International Journal of Forecasting 32 (2016) 411–436

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Forecasting and nowcasting real GDP: Comparing statistical models and subjective forecasts

W. Jos Jansen [a], Xiaowen Jin [b], Jasper M. de Winter [c,*]

[a] *Financial and Economic Policy Department, Ministry of Finance, P.O. Box 20201, The Hague, Netherlands*

[b] *Munich Graduate School of Economics, Ludwig Maximilians Universität, P.O. Box 1111, München, Germany*

[c] *Economic Policy and Research Division, De Nederlandsche Bank, P.O. Box 98, 1000 AB, Amsterdam, Netherlands*

A R T I C L E I N F O

Keywords:
Forecasting competitions
Factor models
Professional forecasters
Judgment

A B S T R A C T

We conduct a systematic comparison of the short-term forecasting abilities of twelve statistical models and professional analysts in a pseudo-real-time setting, using a large set of monthly indicators. Our analysis covers the euro area and its five largest countries over the years 1996–2011. We find summarizing the available monthly information in a few factors to be a more promising forecasting strategy than averaging a large number of single-indicator-based forecasts. Moreover, it is important to make use of all available monthly observations. The dynamic factor model is the best model overall, particularly for nowcasting and backcasting, due to its ability to incorporate more information (factors). Judgmental forecasts by professional analysts often embody valuable information that could be used to enhance the forecasts derived from purely mechanical procedures.

1. Introduction

Information on economic activity and its short-term prospects is very important for decision makers in governments, central banks, financial markets and non-financial firms. Monetary and economic policy makers and economic agents have to make decisions in real time based on incomplete and inaccurate information on current economic conditions. A key indicator of the state of the economy is the growth rate of real GDP, which is available on a quarterly basis only, and is also subject to substantial publication lags. In many countries, an initial estimate of quarterly real GDP is published around six weeks after the end of the quarter. Moreover, real GDP data are subject to revisions that can be substantial, as more data become available to statistical offices over time.

Fortunately, though, a lot of statistical information related to economic activity is published on a more frequent and timely basis. This information includes data on industrial production, prices of goods and services, expenditures, unemployment, financial market prices, loans, and consumer and business confidence. Recently, the forecasting literature has developed several statistical approaches for exploiting this potentially very large information set in order to improve the assessment of both real GDP growth in the current quarter (nowcast) and its development in the near future. Examples of such approaches include bridge models, factor models, mixed-data sampling models (MIDAS) and mixed-frequency vector-autoregressive (MFVAR) models. These models differ in their solutions to the practical problems of dealing with large information sets and the fact that the auxiliary variables are observed at different frequencies and with different publication lags.

Practitioners now have a wealth of statistical models to choose from; but which one should they use? As each model has its own strengths and weaknesses, it is difficult

* Corresponding author.
E-mail addresses: W.J.Jansen@minfin.nl (W. Jos Jansen), xiaowenjin1115@gmail.com (Xiaowen Jin), J.M.de.Winter@dnb.nl (Jasper M. de Winter).

http://dx.doi.org/10.1016/j.ijforecast.2015.05.008

to make a decision on purely theoretical grounds. The ranking of the models in terms of forecasting abilities, and the extent to which this varies with the prediction horizon or the economic circumstances, has to be determined by empirical analysis. On these issues the jury is still out, however, as large-scale comparative studies are scarce. In many papers, the empirical work refers only to a single country, and usually only limited numbers of models are included. Furthermore, studies differ in the size of the information set and the sample period used.[1]

This paper is motivated by this gap in the empirical literature. We undertake a systematic comparison of a broad range of linear statistical models – twelve models in all – that have been applied in the recent literature. For the sake of comparability and robustness, we include five countries (Germany, France, Italy, Spain and the Netherlands) and the euro area in our analysis, and utilize an information set that is as homogeneous as possible across geographical entities. Moreover, our sample includes the volatile episode of the financial crisis of 2008 and its aftermath, which may make it easier to distinguish between the various models. We contrast the models' forecasting abilities before 2008 with those during the crisis period. This may be of great interest for policy makers, financial analysts and economic agents alike, as information on where the economy stands and where it is headed in the immediate short run is particularly valuable at times of great uncertainty.

The provision of cross-country evidence on the relative performances of twelve different statistical forecasting models is our first contribution to the literature. Model forecasts are the result of purely mechanical recipes, and do not incorporate subjective elements. Our second contribution concerns the potential usefulness of forecasts made by professional analysts (published by Consensus Forecasts on a quarterly basis). From a practical point of view, such forecasts are very cheap and easy to use. Moreover, as an expression of the "wisdom of crowds", they may reflect much more information than the statistical information set, which is inevitably limited. A questionnaire conducted by the European Central Bank (ECB) among the participants of the ECB Survey of Professional Forecasters found that the panelists regard 40% of their short-term GDP forecasts as being judgment-based (Meyler & Rubene, 2009). We investigate the extent to which the subjective forecasts by analysts in our sample contain information beyond that generated by the best mechanical statistical models.

The remainder of the paper is structured as follows. Section 2 describes the various statistical models and discusses how they deal with the challenges posed by large and irregularly shaped datasets. Section 3 describes the data, our pseudo real-time forecast design, and other specification issues. Sections 4 and 5 present the results for the mechanical models and the professional forecasts, respectively. Section 6 summarizes our findings and concludes.

2. Linear statistical models for short-term GDP forecasting

2.1. Overview

In practice, taking advantage of auxiliary information for the forecasting of real GDP in the immediate short run poses several challenges. The first challenge is posed by the large size of the information set. There are countless potentially useful variables for forecasting GDP, and often they are also interrelated. The datasets used in the empirical literature vary greatly in size, and may include more than 300 variables. Moreover, the limited length of the time series involved makes over-parametrization a real issue. The second problem relates to the fact that the indicator variables are observed more frequently (monthly, weekly, daily) than GDP. Moreover, the dating of the most recent observation may vary across indicators because of differences in publication lags. This is known as the "ragged edge" problem; see Wallis (1986).

The various statistical approaches in the literature deal with these challenges in different ways. Broadly speaking, a forecasting procedure involves two transformations of the dataset of indicators in order to produce a final forecast: an aggregation and the application of a forecasting tool, which links auxiliary variables to real GDP growth. These two transformations can be executed in either order, representing two fundamentally different strategies. The first strategy begins by computing an indicator-specific GDP forecast for each variable, which are then aggregated into a single final forecast in the second step. We call this strategy the "pooling forecasts strategy". In this approach, it is necessary to specify the weighting scheme for the individual forecasts. A basic scheme is the simple average, which gives each forecast an equal weight, but weights may also be computed recursively depending on the indicators' (recent) forecasting performances. Examples of the forecast pooling strategy include bridge equations and VAR models. In contrast, the "aggregating information strategy" takes the aggregation step first, by summarizing the large dataset in a small number of series. This strategy exploits the fact that the auxiliary variables are correlated. Factor analysis is used to replace a large number of correlated time series with a limited number of uncorrelated (unobserved) factors representing the common information component of the original data series. The implicit weights (factor loadings) are determined from the correlation patterns in the original dataset. The factors serve as inputs for the forecasting procedure in the next step. Examples of this modeling strategy include dynamic factor models and factor augmented versions of forecasting models that pool forecasts. Finally, a recent development is estimation using Bayesian shrinkage on coefficients, which translates a large set of indicators into a single GDP forecast directly, without a clear

[1] Rünstler et al. (2009) form an important exception, comparing three factor models, a bridge model and a quarterly VAR model for ten European countries; however, their study does not include the financial crisis. Kuzin, Marcellino, and Schumacher (2013) analyzed the relative forecasting performances of MIDAS models and dynamic factor models, including part of the crisis years (2008–2009). Liebermann (2012) analyzed the relative forecasting performances of a range of models over the years 2001–2011, but only for the United States.

aggregation step. This approach implicitly aggregates information by applying Bayesian shrinkage to the parameters.

The specification of the forecasting tool is the second feature that distinguishes the approaches. The traditional approaches, such as bridge models and VAR models, rely on forecasting equations that are cast solely in quarterly terms. This means that (forecasts of) monthly indicator variables first have to be aggregated to quarterly averages, before they can be used for forecasting GDP. Moreover, the monthly observations available are not exploited fully by quarterly VAR models. As this may not be an efficient use of the available information, recently developed approaches accommodate both quarterly and monthly data within the same equation or system of equations. These approaches also take publication lags into account. The mixed-frequency VAR (MFVAR) model treats GDP as an unobserved monthly variable in a state space framework. Monthly GDP is related to quarterly GDP via an identity. The quarterly GDP growth rate is observed only in the third month of each quarter. The mixed-data sampling (MIDAS) design relates quarterly GDP directly to a large number of lags of monthly data series, using a parsimonious specification of the lag structure.

A third, and more practical, specification issue is whether or not to include GDP's own past in the forecasting tool. In general, forecasting equations can be augmented easily with autoregressive (AR) terms. Several authors have found that the AR versions of models tend to result in modest improvements in forecasting performances (e.g., Foroni & Marcellino, 2014). In this paper, we analyze twelve statistical models: (1) bridge model (BEQ), (2) BEQ with AR terms (BEQ-AR), (3) quarterly VAR model (QVAR), (4) factor-augmented quarterly VAR (F-VAR), (5) Bayesian quarterly VAR (BVAR), (6) dynamic factor model (DFM), (7) mixed-frequency VAR model (MFVAR), (8) factor-augmented MFVAR (F-MFVAR), (9) mixed data sampling model (MIDAS), (10) MIDAS with AR terms (MIDAS-AR), (11) factor-augmented MIDAS (F-MIDAS), and (12) F-MIDAS with AR terms (F-MIDAS-AR). The next three subsections discuss the forecasting models briefly, starting with the quarterly models. To improve the flow of the discussion, we discuss the selection of the weighting scheme for indicator-based forecasts in Section 3.3. Moreover, we have moved some technical details to Appendix A.2.

We begin by clarifying our notation. Below, $t = 1, \ldots, T$ stands for a quarterly time index and $m = 1, \ldots, T_m + w$ for a monthly time index. T indicates the latest available quarterly observation, and T_m corresponds to the third month of quarter T; hence, $T_m = 3T$. $T_m + w$ indicates the latest available monthly observation, where w is the time difference in months between the most recent observation of the indicator and the GDP on the monthly time scale. Quarterly averages of monthly figures are denoted by the superscript ^{Q}. The quarterly GDP in quarter t, y_t^Q, is assigned to month $3t$ on the monthly time scale. Formally, $y_t^Q = \frac{1}{3}(y_{3t} + y_{3t-1} + y_{3t-2})$, where y_{3t}, y_{3t-1} and y_{3t-2} are unobserved three-month GDP growth rates, i.e., growth rates vis-à-vis the same month of the previous quarter. The matrix of monthly indicators (indexed by $i = 1, \ldots, n$) is defined as $x_m = (x_{1,m}, \ldots, x_{n,m})'$. The monthly series x_m have been transformed as three-month growth rates or differences. The matrix of monthly indicators aggregated to quarterly values is defined as $x_t^Q = (x_{1,t}^Q, \ldots, x_{n,t}^Q)'$. The quarterly GDP growth forecast for quarter $t + h$ at time t is denoted as $y_{t+h|t}^Q$. For quarterly models, time is always measured on a quarterly time scale. In mixed frequency approaches, the quarterly and monthly time scales intermingle.

2.2. Quarterly models for GDP growth

2.2.1. Bridge equation (BEQ)

The quarterly bridge equation is a method that has been used widely for forecasting the GDP using all available observations of monthly indicators; for applications, see Baffigi, Golinelli, and Parigi (2004), Kitchen and Monaco (2003) and Rünstler and Sédillot (2003). Bridge equations are linear regressions that "bridge" monthly variables, such as industrial confidence and retail sales, to quarterly real GDP. Usually, the monthly indicators are not known over the entire projection horizon. Various different specifications are possible within this approach. Here, we analyze a simple version of the bridge equation, proceeding in two steps. Firstly, we obtain predictions of the necessary monthly values of indicator x_i over the forecasting horizon, with the help of univariate autoregressive models, and aggregate these to appropriate quarterly values x_i^Q. Secondly, we use these quarterly aggregates to predict the GDP. The bridge model for x_i is:

$$y_t^Q = \alpha + \sum_{s=0}^{p} \beta_s x_{i,t-s}^Q + \varepsilon_{i,t}^Q, \quad \varepsilon_{i,t}^Q \sim N(0, \sigma_{\varepsilon^Q}^2), \tag{1}$$

where α is a constant, p denotes the number of lags in the bridge equation, and ε_i^Q is a normally distributed error-term. We estimate Eq. (1) for each of the n indicators, and then calculate the final forecast by weighting the n indicator-specific forecasts for each horizon. The lag parameter p is determined recursively by the Schwartz information criterion (SIC), with the maximum number of lags set to four.

