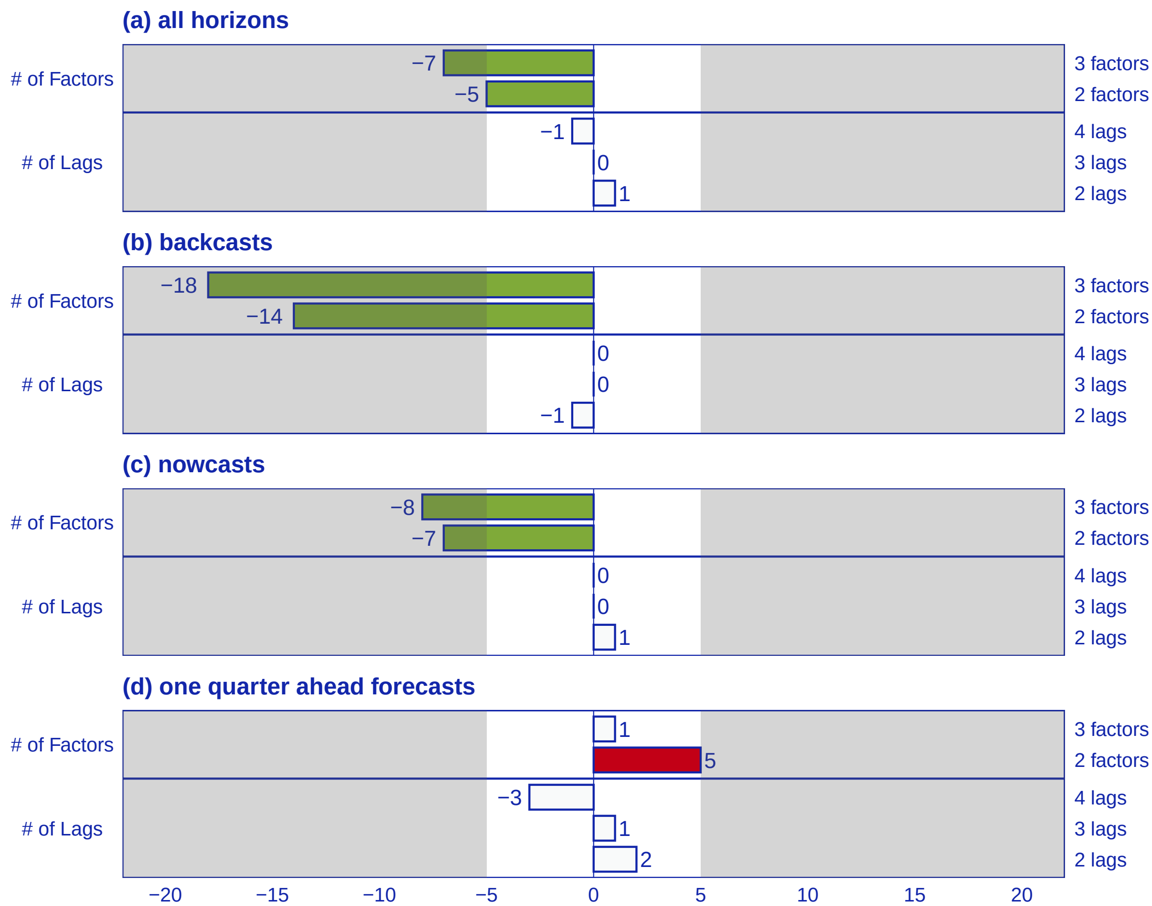
Notes: Squared forecast error regressed on all model features via stacked regression. A colored bar indicates the coefficient is statistically significant at the 5% level. If a bar is inside the shaded the absolute errors differs by more than ±5%. All effects against a model with the following specifications: small dataset, 1 factor, 2 lags, estimation with a rolling window of 10 years, using both public & non−public data, no quarterly variables included, only extreme outliers deleted, surveys in first difference, vintages downloaded on the first of the month.

Estimation period: 2013Q3−2019Q4. Regression diagnostics: Panel a: number of observations = 76,752, adjusted R−squared = 0.081, p−value = 0.000. Panel b: number of observations = 17,784, adjusted R−squared =0.141, p−value =0.000. Panel c: number of observations = 29,484, adjusted R−squared = 0.073, p−value = 0.000. Panel d: number of observations = 29,484, adjusted R−squared = 0.092, p−value = 0.000.

Source: Own calculations.