2.2.2. Vector autoregressive model (QVAR)

The VAR approach is very similar to the bridge equation approach. Unlike bridge equations, VAR models use the information content of GDP itself to produce GDP forecasts (e.g., Camba-Mendez, Kapetanios, Smith, & Weale, 2001). Moreover, it is a system approach, attempting to exploit the interdependence of indicator and real GDP dynamics. However, misspecification anywhere in the system may affect the accuracy of the GDP predictions. More importantly, due to its quarterly time frame, the QVAR model only uses monthly observations that correspond to a full quarter. Consequently, it does not exploit the available monthly information fully. We estimate n quarterly bivariate VAR models that include one of the indicators and GDP growth:

$$z_{i,t}^Q = \alpha + \sum_{s=1}^{p_i} A_s z_{i,t-s}^Q + \varepsilon_{i,t}^Q, \quad \varepsilon_{i,t}^Q \sim N(0, \Sigma_{\varepsilon^Q}), \tag{2}$$

where $z_{i,t}^Q = (y_t^Q, x_{i,t}^Q)'$. From each bivariate VAR, we obtain an indicator-specific GDP forecast $y_{t+h|t}^Q$. As in the case of the bridge model, we form the final forecast as a weighted average of the individual forecasts. The lag parameter p is determined recursively by the SIC, with a maximum of four.

2.2.3. Bayesian VAR model (BVAR)

The number of variables in multivariate VAR models is usually between three and ten, because the number of unrestricted parameters that can be estimated reliably is rather limited. Bayesian vector autoregression with shrinkage is able to handle large unrestricted VARs (Carriero, Clark, & Marcellino, 2012; Giannone, Lenza, & Primiceri, 2012). We estimate a quarterly Bayesian VAR model along the lines proposed by Bańbura, Giannone, and Reichlin (2010). Accordingly, we include all variables in log levels, except for those that are expressed as rates, and derive the GDP forecast $y_{t+h|t}^Q$ from the following VAR system:

$$Z_t^Q = \alpha + \sum_{s=1}^{p} A_s Z_{t-s}^Q + \vartheta_t^Q, \quad \vartheta_t^Q \sim WN(0, \Sigma_{\vartheta^Q}), \tag{3}$$

where $Z_t^Q = (y_t^Q, x_{i,t}^Q \dots x_{n,t}^Q)'$. Moreover, the moments for the prior distribution of the coefficients are:

$$E[(A_k)_{ij}] = \begin{cases} \delta_i, & j = i, k = 1; \\ 0, & \text{otherwise} \end{cases}$$
$$V[(A_k)_{ij}] = \begin{cases} \dfrac{\lambda^2}{k^2}, & j = i; \\ \dfrac{\lambda^2}{k^2}\dfrac{\sigma_i^2}{\sigma_j^2}, & \text{otherwise.} \end{cases} \tag{4}$$

The coefficients $A_1, \dots, A_k$ are assumed a priori to be independent and normally distributed. The covariance matrix of the residuals is assumed to be diagonal, fixed and known: $\Psi = \Sigma$, where $\Sigma = \mathrm{diag}(\sigma_1^2, \dots, \sigma_n^2)$. For non-stationary variables, we use the random walk prior: $\delta_i = 1$. For stationary variables, we use the white noise prior: $\delta_i = 0$. The parameter λ governs the degree of shrinkage. If $\lambda = 0$, the posterior equals the prior and the data do not influence the estimates. At the other extreme, $\lambda = \infty$, the posterior expectations coincide with the ordinary least squares (OLS) estimates. The degree of shrinkage should be chosen so as to prevent over-fitting, while preserving the relevant sample information. Bańbura et al. (2010) argue that λ should be set relative to the size of the system. The reasoning here is that if all data series contain similar information, the relevant signal can still be extracted efficiently from a large data set, despite the higher shrinkage that is required to filter out the systematic component. We determine λ recursively in a fashion similar to Bańbura et al. (2010).[2] The factor $\frac{1}{k^2}$ is the rate at which the prior variance decreases with the lag length, and $\frac{\sigma_i^2}{\sigma_j^2}$ accounts for the differences in the scale and variability of the data. Following Bańbura et al. (2010) we set the maximum number of lags p to five.

2.3. Mixed frequency models

Interest in mixed frequency models has increased among academics and policy makers in recent years, because of the general failure of simple quarterly models to predict or signal the sharp downturn of the economy at the onset of the financial crisis. Here, we investigate the dynamic factor model, the mixed frequency VAR and the MIDAS approach.[3] All of these models utilize the available monthly information fully.

2.3.1. Dynamic factor model (DFM)

Dynamic factor models summarize the information contained in the dataset using a limited number of factors, the dynamic behaviors of which are specified as vector-autoregressive processes. A key feature of this approach is the use of the Kalman filter, which allows for an efficient handling of the unbalancedness of the dataset and the different frequencies of the data. The Kalman filter replaces any missing monthly indicator observations with optimal predictions, and also generates estimates of the unobserved monthly real GDP, subject to a temporal aggregation constraint for the quarterly observation. Dynamic factor models have been shown to produce relatively accurate macroeconomic forecasts for many countries.[4]

In this paper we analyze the dynamic factor model proposed by Bańbura and Rünstler (2011), which is used by several central banks within the euro area. The first equation of the model is

$$x_m = \Lambda f_m + \xi_m, \quad \xi_m \sim N(0, \Sigma_\xi), \tag{5}$$

which relates the n monthly indicators x_m to r monthly static factors $f_m = (f_{1,m}, \dots, f_{r,m})'$ via a matrix of factor loadings Λ and an idiosyncratic component $\xi_m = (\xi_{1,m}, \dots, \xi_{n,m})'$, where $r \ll n$. The DFM assumes that the idiosyncratic components are a multivariate white noise process, hence the covariance matrix Σ_ξ is diagonal. Furthermore, the DFM assumes that the factors follow a vector-autoregressive process of order p:

$$f_m = \sum_{s=1}^{p} A_s f_{m-s} + \zeta_m, \quad \zeta_m \sim (0, \Sigma_\zeta), \tag{6}$$

2 Bańbura et al. (2010) begin by defining FIT, an in-sample measure of fit for three key variables (output, inflation and the short term interest rate). They select λ such that the BVAR model and an unrestricted VAR model featuring the three key variables produce the same numerical value for FIT. In our case, we take real GDP, HICP inflation and the three-month interest rate as the key variables, and calculate FIT and λ recursively. Like Bloor and Matheson (2011), we found preliminary calculations to indicate that the FIT measure thus specified gave rise to over-fitting. We have therefore applied the procedure to FIT/2.

3 Recently, Bayesian mixed frequency regressions and VAR models have been developed and applied to nowcasting. See e.g. Carriero et al. (2012) and Schorfheide and Song (2013). These alternative approaches are outside the scope of this paper.

4 Examples include Giannone, Reichlin, and Small (2008) for the United States; Bańbura, Giannone, and Reichlin (2011), Bańbura and Modugno (2014), Camacho and Perez-Quiros (2010) and Rünstler et al. (2009) for the euro area; Schumacher and Breitung (2008) for Germany; Schneider and Spitzer (2004) for Austria; Cheung and Demers (2007) for Canada; Camacho and Quiros (2011) for Spain; and den Reijer (2013) for the Netherlands.

where A is a square $r \times r$ matrix. Moreover, the covariance matrix of the VAR (σ_ζ) is driven by a q dimensional standardized white noise process η_m:

$$\zeta_m = B\eta_m, \quad \eta_m \sim N(0, I_q), \tag{7}$$

where B is a $r \times q$ matrix and $q \leq r$. The final equation is a forecasting equation linking the factors to (unobserved) mean-adjusted real GDP growth:

$$y_m = \beta' f_m + \varepsilon_m, \quad \varepsilon_m \sim N(0, \sigma_\varepsilon^2), \tag{8}$$

where y_m denotes the unobserved monthly GDP growth rate. The model is estimated in four steps. In the first step, we obtain the factor loadings Λ and the estimated static factors $\hat{f}_m$. In the second step, we estimate the coefficient matrices A_s in Eq. (6) and β in Eq. (8) by OLS using $\hat{f}_m$. In the third step, we compute ζ_m and its covariance matrix Σ_ζ and obtain an estimate of the matrix B by principal components analysis. In the final step, we cast the model in state space and use the Kalman filter and smoother to re-estimate the estimated factors ($\hat{f}_m$) and the monthly GDP growth.[5] To estimate the model, we need to specify the numbers of static and dynamic common factors, denoted by r and q respectively. We set the largest possible value of r to six, based on the scree test of Cattell (1966). Moreover, $q \leq r$ by definition. In view of potential misspecifications and instabilities, we follow Kuzin et al. (2013) and refrain from choosing a particular combination of r and q, but take the (unweighted) average of forecasts over all possible parameterizations in terms of the numbers of static and dynamic factors and the number of lags p in Eq. (6), with $p \leq 6$. The total number of model specifications is $p(r+1)r/2 = 126$.[6]

2.3.2. Mixed frequency vector autoregressive model (MFVAR)

Mixed frequency VAR models (MFVAR) are VAR models that allow for data series with different frequencies. In contrast to the quarterly VAR model, the MFVAR model exploits all of the available monthly information fully. It shares with the QVAR model the strengths and weaknesses of a system approach. In our case, we focus on bivariate MFVAR models featuring a monthly indicator, unobserved monthly GDP and a temporal aggregation scheme.

Let $z_{i,m} = (y_m, x_{i,m})'$ be the vector of the latent monthly real GDP and indicator $x_{i,m}$. The vector follows a VAR model:

$$z_{i,m} - \mu_i = \sum_{s=1}^{p} A_s(z_{i,m-s} - \mu_i) + \varepsilon_{i,m},$$
$$\varepsilon_{i,m} \sim N(0, \Sigma_\varepsilon), \tag{9}$$

where μ_i denotes the mean of $z_{i,t}$. As Kuzin, Marcellino, and Schumacher (2011) documented, the means μ_i are often quite difficult to estimate. Therefore, we work with demeaned GDP and monthly indicators in the estimation procedure, adding the mean back afterwards in order to arrive at the final indicator-based forecast. As in the dynamic factor model, the Kalman filter and smoother fills in any missing monthly indicator observations with optimal predictions, and estimates unobserved monthly real GDP subject to a temporal aggregation constraint for the quarterly observation. The state space setup of the MFVAR is outlined in Appendix A.2.1. We estimate the model using the expectation–maximization algorithm, as detailed by Mariano and Murasawa (2010). As in the case of the QVAR model, we form the final GDP forecast as a weighted average of the individual forecasts derived from the n bivariate MFVAR models. Regarding the number of lags p, it is held fixed for practical reasons ($p = 1$ for the euro area, Germany, Spain and the Netherlands and $p = 2$ for France and Italy).[7]

2.3.3. Mixed data sampling regression model (MIDAS)

The mixed-data sampling model (MIDAS) is a single horizon-specific equation that relates the quarterly GDP to (various lags of) a monthly indicator (Ghysels, Sinko, & Valkanov, 2007; Schumacher, 2014). It generates the GDP forecast in a direct way. The MIDAS model circumvents the ragged edge problem by including as regressors a fixed (fairly large) number of the most recent lagged values of the indicator. In applied work, the MIDAS model economizes on the number of parameters requiring estimation by adopting a parsimoniously parameterized lag polynomial. The efficiency gains of such an approach come at the cost of potential efficiency losses if the implied restrictions on the lag structure between the monthly indicator and quarterly real GDP happen to be invalid. We follow Kuzin et al. (2011) in working with the exponential Almon lag polynomial. Our version of the indicator-specific MIDAS model for forecasting horizon h is defined by the following equations:

$$y^Q_{t+h} = \beta_0 + \beta_1 B(L_M; \theta) x^{(3)}_{i,m} + \varepsilon^Q_{i,t+h} \tag{10}$$

$$B(L_M; \theta) = \sum_{k=0}^{K} c(k, \theta) L_M^k \tag{11}$$

$$c(k, \theta) = \frac{\exp(\theta_1 k + \theta_2 k^2)}{\sum_{k=0}^{K} \exp(\theta_1 k + \theta_2 k^2)}, \tag{12}$$

where L_M is the monthly lag operator and the T observations of the regressor $x^{(3)}_{i,m}$ are skip-sampled from $x_{i,m}$ by including every third observation, starting from the final one.

5 The state space setup of our dynamic factor model is outlined in Appendix A.2.2. See Bańbura and Rünstler (2011) for a more detailed description of the dynamic factor model and the estimation procedure, and also Stock and Watson (2011). See Durbin and Koopman (2012) for a comprehensive discussion of state space models and the use of the Kalman filter and smoother.

6 Applying a different weighting scheme leads to results that are virtually the same; see Table A.9 in Appendix A.3. Alternatively, one could choose the numbers of factors r and q on the basis of in-sample criteria, as described by Bai and Ng (2002, 2007). Our experience, like that of Bańbura and Rünstler (2011), is that these criteria tend to indicate relatively large numbers of factors, leading to volatile and less accurate forecasts.

7 We had difficulty in achieving convergence when estimating MFVAR models with four lags. Moreover, we found that varying the number of lags according to the SIC does not seem to be efficient in terms of forecasting performances, compared to procedures that simply hold the number of lags constant over time. We therefore decided to restrict the maximum number of lags to one, two or three, and to fix it based on the out-of-sample performances of the nowcasts and backcasts in the first quarter of the sample (1996.I–1999.IV). The results are available from the authors upon request.

Thus, $x_{i,m}^{(3)} = x_{i,m}, m = 3 + w, \ldots, 3(T-2) + w, 3(T-1) + w, 3T + w$. Eq. (11) describes a weighting function of lagged values, while Eq. (12) specifies the weight for lag k as a function of k and the two parameters governing the exponential Almon lag polynomial. K is fixed at 11. The model's parameters $(\theta_1, \theta_2, \beta_0, \beta_1)$ are estimated by non-linear least squares, subject to $\theta_1 < 5$ and $\theta_2 < 0$. We compute the final GDP forecast as a weighted average of the individual forecasts derived from the n indicator-specific MIDAS models.

2.4. Factor and AR augmented models

2.4.1. Factor augmented models

We consider versions of the QVAR, MFVAR and MIDAS models in which the independent variable is a factor rather than an observed indicator. We denote these factor augmented versions by F-VAR, F-MFVAR and F-MIDAS, respectively. See Eqs. (13)–(15), which refer to the single-factor case:

$$z_t^Q = \alpha + \sum_{s=1}^{p} A_s z_{t-s}^Q + \varepsilon_t^Q, \quad \varepsilon_t^Q \sim N(0, \Sigma_{\varepsilon Q}) \tag{13}$$

$$z_m - \mu_z = \sum_{s=1}^{p} A_s (z_{m-s} - \mu_z) + \varepsilon_m, \quad \varepsilon_m \sim N(0, \Sigma_\varepsilon) \tag{14}$$

$$y_{t+h}^Q = \beta_0 + \beta_1 B(L_M; \theta) \hat{f}_m^{(3)} + \varepsilon_{t+h}^Q, \tag{15}$$

where $z_t^Q = (y_t^Q, f_t^Q)'$ and $z_m = (y_m, f_m)'$. The number of lags in the F-MFVAR model is the same as in the corresponding MFVAR model. We obtain the factors as follows. For F-VAR, we derive the factor by applying simple principal component analysis to quarterly averages of monthly data. For the mixed-frequency models, we use the Kalman-filtered factors generated by the dynamic factor model, averaged over all possible parameterizations. This facilitates comparisons between the DFM and factor augmented models.

An important specification issue is the number of factors used to summarize the information set. The literature typically restricts the analysis to a single factor (e.g., Marcellino & Schumacher, 2010). We have investigated this issue in three different ways. First, when treating the factors as separate indicators, we found that taking a weighted average of factor-specific forecasts typically weakens the forecasting accuracy, especially for nowcasts and backcasts (see Table A.7 in Appendix A.3). The only exceptions are the MIDAS models for Spain, for which the second and third factors possess substantial predictive power. Second, the marginal forecasting power of additional factors is generally very small (often zero), again with the exception of the MIDAS models for Spain (see Table A.8 in Appendix A.3). Third, when introducing more than one factor into MIDAS equations (multi-factor MIDAS), we found that multi-factor models tend to produce higher RSMFEs for (true) forecasts and nowcasts; the main exception is the two-factor model for Spain. For backcasts, the picture is mixed. For some countries we find losses of accuracy. Moreover, the empirical relationship between the number of factors and the forecasting performance is erratic and difficult to interpret. The marginal forecasting power of an additional factor can be positive or negative, and may be sensitive to the exact dating of the backcast. For example, second-month backcasts may improve, while first-month backcasts deteriorate. These findings suggest that over-fitting issues are a real concern for multi-factor MIDAS models, even when the number of factors is small. We report the estimation results for the multi-factor MIDAS models in the online appendix (see Appendix B). Taken together, the results discussed above hint at a weakness of factor augmented models, namely their limited ability to incorporate extra information into forecasts. We return to this issue in Section 4.3. Based on this preliminary analysis, we follow other authors by using one factor in each of our factor augmented models, except for the MIDAS models for Spain. In the latter case, we use three factors, which we treat as single indicators.[8]

2.4.2. AR augmented models

Finally, we consider versions of the BEQ, MIDAS and F-MIDAS models that feature an AR(1) term, as GDP's own past may contain important information. We denote these models as BEQ-AR, MIDAS-AR and F-MIDAS-AR, respectively. The BEQ-AR model for x_i can be written as

$$y_t^Q = \alpha + \varphi y_{t-1}^Q + \sum_{s=0}^{p_i} \beta_s x_{i,t-s}^Q + \varepsilon_{i,t}^Q. \tag{16}$$

As was proposed by Clements and Galvão (2008), the AR term is introduced as a common factor in the MIDAS-AR and F-MIDAS-AR models:

$$y_{t+h}^Q = \beta_0 + \varphi y_{t-1}^Q + \beta_1 B(L_M; \theta)(1 - \varphi L_M^h) x_{i,m}^{(3)} + \varepsilon_{t+h}^Q \tag{17}$$

$$y_{t+h}^Q = \beta_0 + \varphi y_{t-1}^Q + \beta_1 B(L_M; \theta)(1 - \varphi L_M^h) f_m^{(3)} + \varepsilon_{t+h}^Q. \tag{18}$$

The parameter φ is estimated simultaneously with the other parameters.

3. Data, forecast design and specification issues

This section describes the dataset, the pseudo real-time setup, the weighting scheme that we used for pooling indicator-specific forecasts in the cases of the QVAR, BEQ, BEQ-AR, MFVAR, MIDAS and MIDAS-AR models, and the selection of the numbers of lags and factors in the models.

3.1. Dataset

Our monthly dataset consists of 72 monthly time series variables (using harmonized definitions across the countries), which cover the broad range of information that

[8] The predictive power of the second and third factors in the case of Spain is a persistent feature throughout the sample period, which agents would quickly discover. Thus, the risk of this modeling assumption introducing a hindsight bias is quite small.

is readily available to economic agents. To facilitate cross-country comparisons, we have selected monthly indicators that are available for all countries for a sufficiently long period of time.[9] The indicator variables fall into four categories. The first category is hard, quantitative information on production and expenditures, such as industrial production in various sectors and countries, car sales, world trade and unemployment. The second category refers to input and output prices, such as consumer and producer prices, and oil and commodity prices. The third category contains financial variables, both quantities (money stock and credit volume) and prices (interest rates, stock prices and exchange rates). These determine financing conditions for firms and consumers. Moreover, financial market prices partly reflect financial market expectations on output developments in the near future. The fourth category is soft, qualitative information on expectations derived from surveys among consumers, retailers and firms. Moreover, we also included three composite leading indicators compiled by the OECD.

Appendix A.1 provides details on the sources, availability and applied transformations of the data series. The available monthly data are usually already adjusted for seasonality (and calendar effects). Where necessary, raw data series are seasonally adjusted using the US Census X12-method. All monthly series are made stationary by differencing or log-differencing (in the case of trending data, such as industrial production, retail sales and monetary aggregates). All variables are standardized by subtracting the mean and dividing by the standard deviation. This normalization is necessary to avoid the overweighting of series with large variances in the determination of common factors. The data transformations are the same for all of the statistical models, except for the Bayesian VAR.

3.2. Pseudo real-time design

The forecast design aims to replicate the availability of the data at the time when the forecasts are made in order to mimic the real-time flow of information as closely as possible. To this end, we used a data set downloaded on January 16, 2012, and combined this with the typical data release calendar in order to reconstruct the available dataset on the 16th of each month over the period July 1995–January 2012. All of the monthly indicator series start in January 1985, while the quarterly GDP series start in 1985.I. Thus, we employ a pseudo real-time design, which takes data publication delays into account, but ignores the possibility of data revisions for GDP and some indicators, such as industrial production. The latter implies that we might overestimate the forecasting accuracy of statistical models. However, it is likely that the effects of data revisions on the final forecast will largely cancel out (although the crisis episode may have been atypical in this regard), since statistical methods typically attempt to eliminate noise from the process by either extracting factors from a large data set or pooling large numbers of indicator-based forecasts. For example, using real-time data vintages for Germany, Schumacher and Breitung (2008) did not find any clear impact of data revisions on the forecast errors of factor models. Moreover, the effect on the *relative* performances of models, which is the main focus of this paper, is likely to be quite small (see Bernanke & Boivin, 2003). However, abstracting from data revisions may affect the comparison of mechanical forecasts with forecasts by professional analysts to a greater extent. Analysts' expectations necessarily reflect the inaccurate initial estimates of GDP's recent past, which puts them at a disadvantage vis-à-vis mechanical models in a pseudo-real time setting, as the latter can take data revisions on board.

Table 1
Timing of forecast exercise for third quarter.

No.	Forecast type	Month	Forecast made in middle of
1	Two-quarter-ahead	1	January
2		2	February
3		3	March
4	One-quarter-ahead	1	April
5		2	May
6		3	June
7	Nowcast	1	July
8		2	August
9		3	September
10	Backcast	1	October
11		2	November

We estimate the parameters of all models recursively, using only the information that was available at the time of the forecast. For similar approaches, see Giannone et al. (2008), Kuzin et al. (2011) and Rünstler et al. (2009), among others. We construct a sequence of eleven forecasts for GDP growth in a given quarter, obtained in consecutive months. Table 1 explains the timing of the forecasting exercise, taking the forecast for the third quarter of 2011 as an example. We make the first forecast in January 2011, which is called a two-quarter-ahead forecast in month one. Subsequently, we produce monthly forecasts for the next ten months through November. The last forecast is made just before the first release of GDP in mid-November. Following the conventional terminology, *forecasts* refer to one- or two-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 2011.III, we make two-quarter-ahead forecasts from January to March, one-quarter-ahead forecasts from April to June, nowcasts from July to September, and backcasts in October and November.

3.3. Weighting scheme for indicator-based forecasts

The models BEQ, BEQ-AR, QVAR, MFVAR, MIDAS and MIDAS-AR construct large numbers of different indicator-specific forecasts in the first stage, which have to be aggregated in the second stage in order to obtain the final

[9] As a consequence, no country-specific indicators, such as the Ifo-indicator for Germany, were used for forecasting. Moreover, the Purchasing Managers Index (PMI) time series is not long enough for all countries. Of course, in practice, economic agents will use this country-specific information, meaning that our results might underestimate the forecasting accuracies of mechanical models for certain countries or periods.

forecast. Taking a weighted average of a large number of forecasts may ameliorate the effects of misspecification bias, parameter instability and measurement errors in the data, which may afflict the individual forecasts (Timmermann, 2006). We have investigated three different weighting schemes: (i) equal weights (simple mean); (ii) weights that are inversely proportional to the root mean squared forecast error (RMSFE), measured from the start of the sample period until the previous quarter (recursive RMSFE scheme); and (iii) weights that are inversely proportional to the RMSFE measured over the past four quarters (moving window RMSFE scheme). Equal weights have been proven to work quite well as a pooling mechanism (e.g., Clark & McCracken, 2010 and Stock & Watson, 2004), while the latter two methods assign weights to the indicators based on their (recent) past forecasting performances.