A.2 Results no-COVID period: 2013Q3 – 2019Q4 & 2020Q4-2023Q3

Figure A.5: Impact of model-specification on forecast accuracy: no COVID period
Impact of features on out-of sample mean squared forecast, in percent

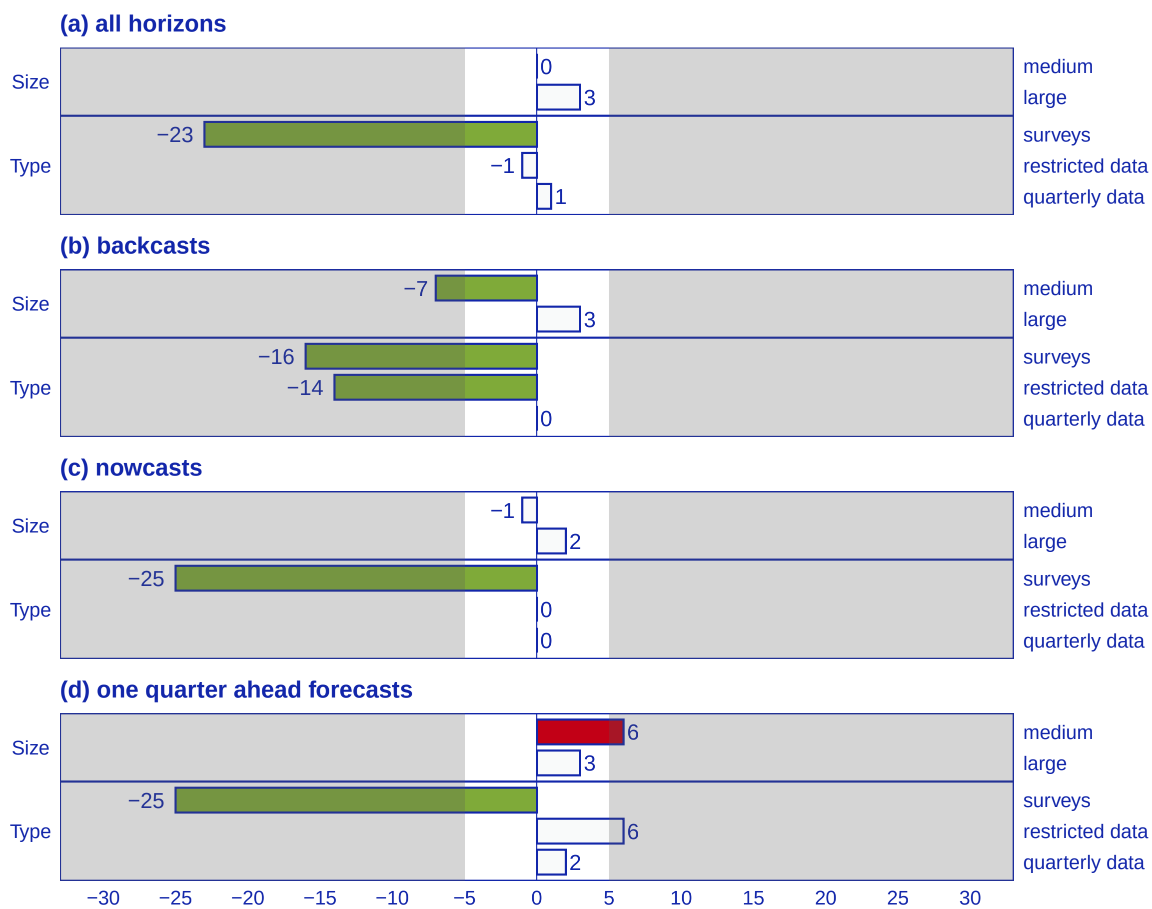


Notes: Squared forecast error regressed on all model features via stacked regression. A colored bar indicates the coefficient is statistically significant at the 5% level. If a bar is inside the shaded the absolute errors differs by more than ±5%. All effects against a model with the following specifications: small dataset, 1 factor, 2 lags, estimation with a rolling window of 10 years, using both public & non−public data, no quarterly variables included, only extreme outliers deleted, surveys in first difference, vintages downloaded on the first of the month.

Estimation period: 2013Q3−2019Q4, 2020Q4−2023Q3. Regression diagnostics: Panel a: number of observations = 113,904, adjusted R−squared = 0.218, p−value = 0.000. Panel b: number of observations = 26,424, adjusted R−squared =0.200, p−value =0.000. Panel c: number of observations = 43,740, adjusted R−squared = 0.198, p−value = 0.000. Panel d: number of observations = 43,740, adjusted R−squared = 0.300, p−value = 0.000.

Source: Own calculations.