Table A.9 in the Appendix A.3 provides an overview of the RMSFEs of the three weighting schemes by horizon and country for BEQ-AR, QVAR, MFVAR and MIDAS-AR. The overall picture is that the moving window RMSFE weighting scheme, which emphasizes performances in the recent past, has the smallest RMSFE on average. However, as the differences between the schemes are very small, our results are not sensitive to the specific weighting method. In the remainder of the paper, we apply the moving window RMSFE weighting scheme to all relevant models and all countries.

4. Empirical results for statistical models

4.1. Forecasting performance

Table 2 presents data on the forecast performances of the statistical models for our five countries and the euro area for the complete sample period 1996.I–2011.III (63 quarters). The underlying empirical analysis has been carried out on a monthly basis for eleven horizons. To save space, Tables 2–5 and Tables A.7–A.13 in Appendix A.3 report results for the two- and one-quarter-ahead forecasts, the nowcast and the backcast, which have been calculated as the averages of the corresponding monthly data. Moreover, Table 2 only reports the AR versions of the BEQ, MIDAS and F-MIDAS models.[10] We measure forecast performances using the root mean square forecast error (RMSFE). We report the results of two benchmark models that have been used in the literature: the random walk (RW) with drift and a pure univariate autoregressive (AR) model.[11] A comparison with the RW benchmark reveals the advantage of using information for forecasting, including GDP's own past. A comparison with the AR benchmark focuses on the value added by monthly auxiliary information. The first column of Table 2 reports the RMSFE of the random walk. For the other statistical models, including the AR model, the entries refer to their RMSFEs relative to that of the RW benchmark, in order to increase the comparability of the results across countries and horizons. Shaded entries indicate the model with the lowest RMSFE in a row (for a particular horizon). Entries in bold indicate models that have RMSFEs that are less than 10% larger than that of the best model, and are also smaller than the RMSFE of the benchmark model.[12] The 10% threshold is meant as a rough indication of the economic significance of differences in forecasting ability. We will call models that meet this condition "competitive models", as they do not differ "too much" from the best model in terms of forecasting performances.[13]

The outcomes in Table 2 point to several interesting results. First, the incorporation of monthly information in statistical forecasting procedures pays off in terms of forecasting accuracy, particularly for nowcasts and backcasts. The large majority of the relative RMSFEs are smaller than one and lower than those of the AR model. They also tend to decline as the horizon shortens and more monthly information has been absorbed. Moreover, models that do not exploit the available monthly observations fully (QVAR and F-VAR) generally have larger RMSFEs than models that do. Second, when forecasting one and two quarters ahead, the gain is rather limited for many models. For the two-quarter-ahead forecast, the best models have RMSFEs that are only 5% lower than the benchmark on average. Only for Spain does the best statistical model deliver an economically significant improvement. For the one-quarter-ahead forecasts, the average improvement by the best models relative to the RW benchmark is 15%, but the other models generally post gains of less than 10%. For the nowcasts and backcasts, the average accuracy gains for the best performing models amount to roughly 30% and 40%, respectively. This pattern suggests that statistical models are of greater value when they can use information that pertains to the relevant quarter. Their relative strength is in improving the assessment of the current state of the economy. Third, the dynamic factor model displays the best performance overall. Looking across countries and horizons, it works best for backcasts and is at least competitive in all other cases. The factor augmented MIDAS-AR and MFVAR models perform relatively well for one-quarter-ahead predictions and nowcasts, but not so well for backcasts. The Bayesian VAR model works best for Spain and relatively well for the Netherlands. It also delivers the best two-quarter-ahead forecasts. Fourth, many models are competitive at the two-quarter-ahead horizon in most of the countries, but the number of competitive models falls quickly as the horizon shortens. For four countries, there are no competitive models left for the backcast. This result is another sign that the predictions

[10] The results for the non-AR versions are very similar to those of their AR counterparts. Section 4.4 discusses the effect of taking an AR term on board. The empirical results for monthly horizons for all of the models analyzed can be found in additional tables that are available in the online appendix (see Appendix B).

[11] The drift parameter is estimated recursively. With regard to the AR model, the number of lags is determined recursively by the SIC, with a maximum of four.

[12] If the best model has an RMSFE of 0.6, the cut-off point is an RMSFE of 0.66.

[13] In addition, we also performed Diebold and Mariano (1995) tests that paint broadly the same picture. The results of these tests are reported in the online appendix (see Appendix B).

Table 2
Forecasting performances of the statistical models (RMSFE), 1996.I–2011.III.

Frequency model	Benchmark		BVAR	Pooling forecasts				Pooling information			
	RW	AR		BEQ-AR	QVAR	MIDAS-AR	MFVAR	DFM	F-VAR	F-MIDAS-AR	F-MFVAR
Euro area											
2Q ahead	*0.64*	1.00	**0.97**	0.97	1.01	**0.97**	**0.97**	**0.99**	**1.00**	1.01	1.01
1Q ahead	*0.63*	1.00	0.93	**0.90**	1.00	0.93	0.93	**0.90**	0.96	0.83	**0.86**
nowcast	*0.63*	0.95	0.83	0.79	0.92	0.79	0.82	**0.69**	0.88	**0.67**	0.66
backcast	*0.63*	0.85	0.71	0.69	0.79	0.62	0.70	0.51	0.79	0.62	0.66
Germany											
2Q ahead	*0.91*	1.05	1.02	0.98	1.00	1.02	**0.99**	**0.99**	1.00	1.01	1.01
1Q ahead	*0.91*	1.03	**0.99**	**0.97**	**1.00**	**0.97**	**0.98**	0.92	**0.98**	**0.94**	**0.92**
nowcast	*0.91*	1.03	0.94	0.90	0.99	0.92	0.90	0.77	0.94	**0.80**	**0.77**
backcast	*0.91*	1.03	0.88	0.85	0.97	0.83	0.79	0.67	0.91	0.74	0.80
France											
2Q ahead	*0.53*	1.02	0.94	**0.96**	1.01	**0.96**	**0.97**	**0.94**	**1.00**	**1.00**	**0.97**
1Q ahead	*0.53*	0.99	**0.90**	**0.90**	1.01	**0.91**	**0.91**	0.84	0.98	**0.85**	**0.85**
nowcast	*0.52*	0.90	0.82	0.78	0.89	0.79	0.80	**0.64**	0.89	0.62	**0.63**
backcast	*0.52*	0.84	0.74	0.73	0.79	0.71	0.72	0.52	0.80	0.62	0.63
Italy											
2Q ahead	*0.75*	1.04	0.95	**0.99**	1.01	**0.98**	**0.99**	**0.99**	**0.99**	1.06	**0.99**
1Q ahead	*0.75*	0.98	**0.94**	**0.95**	0.99	**0.95**	**0.95**	**0.92**	**0.97**	**0.90**	0.89
nowcast	*0.74*	0.95	0.89	0.86	0.94	0.87	0.87	**0.74**	0.90	**0.72**	0.72
backcast	*0.74*	0.92	0.81	0.80	0.85	0.80	0.78	0.64	0.80	**0.70**	**0.67**
Spain											
2Q ahead	*0.64*	**0.92**	**0.87**	**0.91**	**0.92**	**0.9**	**0.93**	**0.88**	**0.90**	0.86	**0.93**
1Q ahead	*0.63*	0.83	**0.77**	0.85	0.85	**0.78**	0.89	0.75	**0.81**	0.85	0.84
nowcast	*0.63*	0.74	0.62	0.78	0.76	**0.68**	0.83	**0.64**	0.75	0.71	0.82
backcast	*0.62*	0.79	0.49	0.79	0.72	0.75	0.84	**0.57**	0.75	0.66	0.85
Netherlands											
2Q ahead	*0.72*	**0.99**	0.92	**0.96**	**0.99**	**0.97**	**0.98**	**0.99**	**0.96**	**0.95**	1.03
1Q ahead	*0.71*	0.99	**0.89**	**0.90**	0.98	0.92	0.96	**0.90**	0.94	0.83	**0.89**
nowcast	*0.71*	0.96	0.83	0.82	0.93	0.85	0.88	**0.76**	0.89	0.74	**0.78**
backcast	*0.71*	0.90	0.75	0.77	0.84	0.76	0.84	0.68	0.81	**0.72**	0.79

Notes: RW: random walk; AR: autoregressive model; BEQ-AR: bridge equation with an AR term; QVAR: quarterly vector autoregressive model; BVAR: Bayesian QVAR model; F-VAR: factor augmented QVAR model; DFM: dynamic factor model; MFVAR: mixed frequency vector autoregressive model; F-MFVAR: factor augmented MFVAR model; MIDAS-AR: mixed data sampling model with AR-term; F-MIDAS-AR: factor augmented MIDAS-AR.
For RW (in italics), the entries refer to the RMSFE; for all other models, they refer to the RMSFE relative to the RW model's RMSFE. Grey cells indicate the models with the lowest RMSFEs. Figures in boldface indicate models whose RMSFEs are no more than 10% larger than the RMSFE of the best model.

from statistical models incorporate little information at the two-quarter-ahead horizon. Models that exploit all of the available monthly information fully, including the traditional bridge model, generally stay competitive up to the first quarter ahead horizon. Fifth, Spain is an exceptional case within our sample of countries, as most of the statistical models perform poorly relative to the AR model, which happens to forecast pretty well. Thus, most of the models appear to have difficulty in capitalizing on their comparative advantage, the monthly information set. In contrast, the Bayesian VAR and the dynamic factor model both perform strongly in the Spanish case.

4.2. The marginal value of statistical models

Ranking models according to their RMSFEs gives an initial indication of the relative usefulness of each. This subsection focuses on the marginal value of the various models by investigating whether the forecasts generated by different models differ in their information content. As the various statistical approaches follow different strategies for extracting monthly information, it is conceivable that some models may be complementary. In that case, taking a weighted average of their respective forecasts may improve the forecast accuracy. Even a model that performs badly may have a positive marginal value if it is able to pick up specific useful information. We establish the marginal value of the models relative to the best statistical model (lowest RMSFE) by running an encompassing test (e.g., Rünstler et al., 2009 and Stekler, 1991). The test regression is:

$$y^Q_{t+h|t} = \lambda \hat{y}^Q_{\mathrm{a}(t+h|t)} + (1-\lambda)\hat{y}^Q_{\mathrm{b}(t+h|t)} + \varepsilon_t, \tag{19}$$

where y^Q_t is GDP growth in t, $\hat{y}^Q_{\mathrm{a}(t+h|t)}$ and $\hat{y}^Q_{\mathrm{b}(t+h|t)}$ are the forecasts for quarter $t+h$ at time t from the alternative and

Table 3
Marginal value of the statistical models (evaluation period 1996.I–2011.III).