Figure A.6: Impact of choice of data on forecast accuracy: no COVID period
Impact of features on out-of-sample mean squared forecast, in percent



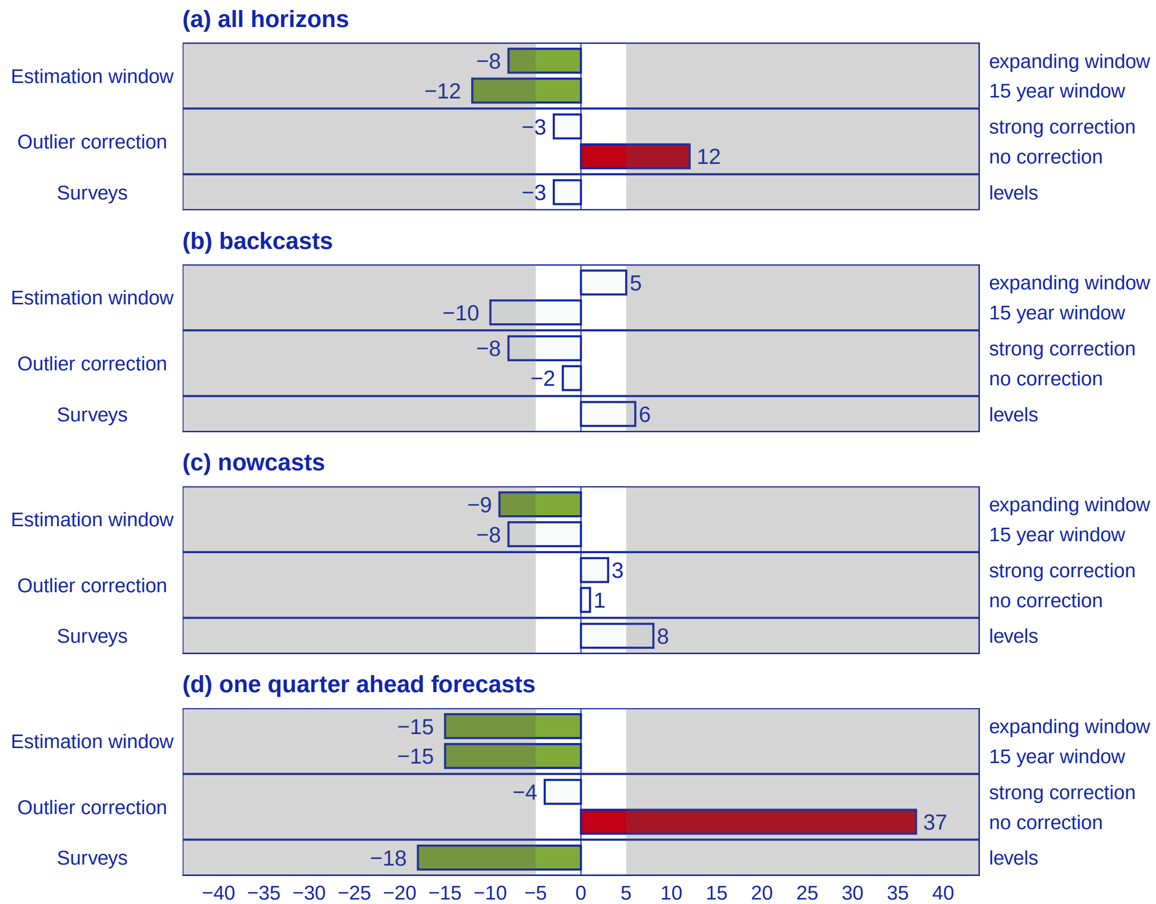
Notes: Squared forecast error regressed on all model features via stacked regression. A colored bar indicates the coefficient is statistically significant at the 5% level. If a bar is inside the shaded the absolute errors differs by more than ±5%. All effects against a model with the following specifications: small dataset, 1 factor, 2 lags, estimation with a rolling window of 10 years, using both public & non−public data, no quarterly variables included, only extreme outliers deleted, surveys in first difference, vintages downloaded on the first of the month.

Estimation period: 2013Q3−2019Q4, 2020Q4−2023Q3. Regression diagnostics: Panel a: number of observations = 113,904, adjusted R−squared = 0.218, p−value = 0.000. Panel b: number of observations = 26,424, adjusted R−squared =0.200, p−value =0.000. Panel c: number of observations = 43,740, adjusted R−squared = 0.198, p−value = 0.000. Panel d: number of observations = 43,740, adjusted R−squared = 0.300, p−value = 0.000.

Source: Own calculations.

Figure A.7: Impact of estimation & transformation on forecast accuracy: no COVID period

Impact of features on out-of sample mean squared forecast, in percent



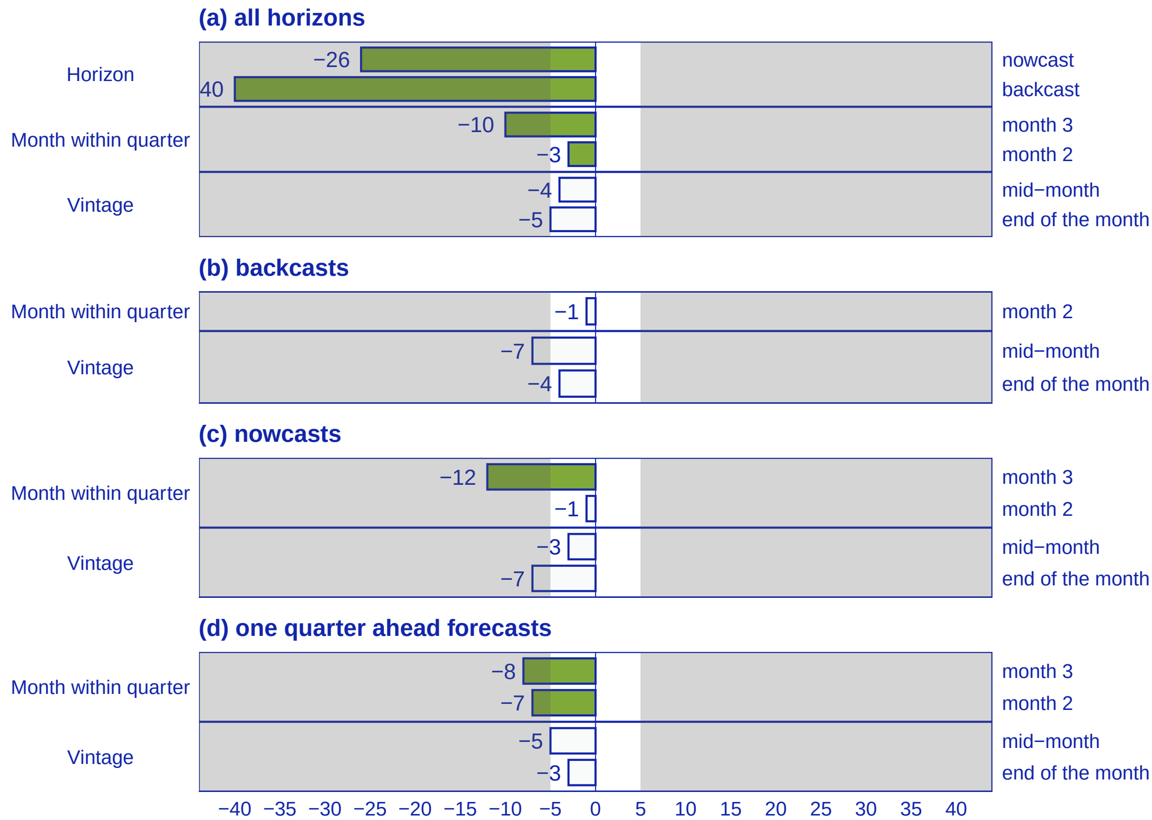
Notes: Squared forecast error regressed on all model features via stacked regression. A colored bar indicates the coefficient is statistically significant at the 5% level. If a bar is inside the shaded the absolute errors differs by more than ±5%. All effects against a model with the following specifications: small dataset, 1 factor, 2 lags, estimation with a rolling window of 10 years, using both public & non−public data, no quarterly variables included, only extreme outliers deleted, surveys in first difference, vintages downloaded on the first of the month.

Estimation period: 2013Q3−2019Q4, 2020Q4−2023Q3. Regression diagnostics: Panel a: number of observations = 113,904, adjusted R−squared = 0.218, p−value = 0.000. Panel b: number of observations = 26,424, adjusted R−squared =0.200, p−value =0.000. Panel c: number of observations = 43,740, adjusted R−squared = 0.198, p−value = 0.000. Panel d: number of observations = 43,740, adjusted R−squared = 0.300, p−value = 0.000.

Source: Own calculations.

Figure A.8: Impact of timing on forecast accuracy: no COVID period

Impact of features on out-of sample mean squared forecast, in percent



Notes: Squared forecast error regressed on all model features via stacked regression. A colored bar indicates the coefficient is statistically significant at the 5% level. If a bar is inside the shaded the absolute errors differs by more than ±5%. All effects against a model with the following specifications: small dataset, 1 factor, 2 lags, estimation with a rolling window of 10 years, using both public & non−public data, no quarterly variables included, only extreme outliers deleted, surveys in first difference, vintages downloaded on the first of the month.