Frequency model	AR	BVAR	Pooling forecasts				Pooling information			
			BEQ-AR	QVAR	MIDAS-AR	MFVAR	DFM	F-VAR	F-MIDAS-AR	F-MFVAR
Euro area										
2Q ahead	–	–		–	–	–	**0.99**	–	–	**0.98**
1Q ahead	–	–	–	–	–	–	–	–		–
nowcast	–	**0.97**	**0.99**	–	**0.99**	–	–	–	–	
backcast	–	–	–	–	–	–		–	–	–
Germany										
2Q ahead	–	–		–	–	–	–	–	–	**0.99**
1Q ahead	–	–	–	–	–	–		–	–	0.99
nowcast	–	–	–	–	–	–		–	–	**0.99**
backcast	–	–	–	–	–	–		–	–	–
France										
2Q ahead	–		–	–	–	–	**0.98**	–	–	**0.99**
1Q ahead	–	**0.99**	–	–	–	–		–	**0.99**	**0.99**
nowcast	–	**0.99**	–	–	–	–	**0.99**	–		0.99
backcast	–	**0.99**	–	–	–	–		–	–	–
Italy										
2Q ahead	–		–	–	–	–	–	–	–	–
1Q ahead	**0.99**	**0.98**	–	**0.99**	**0.98**	–	–	**0.99**	**0.98**	
nowcast	–	–	–	–	–	–	–	–	–	
backcast	–	–	–	–	–	–		–	–	0.99
Spain										
2Q ahead	**0.98**	**0.98**	–	0.99	**0.99**	–	–	**0.99**		–
1Q ahead	**0.95**	**0.96**	–	**0.99**	**0.97**	–		**0.99**	–	–
nowcast	–		–	–	–	–	**0.91**	–	**0.93**	**0.96**
backcast	–		–	–	–	–	**0.91**	–	**0.97**	0.99
Netherlands										
2Q ahead	–		–	–	–	–	–	–	–	–
1Q ahead	–	**0.99**	–	–	–	–	–	–		–
nowcast	–	**0.99**	**0.99**	–	–	–	**0.99**	–		**0.99**
backcast	–	**0.97**	–	–	–	–		–	–	–

Notes: AR: autoregressive model; BEQ-AR: bridge equation with an AR term; QVAR: quarterly vector autoregressive model; BVAR: Bayesian QVAR model; F-VAR: factor augmented QVAR model; DFM: dynamic factor model; MFVAR: mixed frequency vector autoregressive model; F-MFVAR: factor augmented MFVAR model; MIDAS-AR: mixed data sampling model with AR-term; F-MIDAS-AR: factor augmented MIDAS-AR.
Grey cells indicate the model with the lowest RMSFE. Figures in boldface indicate that the encompassing test is statistically significant at the 5% level.

best models respectively; λ is the weight of the alternative model; and $(1 - \lambda)$ is the weight of the best model. In order to obtain interpretable results, we impose the restriction that λ must lie between 0 and 1. The alternative model contains additional information compared to the best model if $\lambda > 0$. We estimate λ and its standard error on the interval [0, 1] by Maximum Likelihood (ML), and perform a one-sided (asymptotically valid) test of the hypothesis $\lambda = 0$ at the 5% significance level. All of the calculations refer to the complete sample period 1996.I–2011.III (63 quarters).

Table 3 reports the results of our encompassing test. The entries are the RMSFEs of the forecast combination relative to those of the best model, as a measure of the potential gains from using forecast combinations. The estimated weight λ itself is not reported; entries in boldface signify λ estimates that are statistically greater than zero. A blank entry means that the ML algorithm returned the corner solution $\lambda = 0$.

The main message of Table 3 is that, in economic terms, the gains from combining forecasts from different statistical models are very limited for all countries except for Spain. In the majority of cases, the accuracy gain is zero. Although the gains tend to increase as the horizon shortens and the models have absorbed more information, they typically do not exceed 3%, even for backcasts. Moreover, no model emerges as a clear winner, although the models with the lowest RMSFEs (DFM, F-MIDAS-AR and BVAR) also tend to be the most promising in terms of marginal value. Thus, it appears that the various approaches do not differ greatly with respect to the types of information that they extract from large-scale monthly datasets. Finally, Table 3 shows that statistical significance and economic importance are different concepts. Most of the non-zero entries reflect a significant test result for the encompassing test, while most of the gains in forecast accuracy are very small.

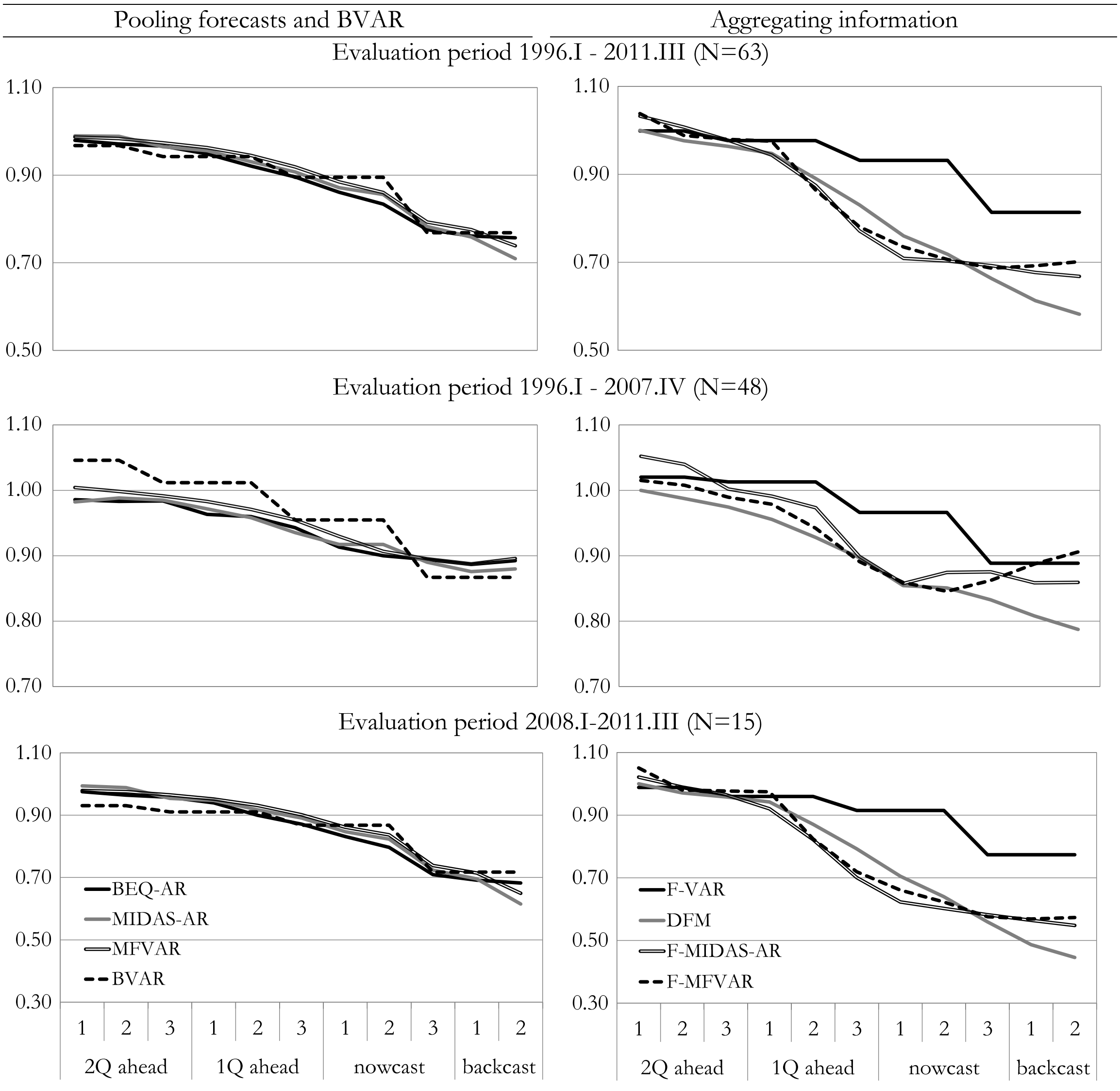

Fig. 1. Learning curves of statistical models. Notes: RMSFE first 2Q ahead forecast of DFM = 1; all lines excluding Spain.

4.3. Splitting the sample: the Great Moderation versus the financial crisis

Our sample includes the financial crisis, when real GDP went through a particularly volatile phase across the industrialized countries. An obvious question is whether and to what extent the performances of statistical forecasting models differ between the financial crisis period and the period before the financial crisis, which was characterized by a large degree of macroeconomic stability. The latter period has been labeled the Great Moderation. Most of the existing literature on short term forecasting is based on data from the Great Moderation period. Of course, forecasting in volatile times poses greater challenges, meaning that the results of a comparative analysis will be more informative on the issue of which models are most apt at absorbing monthly information. Moreover, good forecasts and nowcasts are of greater importance to economic agents and policy makers in a volatile environment.

We divide the sample period into two parts: 1996.I–2007.IV (Great Moderation) and 2008.I–2011.III (financial crisis). We discuss the models' performances on the basis of their learning curve, which shows the relative decline in the RMSFE as the forecasting horizon shortens, averaged over four countries and the euro area.[14] We calculate a model's learning curve as its RMSFE standardized by the RMSFE for the first month of the two-quarter-ahead DFM-forecast. Fig. 1 shows the learning curves of selected

[14] We leave out Spain because virtually all statistical models fail to beat the AR benchmark for the period 1996.I–2007.IV. Moreover, this avoids any possible hindsight bias due to the fact that the F-MIDAS-AR model for Spain employs a different number of factors. Country details can be found in Tables A.10–A.13 in Appendix A.3, which are the counterparts of Tables 2 and 3 for both subperiods.

models for the complete sample period and the two subperiods (in the rows). The graphs on the left refer to the three models that aggregate indicator-specific forecasts and the Bayesian VAR, while the graphs on the right refer to the models that rely on factor analysis to summarize the indicators.

For the complete sample period, we find that the dynamic factor model displays the steepest learning curve. Its RMSFE falls by an average of 42% within 11 months. In addition, models involving factor analysis have steeper learning curves up to the nowcast, on average, than models that aggregate indicator-specific forecasts. Moreover, the learning curve is rather flat until month four for all models, reflecting the fact that the scope for predicting GDP for horizons beyond one quarter in the future is very limited. This is a stable pattern that holds during both the Great Moderation episode and the crisis episode (and also across countries). The Bayesian VAR is the fastest learning VAR model. Despite the fact that it does not use all available monthly observations, it holds up well against many of the models that do use all of the monthly information. This suggests that the model's implicit aggregation by Bayesian shrinkage on coefficients, when applied in a mixed-frequency setting, may turn out to be a viable alternative approach to the information aggregation strategy of factor-based models, especially for backcasts in stable periods. For the US, Schorfheide and Song (2013) find that using within-quarter monthly information leads to drastic improvements in short-horizon forecasting performances.

Predicting GDP is much more difficult in the crisis period (see Tables A.10–A.13 in Appendix A.3). The RMSFE of the benchmark model is two to three times as large during the crisis period as during the Great Moderation. However, part of this deterioration can be offset, as the scope for improving forecasts through the utilization of monthly information appears to be larger in volatile times, particularly for nowcasting and backcasting. For example, the RMSFE of the dynamic factor model falls by 21% on average over the course of 11 months in the period before the crisis, as compared to 55% in the crisis period. The differences in forecast accuracy across models are considerably larger after the crisis than before the crisis. This also means that the number of competitive models during the Great Moderation is much larger than after the financial crisis, especially for the nowcasting and backcasting horizons. This finding is consistent with the results of D'Agostino and Giannone (2012), who show that the gain from using factor models is substantial, especially in periods of high comovement, as was the case during the financial crisis. The crisis episode poses a more demanding test to models, and consequently, fewer models manage to pass. This finding also implies that the cost of employing a suboptimal model increased after the crisis. Finally, the potential gains of combining statistical models (marginal value) tend to be markedly smaller during the financial crisis than in the preceding period.[15]

Looking at the models that rely on factor analysis, a remarkable result is the strikingly different shapes of the learning curves of factor augmented mixed-frequency models on the one hand, and the dynamic factor model on the other hand. F-MIDAS-AR and F-MFVAR learn faster than the dynamic factor model between months four and seven, but the pace of improvement quickly levels off beyond that point. As a result, their one-quarter-ahead forecasts and early nowcasts are more accurate than their DFM counterparts on average. This good performance is attributable entirely to the financial crisis episode. During the Great Moderation, their forecasts do not show any improvement at all after month seven. In contrast, the DFM improves its forecast steadily as more and more information is absorbed, producing superior backcasts and late nowcasts in both quiet and volatile times. Moreover, the DFM delivered better or equally good forecasts over the whole horizon during the Great Moderation period.

This pattern reflects the comparative strengths and weaknesses of the two model approaches, which may play out differently at different horizons and in different circumstances. To gain additional insights into why the DFM works, Fig. 2 shows its learning curve for different numbers of lags (p) and (static) factors (r). The learning curves indicate that the DFM's performance for nowcasts and backcasts is linked strongly to the number of factors. However, additional factors do not offer any benefits for predictions beyond the current quarter; then, one factor is sufficient. When the DFM is restricted to only one factor, its learning curve has the same shape as that of the F-MIDAS-AR model. Moreover, the number of lags in the factor VAR process is only a minor determinant of the forecast quality at any horizon. As a consequence, the comparative strength of the DFM is its ability to include more information in the forecasting procedure. In contrast, this aspect is the weak point of factor augmented models, which can exploit only one factor in practice (see Section 2.4.1). This hurts their performances for late nowcasts and backcasts. On the other hand, factor augmented models (and MIDAS models in particular) may exploit their richer dynamic specification of the relationship between an indicator and GDP. This may give them an edge over the DFM, particularly for one-quarter-ahead predictions and early nowcasts, for which the single-factor restriction is not a disadvantage. However, a flexible dynamic specification is an asset for forecasting only if it is feasible to identify stable dynamic relationships reliably. Otherwise, these models may estimate spurious dynamic relationships in-sample, which may actually reduce the accuracy of the out-of-sample predictions. The evidence in Fig. 1 highlights this identification problem. During the Great Moderation, factor augmented models were unable to capitalize on their comparative advantage; however, that changed dramatically after the crisis hit. In volatile times, when it is easier to identify dynamic relationships, factor augmented models may deliver the best forecasts for specific horizons: (late) one-quarter-ahead forecasts and early nowcasts. Dynamic flexibility tends to increase RMSFEs for two-quarter-ahead forecasts in all environments, due to the very limited scope for forecasting that far into the future. Thus, our findings suggest that the question of whether or not to use factor augmented models in practice for one-quarter-ahead forecasts and early nowcasts should depend on the researcher's or practitioner's confidence in the models' abilities to uncover useful dynamic relationships.

15 Moreover, the encompassing test is significant in only a few cases, but this can be attributed partly to the small number of observations.

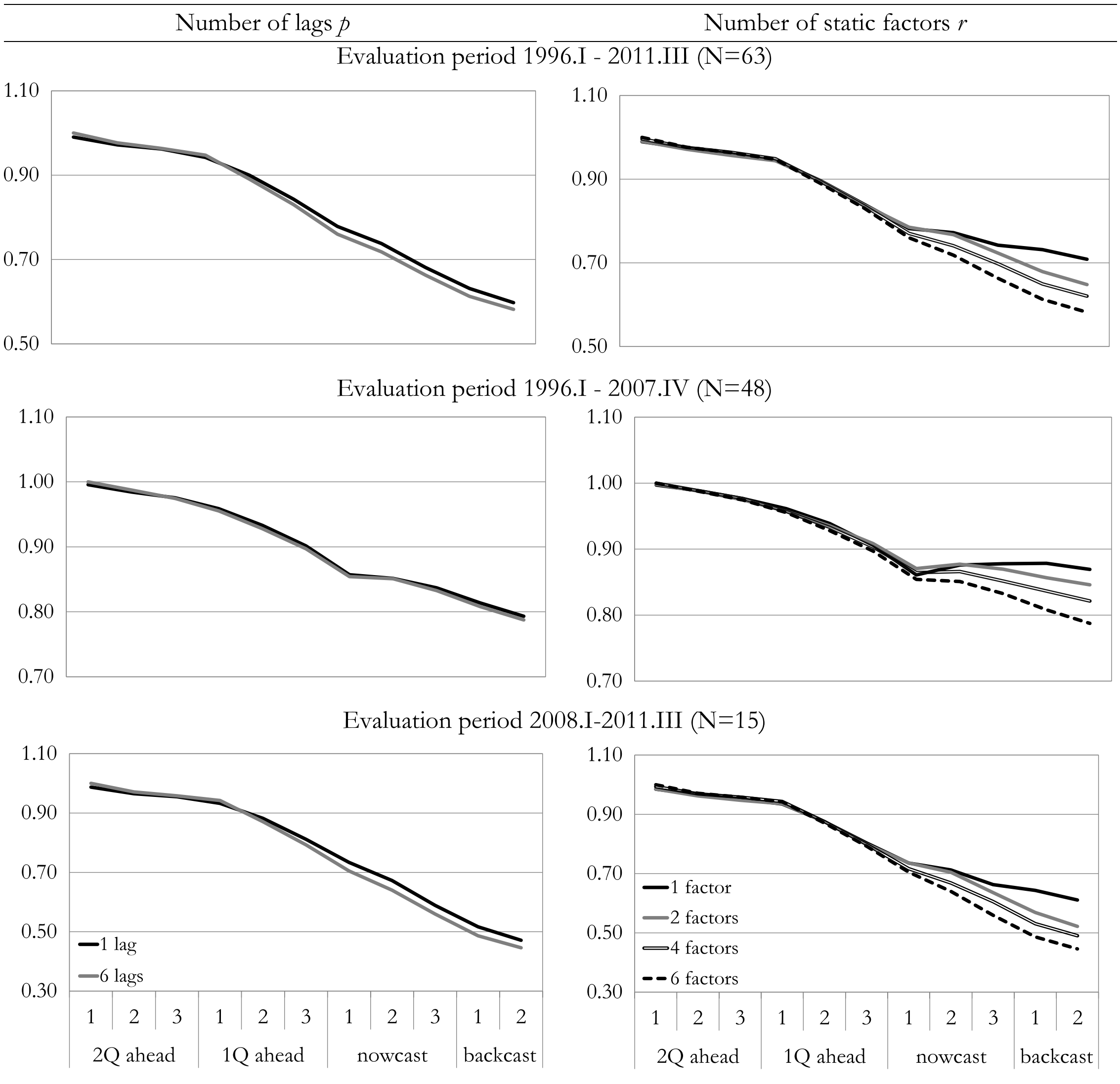

Fig. 2. Learning curve of the dynamic factor model by the numbers of lags and factors. Notes: RMSFE first 2Q ahead forecast of DFM with 6 lags ($p = 6$) or 6 static factors ($r = 6$) = 1; all lines excluding Spain.

4.4. Assessing model features

The fact that our analysis includes many models and five countries plus the euro area allows us to shed some light on the issue of which model features are most valuable for forecasting and nowcasting. We focus on the following modeling choices: (1) employing factor analysis to summarize monthly information; (2) using all available monthly information; (3) exploiting GDP's own past by adding an autoregressive term to the forecasting equation. To assess the effect of a specific model feature on the RMSFE, we compare (pairs of) models that differ only in that aspect. Moreover, we take the average over four countries and the euro area (excluding Spain again) to average out the country-specific component.

To measure the impact of utilizing factor analysis for the aggregation of monthly information, rather than aggregating indicator-specific forecasts, we can compare four pairs of models: (F-VAR, QVAR), (F-MIDAS, MIDAS), (F-MIDAS-AR, MIDAS-AR) and (F-MFVAR, MFVAR). The effect of using all available information on indicators can be measured by comparing the quarterly VAR models with their mixed-frequency counterparts: (MFVAR, QVAR) and (F-MFVAR, F-VAR). This comparison also relates to the effect of making GDP a monthly latent variable in a system. For the AR effect, we look at three pairs: (BEQ-AR, BEQ), (MIDAS-AR, MIDAS) and (F-MIDAS-AR, F-MIDAS).

Table 4 reports the impacts of the three model features (averaged over four countries and the euro area) for the complete sample period and the two subperiods. Starting with the effect of utilizing factor analysis, we find that this improves the forecasting accuracy substantially for all horizons, and for nowcasts in particular. The gains are much larger for mixed frequency models than for quarterly models. For the complete sample, the average gain is 14% for nowcasts, 8% for backcasts, and 6% for one-quarter-ahead forecasts. This suggests that summarizing

Table 4
Effects of model features on the forecast accuracy.

Evaluation period		1996.I–2011.III ($N = 63$)				1996.I–2007.IV ($N = 48$)				2008.I–2011.III ($N = 15$)			
Forecast horizon		2Q ahead	1Q ahead	Nowcast	Backcast	2Q ahead	1Q ahead	Nowcast	Backcast	2Q ahead	1Q ahead	Nowcast	Backcast
Aggregate information first													
F-VAR	QVAR	−0.02	−0.03	−0.04	−0.03	0.01	0.01	−0.02	−0.04	−0.03	−0.05	−0.04	−0.02
F-MFVAR	MFVAR	0.02	−0.07	−0.18	−0.08	0.01	−0.03	−0.07	0.01	0.03	−0.09	−0.27	−0.18
F-MIDAS	MIDAS	0.00	−0.07	−0.18	−0.13	0.02	−0.01	−0.06	−0.04	−0.01	−0.10	−0.28	−0.23
F-MIDAS-AR	MIDAS-AR	0.02	−0.07	−0.18	−0.09	0.04	0.00	−0.04	−0.02	0.01	−0.12	−0.28	−0.16
Average		**0.01**	**−0.06**	**−0.14**	**−0.08**	**0.02**	**−0.01**	**−0.05**	**−0.02**	**0.00**	**−0.09**	**−0.22**	**−0.15**
Full use of monthly information													
F-MFVAR	F-VAR	0.01	−0.09	−0.24	−0.15	−0.01	−0.06	−0.11	0.02	0.02	−0.11	−0.34	−0.30
MFVAR	QVAR	−0.02	−0.05	−0.09	−0.10	0.00	−0.01	−0.05	−0.04	−0.03	−0.07	−0.11	−0.14
Average		**−0.01**	**−0.07**	**−0.16**	**−0.12**	**−0.01**	**−0.04**	**−0.08**	**−0.01**	**0.00**	**−0.09**	**−0.23**	**−0.22**
Autoregressive term													
MIDAS-AR	MIDAS	0.01	−0.01	−0.03	−0.05	0.01	0.00	−0.02	−0.01	0.01	−0.01	−0.04	−0.08
F-MIDAS-AR	F-MIDAS	0.03	−0.01	−0.03	0.00	0.03	0.01	−0.01	0.00	0.04	−0.02	−0.05	−0.01
BEQ-AR	BEQ	0.00	0.00	−0.02	−0.04	0.00	0.00	−0.01	0.01	0.00	0.00	−0.03	−0.08
Average		**0.02**	**−0.01**	**−0.03**	**−0.03**	**0.01**	**0.00**	**−0.01**	**0.00**	**0.02**	**−0.01**	**−0.04**	**−0.06**

Notes: The effects are calculated as (RMSFE (model)—RMSFE (base model))/RMSFE (base model), averaged across the euro area, Germany, France, Italy and the Netherlands.
BEQ: bridge equation; BEQ-AR: BEQ with an AR term; QVAR: quarterly vector autoregressive model; BVAR: Bayesian QVAR model; F-VAR: factor augmented QVAR model; MFVAR: mixed frequency vector autoregressive model; F-MFVAR: factor augmented MFVAR model; MIDAS: mixed data sampling model; MIDAS-AR: MIDAS with AR-term; F-MIDAS: factor augmented MIDAS; F-MIDAS-AR: F-MIDAS with an AR term.

the information from monthly data is especially helpful when the information pertains to the quarter of interest itself. When forecasting or backcasting, the inevitable loss of information due to summarizing appears to partly offset any gains that arise from the removal of noise. Moreover, we find an interesting difference between tranquil and volatile times. Using factors produces only modest gains in tranquil times, when GDP develops rather smoothly. In such periods, there is little information available in the first place, and the information losses due to summarizing may be comparatively severe relative to the gains from the removal of noise. In volatile times, when the indicators display a larger degree of comovement, the gains are much larger: up to 22% for nowcasts and 15% for backcasts. Next, we discuss the effect of using all available monthly observations. This effect is also sizable for all horizons except for the two-quarter-ahead forecast. For the full sample, the RMSFEs of nowcasts decrease by 16% and those of backcasts by 12%. Again, there is a large difference between the pre-crisis and the crisis periods. The gains from using monthly information are realized primarily in volatile episodes, as is evidenced by the 23% and 22% gains in accuracy for nowcasts and backcasts, respectively, in the crisis period. In contrast, the gains during the Great Moderation period are (very) modest, once again suggesting that the information content of the monthly dataset is low in stable environments. Finally, exploiting GDP's own past by adding an AR term has small positive effects on the forecasting accuracy of nowcasts and backcasts, but only during the crisis episode; the nowcasts improve by 3% and the backcasts by 6%.