Estimation period: 2013Q3−2019Q4, 2020Q4−2023Q3. Regression diagnostics: Panel a: number of observations = 113,904, adjusted R−squared = 0.218, p−value = 0.000. Panel b: number of observations = 26,424, adjusted R−squared =0.200, p−value =0.000. Panel c: number of observations = 43,740, adjusted R−squared = 0.198, p−value = 0.000. Panel d: number of observations = 43,740, adjusted R−squared = 0.300, p−value = 0.000.

Source: Own calculations.

A.3 Dataset

Table A.1: Overview of variables in small, medium-sized and large datasets

Nr.	Series	Group	Freq.	Dataset size S	Dataset size M	Dataset size L	Trans. Ln.	Trans. Diff.	Source	Start	Publ.lag (days)
1	Unemployment rate: 15-75 years	Labor market	M	X	X	X	O	X	CBS	jan-'83	17
2	Collectively agreed wages: all sectors	Prices & wages	M	O	X	X	X	X	CBS	jan-'00	7
3	Bankruptcies: total	Output & income	M	O	O	X	X	X	CBS	jan-'81	12
4	Construction: value added	Output & income	M	O	X	X	X	X	CBS	jan-'00	48
5	Building permits: firms	Output & income	M	O	X	X	X	X	CBS	jan-'90	54
6	Building permits: hiousing	Housing	M	O	X	X	X	X	CBS	jan-'95	54
7	House price index	Housing	M	X	X	X	X	X	CBS	jan-'95	22
8	Houses sold	Housing	M	X	X	X	X	X	CBS	jan-'95	22
9	Consumption: gas	Cons., orders & inv.	M	O	X	X	X	X	CBS	jan-'82	16
10	Industrial production: manufacturing	Output & income	M	X	X	X	X	X	CBS	jan-'95	39
11	Industrial production: food, bev. & tobacco	Output & income	M	O	X	X	X	X	CBS	jan-'95	39
12	Industrial production: chemical (products)	Output & income	M	O	X	X	X	X	CBS	jan-'95	39
13	Industrial production: machinery & equipment	Output & income	M	O	X	X	X	X	CBS	jan-'95	39
14	Consumption households: domestic	Cons., orders & inv.	M	X	X	X	X	X	CBS	jan-'95	38
15	Consumer confidence: headline	Cons., orders & inv.	M	X	X	X	O	X	CBS	apr-'86	-8
16	Consumer confidence: economic sit. < 12 m.	Cons., orders & inv.	M	O	O	X	O	X	CBS	apr-'86	-8
17	Consumer confidence: economic sit. > 12 m.	Cons., orders & inv.	M	O	O	X	O	X	CBS	apr-'86	-8
18	Consumer confidence: unemployment > 12 m.	Labor market	M	O	X	X	O	X	CBS	apr-'86	-8
19	Consumer conf.: purchase of dur. goods > yr.	Cons., orders & inv.	M	O	X	X	O	X	CBS	apr-'86	-8
20	Cons. confidence: buying/build.a home > 2 yr.	Housing	M	O	O	X	O	X	CBS	apr-'86	-8
21	Import of goods	Output & income	M	X	X	X	X	X	CBS	jan-'95	42
22	Export of goods	Output & income	M	X	X	X	X	X	CBS	jan-'95	42
23	Terms of trade	Prices & wages	M	O	O	X	X	X	CBS	jan-'95	42
24	Consumer price index: headline excl. energy	Prices & wages	M	O	X	X	X	X	CBS	jan-'96	1
25	Harmonised index of consumer prices: total	Prices & wages	M	X	X	X	X	X	CBS	jan-'96	1
26	Producer prices: manufacturing	Prices & wages	M	X	X	X	X	X	CBS	jan-'81	30
27	Flights: tonnes of goods	Output & income	M	O	O	X	X	X	CBS	jan-'99	32
28	Industrial confidence: headline	Output & income	M	X	X	X	O	X	Eurostat	jan-'80	0
29	Industrial confidence: employment > 3 m.	Labor market	M	O	X	X	O	X	Eurostat	apr-'82	0
30	Industrial confidence: export order books	Cons., orders & inv.	M	O	X	X	O	X	Eurostat	feb-'90	0
31	Industrial confidence: order-book	Cons., orders & inv.	M	O	O	X	O	X	Eurostat	jan-'80	0
32	Industrial confidence: production > 3 m.	Output & income	M	O	X	X	O	X	Eurostat	jan-'80	0
33	Industrial confidence: production < 3 m.	Output & income	M	O	O	X	O	X	Eurostat	jan-'80	0
34	Industrial confidence: stocks finished products	Cons., orders & inv.	M	O	O	X	O	X	Eurostat	jan-'80	0
35	Construction confidence: headline	Housing	M	O	X	X	O	X	Eurostat	jan-'80	0
36	Construction confidence: employment > 3 m.	Labor market	M	O	O	X	O	X	Eurostat	jan-'80	0