5. Analysis of forecasts by professional analysts

The views of professional forecasters are an alternative and convenient source of information for policy makers and market participants. There are currently several surveys on the economic outlook that are available on a regular basis. The European Central Bank undertakes a quarterly survey among professional forecasters to obtain information on inflation expectations and growth prospects for the euro area. In the US, the Federal Reserve Bank of Philadelphia runs a well-known survey. Moreover, the private sector firm Consensus Economics collects and publishes economic forecasts on a monthly basis in the publication *Consensus Forecasts*. Consensus Forecasts offers an overview of private sector analysts' expectations for a set of key macroeconomic variables for a broad range of countries. Consensus Forecasts is best known for its expectations on annual GDP growth for the current and next year. However, it also provides quarterly forecasts for GDP, which we will use in this paper.[16] The panelists supply forecasts for six consecutive quarters, starting from the first unpublished quarter. The numbers of respondents vary somewhat over time, but on average about nine institutions participate in the poll for the Netherlands, fifteen each for Italy and Spain, twenty for France, and thirty for Germany and the euro area.

This section investigates two issues. The first issue is the quality of Consensus forecasts as a separate forecasting device, relative to the best statistical model. The second issue is the marginal value of Consensus forecasts, based on an encompassing test versus the three best models (DFM, BVAR and F-MIDAS-AR). In forming their expectations, analysts include subjective assessments of (potentially) a multitude of relevant factors, alongside presumably model-based predictions. If a mixture of model-based and (subjective) Consensus forecasts improves the forecast

[16] The annual Consensus forecasts have been analyzed in several papers (e.g., Ager, Kappler, & Osterloh, 2009, Batchelor, 2001, Lahiri, Isiklar, & Loungani, 2006 and Loungani & Rodriguez, 2008). The quarterly forecasts have not been used before, except in a case study for the Netherlands by de Winter (2011).

Table 5
Comparison of the Consensus forecasts with those of the best statistical models.

Evaluation period	1996.I–2011.III ($N = 63$)					1996.I–2007.IV ($N = 48$)					2008.I–2011.III ($N = 15$)				
Indicator	rRMSE Best	Rank	Gain BVAR	Gain DFM	Gain F-MIDAS-AR	rRMSE Best	Rank	Gain BVAR	Gain DFM	Gain F-MIDAS-AR	rRMSE Best	Rank	Gain BVAR	Gain DFM	Gain F-MIDAS-AR
Euro area															
2Q ahead	0.99	1	**0.99**	**0.95**	**0.92**	1.21	10	–	–	–	0.99	1	0.99	**0.93**	**0.91**
1Q ahead	1.07	3	**0.95**	**0.95**		1.42	10	–	–	–	1.06	2	**0.94**	**0.94**	
Nowcast	1.18	4	**0.92**	**0.99**	**0.97**	1.42	10	–		**0.98**	1.17	4	**0.89**	**0.98**	**0.98**
Backcast	1.37	5	**0.89**	0.99	**0.92**	1.70	10	–		**0.94**	1.32	4	**0.84**		**0.91**
Germany															
2Q ahead	1.02	4	**0.97**	**0.99**	**0.98**	1.03	6	**0.96**	**0.98**	**0.93**	1.03	5	0.97	–	
1Q ahead	1.02	3	**0.94**	**0.98**	**0.97**	0.98	1	**0.95**	**0.96**	**0.91**	1.06	4	0.92	–	0.99
Nowcast	1.07	4	**0.87**	**0.98**	**0.95**	0.93	1	**0.90**	**0.89**	**0.84**	1.26	4	0.85	–	–
Backcast	1.01	2	**0.78**	**0.93**	**0.88**	0.77	1	**0.77**	**0.76**	**0.72**	1.40	4	0.78		0.99
France															
2Q ahead	1.05	7		**0.99**	**0.96**	1.22	10	**0.99**		**0.98**	1.06	2		**0.96**	**0.92**
1Q ahead	1.11	8	**0.99**		**0.99**	1.27	10	**0.98**		**0.98**	1.06	5	–	**0.98**	0.99
Nowcast	1.31	7	**0.96**	–	**0.98**	1.34	10	**0.95**		**0.96**	1.32	4	**0.95**	0.99	0.99
Backcast	1.29	4	**0.89**	**0.98**	**0.91**	1.34	8	**0.91**	**0.99**	**0.90**	1.23	4	**0.82**	**0.98**	**0.92**
Italy															
2Q ahead	1.13	10		–	**0.97**	1.31	10	–	–	–	1.03	2		**0.96**	**0.91**
1Q ahead	1.15	10	–	–	–	1.38	10	–	–	–	1.01	2	**0.96**	**0.95**	0.99
Nowcast	1.30	9	**0.98**	–	–	1.46	10	–	–	–	1.23	4	**0.92**	–	
Backcast	1.35	10	**0.97**		**0.98**	1.40	10	**0.98**	0.99	**0.98**	1.34	4	**0.92**		0.98
Spain															
2Q ahead	0.93	1	**0.92**	**0.89**	**0.92**	1.23	10		–	–	0.84	1	**0.82**	**0.81**	**0.84**
1Q ahead	0.98	1	**0.93**	**0.91**	**0.82**	1.31	10		–	**0.96**	0.85	1	**0.80**	**0.82**	**0.74**
Nowcast	1.03	3	**0.95**	**0.86**	**0.78**	1.44	10		**0.98**	**0.96**	0.81	1	**0.76**	**0.71**	**0.59**
Backcast	1.13	2	**0.97**	**0.80**	**0.74**	1.48	7		**0.94**	**0.93**	0.69	1	**0.69**	**0.51**	**0.43**
Netherlands															
2Q ahead	1.11	9		**0.97**	**0.98**	1.31	10	–	–		0.95	1	**0.95**	**0.85**	**0.89**
1Q ahead	1.08	4	**0.97**	**0.95**	**0.97**	1.16	10	–	**0.99**	**0.99**	1.01	2	**0.88**	**0.88**	**0.94**
Nowcast	1.17	7	**0.96**	**0.97**	**0.96**	1.25	10	0.99	**0.98**	**0.98**	1.11	4	**0.83**	**0.93**	**0.92**
Backcast	1.30	10	**0.97**	**0.97**	**0.96**	1.44	10		0.99	–	1.14	2	**0.82**	**0.91**	**0.82**

Notes: rRMSFE: RMSFE (Consensus)/RMSFE (best statistical model); rank: ranking among ten procedures (nine statistical models and Consensus forecasts); gain: RMSFE (combination of Consensus and statistical model)/RMSFE (statistical model). Figures in boldface indicate that the encompassing test is statistically significant at the 5% level. Grey cells indicate the statistical models with the lowest RMSFEs. For the euro area, both the full sample and the Great Moderation sample start in 2003.III.
BVAR: Bayesian quarterly VAR model; DFM: dynamic factor model; F-MIDAS-AR: factor augmented MIDAS model with AR-term.

accuracy, this can be viewed as evidence that forecasts by analysts do indeed embody a different type of valuable information (subjective judgments).

In our analysis, we use the mean quarterly forecast as the measure of private sector expectations. Fresh Consensus forecasts become available only once a quarter, in the second week of the last month of the quarter. For our information set, this means that Consensus forecasts are not updated in the first and second months of a quarter, while the monthly indicator series are updated every month. Moreover, at the time when panelists form their expectations, they have information on GDP growth in the preceding quarter. Thus, the Consensus backcast for quarter t is equal to the non-updated Consensus forecast published in the last month of quarter t.

Table 5 presents the results for Consensus forecasts for the complete sample period, the pre-crisis period and the crisis period.[17] For the one- and two-quarter-ahead forecasts, the Consensus forecasts are better than the best statistical model in the case of Spain, and competitive for three other countries (measured over the whole sample). However, the performance relative to the best model is weak for nowcasts and particularly backcasts for all countries except for Germany and Spain. Consequently, purely mechanical models seem to be more adept at

[17] Consensus forecasts for the euro area are available from March 2002 onward only, so the results in Table 5 for the euro area refer to the period 2003.III–2011.III.

learning when monthly information regarding the quarter of interest becomes available. In the relatively stable pre-crisis period, the Consensus forecasts fare very poorly, usually ranking at the bottom of the list. However, they do very well in the case of Germany, which suggests that analysts were able to assess the economic conditions better during and after the extraordinary episode of German reunification in the early 1990s. In contrast, Consensus forecasts perform much better during the crisis period, when GDP and the monthly indicators displayed extreme fluctuations. At the one- and two-quarter-ahead horizons, the Consensus forecasts belong in the top three models in most cases. For Spain and the Netherlands, the difference in forecasting precision is substantial. This suggests that analysts are able to handle extreme observations of GDP growth and auxiliary indicators better once they have occurred, while the quality of recursively estimated models in mechanical procedures is more susceptible to extreme observations in the sample, particularly when truly forecasting. Our findings support those of Lundquist and Stekler (2012), who conclude that professional forecasters are very responsive to the latest information about the state of the economy and adjust their predictions quickly. We find that, in spite of this head start, private sector forecasts still fall behind the best model in most cases as the horizon becomes shorter and more timely monthly information becomes available to improve forecasts. For example, leaving aside Spain, the RMSFEs of backcasts by Consensus forecasts are between 14% and 40% larger than those associated with the best model (dynamic factor model).

Despite the fact that the Consensus forecasts are a rather poor predictor of GDP on their own, the results for the encompassing test show that they often still contain valuable extra information. The effects are generally smaller for the best statistical model, which usually differs across horizons. This implies that a combination of model-based predictions and Consensus forecasts narrows the differences between the best models. This would lower the cost of using a single model for all horizons for practitioners. In many cases, an accuracy improvement of around 10% is feasible. The effects tend to be stronger for backcasts by analysts, even though these actually reflect relatively dated information. During the crisis period, Consensus forecasts, unlike their statistical competitors, still offer great added value compared to the best statistical model for Spain, and to a lesser extent for the Netherlands. Moreover, they can be used to improve the predictions of near-best models significantly for almost all countries. The added value is smaller in the pre-crisis period, except for Germany. All in all, the outcomes of the encompassing test suggest that subjective private sector forecasts potentially contain information that sophisticated mechanical forecasting procedures are unable to pick up.[18]

6. Conclusion

This paper makes two contributions to the empirical literature on forecasting real GDP in the short run. The first contribution is a systematic comparison of twelve statistical linear models for five countries (Germany, France, Italy, Spain and the Netherlands) and the euro area, utilizing the same information set across countries and the euro area. Our sample period (1996.I–2011.III) allows us to compare the models' forecasting abilities in the period before the financial crisis of 2008 (Great Moderation) and in the much more volatile period that followed (the financial crisis and its aftermath). The second contribution concerns the potential usefulness of (subjective) forecasts made by professional analysts. Such forecasts are very cheap and easy to use, and may incorporate valuable information that goes beyond purely statistical data.

We summarize our findings in five points. First, monthly indicators contain valuable information that can be extracted by mechanical statistical procedures, particularly as the horizon shortens and more monthly information is processed. The largest accuracy gains are for nowcasts and backcasts, suggesting that statistical models are especially helpful when they are able to use information that pertains to the quarter of interest. Moreover, statistical models are generally more efficient at extracting monthly information in volatile periods. Thus, their relative strength is to improve the assessment of the current state of the economy. In contrast, the predictions from statistical models generally incorporate little information at the two-quarter-ahead horizon.

Second, the dynamic factor model displays the best forecasting capabilities overall, particularly for backcasts and nowcasts. Its ability to incorporate more than one factor, and thus, more information, is key to this result. Factor-augmented MFVAR and MIDAS models produce better one-quarter-ahead predictions after the financial crisis, due to their richer dynamic specifications. However, the latter feature does not appear to be an advantage in stable times. The Bayesian quarterly VAR is the best quarterly model. It performs quite well for Germany, the Netherlands and Spain in the more stable period of the Great Moderation. Remarkably, all of the other models, including the dynamic factor model, perform (very) poorly in the case of Spain during the Great Moderation. These findings suggest that Bayesian estimation is a fundamentally different way of extracting information from a large data set, which may deliver benefits, even if the model is inefficient in its use of the available monthly information.