37	Construction confidence: evolution order books	Housing	M	O	O	X	O	X	Eurostat	jan-'80	0
38	Construction confidence: building activity < 3m.	Housing	M	O	O	X	O	X	Eurostat	jan-'80	0
39	Retail confidence: stock of finished products	Cons., orders & inv.	M	O	O	X	O	X	Eurostat	jan-'86	0
40	Retail confidence: headline	Output & income	M	O	X	X	O	X	Eurostat	jan-'86	0
41	Retail confidence: activity > 3 m.	Output & income	M	O	O	X	O	X	Eurostat	jan-'86	0
42	Retail confidence: employment > 3 m.	Labor market	M	O	O	X	O	X	Eurostat	jan-'86	0
43	Retail confidence: activity < 3 m.	Output & income	M	O	O	X	O	X	Eurostat	jan-'86	0
44	Economic sentiment: headline	Output & income	M	X	X	X	O	X	Eurostat	jan-'80	0
45	Services confidence: business situation < 3 m.	Output & income	M	O	X	X	O	X	Eurostat	apr-'93	0
46	Services confidence: demand > 3 m.	Output & income	M	O	O	X	O	X	Eurostat	jan-'96	0
47	Services confidence: demand < 3 m.	Output & income	M	O	O	X	O	X	Eurostat	apr-'93	0
48	Services confidence: headline	Output & income	M	X	X	X	O	X	Eurostat	jan-'96	0
49	Services confidence: employment > 3 m.	Labor market	M	O	X	X	O	X	Eurostat	apr-'93	0
50	Services confidence: employment < 3 m.	Labor market	M	O	O	X	O	X	Eurostat	apr-'93	0
51	Unemployment rate: < 25 years	Labor market	M	O	X	X	O	X	Eurostat	jan-'83	31
52	10-year government bond yield	Interest & exch. rates	M	X	X	X	O	X	Eurostat	apr-'86	10
53	Exchange rate eur/usd	Interest & exch. rates	M	O	X	X	X	X	Eurostat	jan-'99	2
54	Nominal effective exchange rate	Interest & exch. rates	M	X	X	X	X	X	Eurostat	jan-'99	30
55	Real effective exchange rate	Interest & exch. rates	M	O	O	X	X	X	Eurostat	jan-'99	30
56	Retail sales: mail orders & online stores	Output & income	M	O	X	X	X	X	Eurostat	jan-'00	39
57	Retail sales: total	Output & income	M	X	X	X	X	X	Eurostat	jan-'00	39
58	Consumer confidence: Belgium headline	Output & income	M	O	O	X	O	X	Eurostat	jan-'85	0
59	Consumer confidence: Germany headline	Output & income	M	O	O	X	O	X	Eurostat	jan-'85	0
60	Consumer confidence: eurozone headline	Output & income	M	O	X	X	O	X	Eurostat	jan-'85	0
61	Consumer confidence: Spain headline	Output & income	M	O	O	X	O	X	Eurostat	jun-'86	0
62	Consumer confidence: France headline	Output & income	M	O	O	X	O	X	Eurostat	jan-'85	0
63	Consumer confidence: Italy headline	Output & income	M	O	O	X	O	X	Eurostat	jan-'85	0
64	Producer confidence: Belgium headline	Output & income	M	O	O	X	O	X	Eurostat	jan-'85	0
65	Producer confidence: Germany headline	Output & income	M	O	O	X	O	X	Eurostat	jan-'80	0
66	Producer confidence: eurozone headline	Output & income	M	O	O	X	O	X	Eurostat	jan-'80	0
67	Producer confidence: Spain headline	Output & income	M	O	O	X	O	X	Eurostat	apr-'87	0
68	Producer confidence: France headline	Output & income	M	O	O	X	O	X	Eurostat	feb-'85	0
69	Producer confidence: Italy headline	Output & income	M	O	O	X	O	X	Eurostat	jan-'80	0
70	Passenger car registration	Cons., orders & inv.	M	O	X	X	X	X	ECB	jan-'90	48
71	Industrial production: comp. & electronics	Output & income	M	O	O	X	X	X	ECB	jan-'90	41
72	Credit to the domestic private sector	Money & credit	M	O	X	X	X	X	ECB	mrt-'70	26
73	Loans: households	Money & credit	M	X	X	X	X	X	ECB	sep-'97	26

Freq.: frequency of series (M=monthly, Q= Quarterly), Trans.: transformation of series, Ln: take logarithm (X= year, O= no), Diff.: take first difference of series (X= 0, O= yes), Source: source of series, Start: start mont/quarter of series, Publ. Lag: publication lag of series in days; negative publ. lag implies serie is released *n* days before the end of the month.

Continued on next page ...

Table A.1: Continued from previous page

Nr.	Series	Group	Freq.	Dataset size S	Dataset size M	Dataset size L	Trans. Ln.	Trans. Diff.	Source	Start	Publ.lag (days)
74	Loans: non-financial corporations	Money & credit	M	O	X	X	X	X	ECB	sep-'97	26
75	M1	Money & credit	M	O	O	X	X	X	ECB	jan-'80	26
76	M3	Money & credit	M	X	X	X	X	X	ECB	jan-'70	26
77	Int.% new loans nfc's: < 1 y. & <= eur 1 mil.	Interest & exch. rates	M	O	X	X	O	X	ECB	jan-'80	33
78	Int. % new loans hh. house purchase: < 1 y.	Interest & exch. rates	M	O	X	X	O	X	ECB	jan-'80	33
79	Interest rate new debt securityr: 10 years	Interest & exch. rates	M	O	X	O	O	X	ECB	jan-'93	12
80	3-month interest rate	Interest & exch. rates	M	O	X	X	O	X	ECB	okt-'72	1
81	country level index of financial stress	Stock market	M	O	O	X	O	X	ECB	jan-'70	25
82	Crude oil price: brent spot free on board	Prices & wages	M	O	X	X	X	X	ECB	feb-'99	1
83	1-year euribor interest rate	Interest & exch. rates	M	O	O	X	O	X	ECB	jan-'94	1
84	Industrial production: eurozone	Output & income	M	O	X	X	X	X	ECB	jan-'70	45
85	Retail trade: eurozone	Output & income	M	O	X	X	X	X	ECB	jan-'00	39
86	Unemployment rate: eurozone	Labor market	M	O	O	X	O	X	ECB	apr-'98	34
87	Gold price	Stock market	M	O	O	X	X	X	ECB	jan-'70	5
88	composite indicator of systemic stress	Stock market	M	O	O	X	O	X	ECB	jan-'99	1
89	Dow Jones Euro Stoxx: consumer services	Stock market	M	O	X	X	X	X	ECB	dec-'91	1
90	Commodity Price index: non-energy comm.	Prices & wages	M	X	X	X	X	X	ECB	jan-'96	5
91	PMI manufacturing: eurozone headline	Output & income	M	O	O	X	O	X	Refinitiv	jun-'97	3
92	PMI services: eurozone headline	Output & income	M	O	O	X	O	X	Refinitiv	jul-'98	5
93	PMI manufacturing: headline	Output & income	M	X	X	X	O	X	Refinitiv	mrt-'00	1
94	PMI manufacturing: production	Output & income	M	O	O	X	O	X	Refinitiv	mrt-'00	1
95	PMI manufacturing: new orders	Cons., orders & inv.	M	X	X	X	O	X	Refinitiv	mrt-'00	1
96	PMI manufacturing: employment	Labor market	M	O	O	X	O	X	Refinitiv	mrt-'00	1
97	AEX-midkap	Stock market	M	O	O	X	X	X	Refinitiv	jan-'83	1
98	Dow Jones Euro Stoxx: 50	Stock market	M	O	X	X	X	X	Refinitiv	jan-'87	1
99	Dow Jones Euro Stoxx: basic materials	Stock market	M	O	O	X	X	X	Refinitiv	jan-'87	1
100	Dow Jones Euro Stoxx: technology	Stock market	M	O	O	X	X	X	Refinitiv	jan-'87	1
101	Dow Jones Euro Stoxx: industrials	Stock market	M	O	O	X	X	X	Refinitiv	jan-'87	1
102	AEX	Stock market	M	X	X	X	X	X	Refinitiv	jan-'83	1
103	PMI composite: world (headline)	Output & income	M	X	X	X	O	X	Refinitiv	jul-'98	5
104	World steel production	Output & income	M	O	X	X	X	X	Refinitiv	jan-'80	21
105	Baltic exchange dry index	Prices & wages	M	O	O	X	X	X	Refinitiv	mei-'85	1
106	VIX Europe	Stock market	M	O	O	X	X	X	Refinitiv	feb-'99	1
107	Serv. conf.: ev. of de. < 3 m.; pc. progr. & conslt.	Output & income	M	O	X	X	O	X	EC	apr-'97	0
108	Serv. conf.: ev. of de. < 3 m.: conslt. & head off.	Output & income	M	O	X	X	O	X	EC	apr-'97	0
109	Serv. conf.: ev. of de. < 3 m.: empl. activities	Output & income	M	O	X	X	O	X	EC	apr-'97	0
110	Global supply chain pressure index	Output & income	M	O	O	X	O	X	FED NY	jan-'98	4
111	Debit card payments: total	Cons., orders & inv.	M	O	X	X	X	X	DNB	jan-'95	10
112	Debit card payments: services	Cons., orders & inv.	M	O	O	X	X	X	DNB	jan-'98	10
113	US high yield bond spread	Stock market	M	O	O	X	X	X	DNB	jan-'97	0
114	Eurozone high yield bond spread	Stock market	M	O	X	X	X	X	DNB	jan-'98	0
115	FD: headline	Output & income	M	X	O	O	O	O	FD	jan-'85	0
116	FD: financial markets (FM)	Stock market	M	O	X	O	O	O	FD	jan-'85	0
117	FD: companies (COM)	Output & income	M	O	X	O	O	O	FD	jan-'85	0
118	FD: economics (ECO)	Output & income	M	O	X	O	O	O	FD	jan-'85	0
119	FD: politics (POL)	Output & income	M	O	X	O	O	O	FD	jan-'85	0
120	FD FM: financial markets	Stock market	M	O	O	X	O	O	FD	jan-'85	0
121	FD FM: financials	Stock market	M	O	O	X	O	O	FD	jan-'85	0
122	FD FM: news	Stock market	M	O	O	X	O	O	FD	jan-'85	0
123	FD FM: financial indices	Stock market	M	O	O	X	O	O	FD	jan-'85	0
124	FD COM: infrastructure	Output & income	M	O	O	X	O	O	FD	jan-'85	0
125	FD COM: multinationals	Output & income	M	O	O	X	O	O	FD	jan-'85	0
126	FD COM: construction & energy	Output & income	M	O	O	X	O	O	FD	jan-'85	0
127	FD COM: demography	Output & income	M	O	O	X	O	O	FD	jan-'85	0
128	FD ECO: elections	Output & income	M	O	O	X	O	O	FD	jan-'85	0
129	FD ECO: indicators	Output & income	M	O	O	X	O	O	FD	jan-'85	0
130	FD ECO: raw materials	Output & income	M	O	O	X	O	O	FD	jan-'85	0
131	FD ECO: EU	Output & income	M	O	O	X	O	O	FD	jan-'85	0
132	FD POL: parliament	Output & income	M	O	O	X	O	O	FD	jan-'85	0
133	FD POL: national	Output & income	M	O	O	X	O	O	FD	jan-'85	0
134	FD POL: lower government	Output & income	M	O	O	X	O	O	FD	jan-'85	0
135	FD POL: social partners	Output & income	M	O	O	X	O	O	FD	jan-'85	0
136	Consumption	Cons., orders & inv.	Q	O	X	X	X	X	CBS	mrt-'96	45
137	Gross private investment	Output & income	Q	O	X	X	X	X	CBS	mrt-'96	45
138	Exports	Output & income	Q	O	X	X	X	X	CBS	mrt-'96	45
139	Imports	Output & income	Q	O	X	X	X	X	CBS	mrt-'96	45
140	Hours worked	Labor market	Q	X	X	X	X	X	CBS	mrt-'95	45
141	Labour productivity	Labor market	Q	O	X	X	X	X	Eurostat	mrt-'96	45
142	Capacity utilization	Output & income	Q	X	X	X	O	X	Eurostat	mrt-'80	-3
143	Gross domestic product: eurozone	Output & income	Q	X	X	X	X	X	Eurostat	mrt-'95	45
144	Gross domestic product	Output & income	Q	X	X	X	X	X	CBS	mrt-'96	45
	N (monthly)			24	67	129					
	N (quarterly)			4	9	9					
	N (total)			28	76	138					

Freq.: frequency of series (M=monthly, Q= Quarterly), Trans.: transformation of series, Ln: take logarithm (X= year, O= no), Diff.: take first difference of series (X= 0, O= yes), Source: source of series, Start: start mont/quarter of series, Publ. Lag: publication lag of series in days; negative publ. lag implies serie is released *n* days before the end of the month.