Third, regarding crucial model features, we find that employing factor analysis to summarize the available

[18] As fresh Consensus forecasts become available only in the third month of the quarter, the month-by-month pattern of the results is also interesting. We find that the relative RMSFE of Consensus forecasts (versus the best statistical model) is better in third months, when they are new, and deteriorates in other months, as the statistical models can take advantage of newly available monthly information. However, the results for the encompassing test on a month-by-month basis show that the value added by Consensus forecasts does not decrease significantly with age. Even Consensus forecasts that are one or two months old often contain valuable extra information. This finding reinforces our conjecture that analysts' forecasts embody information that differs in nature from the information that can be filtered out of statistical data. The monthly version of Table 5 can be found in the online appendix (see Appendix B).

monthly information clearly delivers better results than the alternative of averaging single-indicator-based forecasts in the case of one-quarter-ahead forecasts and nowcasts. Strategies that aggregate information work better than strategies that pool forecasts. Moreover, it is important for a model to make use of all of the available monthly observations. On the other hand, allowing for autoregressive terms (GDP's own past) in forecasting equations leads to only minor improvements in forecast reliability. All of these effects are more pronounced during the crisis period, implying that the cost of employing a suboptimal forecasting model is larger in periods of high volatility.

Fourth, statistical models differ significantly in the rates at which they are able to absorb monthly information as time goes by. However, the information content of the resulting forecasts appears to overlap to a large extent, and the unique model-specific component appears to be small (relative to the best model). The different models do not seem to have any comparative advantage of extracting certain types of information, offering perspectives that complement each other. The scope for improving GDP forecasts by combining the 'views' of various models is rather limited in economic terms, although there are some exceptions. This is particularly true during volatile episodes, when reliable assessments of the current situation and short run prospects are most needed, unfortunately.

Lastly, forecasts by professional analysts, which contain judgmental elements, appear to fall into a different category. In many cases, such forecasts are rather poor predictors of GDP compared to the best statistical model. However, they tend to perform better during the crisis, when it really counts, and they often embody information that sophisticated mechanical forecasting procedures fail to pick up. Thus, subjective private sector analysts' forecasts seem to offer the potential of enhancing mechanical forecasts.

The results of our large-scale comparative analysis may be useful to policy makers, financial analysts and economic agents, as information on where the economy stands and where it is heading in the short run is particularly valuable in times of great uncertainty. In practice, the dynamic factor model and factor-augmented statistical models are obvious candidate models for generating short-term forecasts.

An interesting topic for future research is the investigation of the ways in which the potential of judgmental forecasts may be taken on board in mechanical statistical procedures in a truly real-time context, in which both models and analysts have to deal with possibly inaccurate initial GDP estimates. Furthermore, considering the results of our Bayesian quarterly VAR model, more research on Bayesian mixed-frequency VAR models is needed, as these could develop into viable and practical alternatives to factor models. Another relevant issue is the finding of ways to incorporate more factors into the factor MIDAS model in order to improve its capabilities for nowcasting and backcasting. Finally, the construction of optimal weighting schemes for models that rely on pooling forecasts deserves more attention. Although we find that factor-based models generally perform better than models that pool indicator-specific forecasts, this may be due to suboptimal weighting schemes. As the latter category of models offers the attractive opportunity of calibrating weights on the basis of recent forecasting performances, the issue of optimal weights should be investigated further.

Acknowledgments

The opinions expressed in this paper are the authors' personal views and do not necessarily reflect the position of De Nederlandsche Bank or the Ministry of Finance. We are grateful to Jakob de Haan, Jan Jacobs, an anonymous Associate Editor, three anonymous referees and seminar and conference participants at De Nederlandsche Bank, the Computational and Financial Econometrics Conference 2012 in Oviedo and the European Meeting of the Econometrics Society 2013 in Gothenburg for their helpful comments. We are particularly indebted to Marta Bańbura for supplying us with the Matlab code for the estimation of the Bayesian VAR model, and to Job Swank for his valuable comments and stimulating suggestions.

Appendix A

A.1. Data

Quarterly GDP data for France, Italy, the Netherlands and Spain were taken from the OECD release data and revisions database (http://stats.oecd.org/mei/default.asp?rev=1). The source of the German GDP data is the Deutsche Bundesbank (http://www.bundesbank.de/statistik/statistik_zeitreihen.en.php?lang=en&open=&func=row&tr=JB5000). They refer to re-unified Germany from 1991.I onwards and to West Germany before 1991.I. We constructed a synthetic GDP series for the euro area using the database underlying the ECB's Area-Wide Model, supplemented with data from the OECD database. See http://www.eabcn.org/data/awm/index.htm.

The primary source of the monthly data is the ECB Statistical Datawarehouse (http://sdw.ecb.europa.eu/). World

Table A.6
Description of the monthly database.

No.	Description	Type	Transform			Country					
			log	dif.	filt.	EA	DE	FR	IT	ES	NL
1	World Trade (CPB)	Sales	1	1	3	'77	'77	'77	'77	'77	'77
2	World Industrial Production (CPB)	Sales	1	1	3	'91	'91	'91	'91	'91	'91
3	Ind. production: United States	Sales	1	1	3	'77	'77	'77	'77	'77	'77
4	Ind. production: United Kingdom	Sales	1	1	3	'77	'77	'77	'77	'77	'77

(continued on next page)

Table A.6 (*continued*)

No.	Description	Type	Transform			Country					
			log	dif.	filt.	EA	DE	FR	IT	ES	NL
5	Ind. production (excl. construction)	Sales	1	1	3	’77	’77	’77	’77	’77	’77
6	Ind. production, cons. goods ind.	Sales	1	1	3	’80	’80	’77	’77	’77	’90
7	Ind. production, energy	Sales	1	1	3	’80	’80	’77	’80	’80	’90
8	Ind. production, interm. goods ind.	Sales	1	1	3	’77	’80	’77	’77	’77	’95
9	Ind. production, capital goods	Sales	1	1	3	’77	’80	’77	’77	’77	’77
10	Ind. production, manufacturing	Sales	1	1	3	’77	’78	’77	’77	’80	’77
11	Ind. production, construction	Sales	1	1	3	’85	’78	’85	’95	’88	’85
12	New orders manufacturing	Sales	1	1	3	’95	’91	’00	’90	’00	’95
13	New passenger cars (reg.)	Sales	1	1	3	’90	’90	’90	’90	’90	’90
14	New commercial vehicles (reg.)	Sales	1	1	3	’90	’90	’90	’90	’90	’90
15	Retail trade volume	Sales	1	1	3	’77	’77	’77	’90	’95	’77
16	Unemployment rate	Sales	0	1	3	’83	’91	’83	’83	’86	’83
17	Unemployment rate: United Kingdom	Sales	0	1	3	’83	’83	’83	’83	’83	’83
18	Unemployment rate: United States	Sales	0	1	3	’83	’83	’83	’83	’83	’83
19	Exports	Sales	1	1	3	’00	’89	’89	’89	’89	’89
20	Imports	Sales	1	1	3	’00	’89	’89	’89	’89	’89
21	Total HICP index	Prices	1	2	3	’77	’77	’77	’77	’77	’77
22	Core HICP index	Prices	1	2	3	’77	’77	’77	’77	’77	’77
23	CPI, food	Prices	1	2	3	’90	’77	’77	’77	’93	’77
24	CPI, energy	Prices	1	2	3	’90	’77	’77	’77	’77	’77
25	HICP, services	Prices	1	2	3	’90	’85	’90	’87	’92	’87
26	Producer prices (total, excl. constr.)	Prices	1	2	3	’81	’77	’77	’77	’77	’77
27	World commodity prices, total	Prices	1	2	3	’77	’77	’77	’77	’77	’77
28	World commodity prices, raw mat.	Prices	1	2	3	’77	’77	’77	’77	’77	’77
29	World commodity prices, food	Prices	1	2	3	’77	’77	’77	’77	’77	’77
30	World commodity prices, metals	Prices	1	2	3	’77	’77	’77	’77	’77	’77
31	World commodity prices, energy	Prices	1	2	3	’77	’77	’77	’77	’77	’77
32	Oil price (1 month future Brent)	Prices	1	2	3	’77	’77	’77	’77	’77	’77
33	M1	Finan.	1	1	3	’77	’80	’77	’80	’80	’80
34	M3	Finan.	1	1	3	’77	’77	’77	’77	’77	’77
35	Interest rate on mortgage	Finan.	0	1	3	’03	’82	’80	’95	’84	’80
36	3 month interest rate euro	Finan.	0	1	3	’94	’77	’77	’77	’77	’77
37	10 year government bond yield	Finan.	0	1	3	’77	’94	’77	’77	’80	’77
40	Headline stock index	Finan.	1	1	3	’77	’77	’77	’77	’87	’77
41	Basic material index	Finan.	1	1	3	’77	’77	’77	’77	’87	’77
42	Industrials stock index	Finan.	1	1	3	’77	’77	’77	’77	’87	’77
43	Consumer goods stock index	Finan.	1	1	3	’77	’77	’77	’77	’87	’77
44	Consumer services stock index	Finan.	1	1	3	’77	’77	’77	’87	’77	’77
45	Financials stock index	Finan.	1	1	3	’77	’77	’77	’77	’87	’77
46	Technology stock index	Finan.	1	1	3	’77	’88	’77	’86	’99	’85
47	Loans to the private sector	Finan.	1	1	3	’91	’80	’80	’83	’80	’82
48	Exchange rate, US Dollars per Euro	Finan.	1	1	3	’80	’80	’80	’80	’80	’80
49	Real effective exchange rate (CPI)	Finan.	1	1	3	’77	’77	’77	’77	’77	’77
50	Ind. conf.: headline	Survey	0	1	3	’85	’85	’85	’85	’87	’85
51	Ind. conf.: Order-book expect.	Survey	0	1	3	’85	’85	’85	’85	’87	’85
52	Ind. conf.: Stocks expect.	Survey	0	1	3	’85	’85	’85	’85	’87	’85
53	Ind. conf.: Production expect.	Survey	0	1	3	’85	’85	’85	’85	’87	’85
54	Ind. conf.: Employment expect.	Survey	0	1	3	’85	’85	’85	’85	’87	’85
55	Cons. conf.: headline	Survey	0	1	3	’85	’85	’85	’85	’86	’85
56	Cons. conf.: Financial sit.	Survey	0	1	3	’85	’85	’85	’85	’86	’85
57	Cons. conf.: General ec. sit.	Survey	0	1	3	’85	’85	’85	’85	’86	’85
58	Cons. conf.: Unemployment expect.	Survey	0	1	3	’85	’85	’85	’85	’86	’85
59	Cons. conf.: Major purchases expect.	Survey	0	1	3	’85	’85	’85	’85	’86	’85
60	Constr. conf.: Headline	Survey	0	1	3	’85	’85	’85	’85	’89	’85
61	Constr. conf.: Order book (evolution)	Survey	0	1	3	’85	’85	’85	’85	’89	’85
62	Constr. conf.: Employment expect.	Survey	0	1	3	’85	’85	’85	’85	’89	’85
63	Retail conf.: Headline	Survey	0	1	3	’85	’85	’85	’85	’88	’86
64	Retail conf.: Current stocks (volume)	Survey	0	1	3	’85	’85	’85	’85	’88	’86
65	Retail conf.: Orders expectations	Survey	0	1	3	’85	’85	’85	’85	’88	’86
66	Retail conf.: Business expect.	Survey	0	1	3	’85	’85	’85	’85	’88	’86
67	Retail conf.: Employment expect.	Survey	0	1	3	’86	’85	’85	’86	’88	’86
68	PMI: United States	Survey	0	1	3	’77	’77	’77	’77	’77	’77
69	PMI: United Kingdom	Survey	0	1	3	’92	’92	’92	’92	’92	’92
70	OECD composite leading ind. UK	Other	0	1	3	’77	’77	’77	’77	’77	’77
71	OECD composite leading ind. US	Other	0	1	3	’77	’77	’77	’77	’77	’77
72	OECD composite leading ind.	Other	0	1	3	’77	’77	’77	’77	’77	’77

Notes: type: sales = production and sales; finan. = monetary and financial; price = price data; survey = surveys. Transform: log: 0 = no logarithm, 1 = logarithm; dif.: degree of differencing 1 = first difference, 2 = second difference; filt.: 3 = change against the same month of the previous quarter.

trade and world industrial production are from the CPB World trade monitor (http://www.cpb.nl/en/world-trade-monitor). Commodity prices and most of the financial market variables were taken from Thomson Reuters Datastream, and most of the survey data from the European Commission (http://ec.europa.eu/economy_finance/db_indicators/surveys/index_en.htm). Table A.6 provides an overview of all monthly variables, with the transformations applied and the starting dates of the monthly series at the data sources. Note that only observations from January 1985 onwards have been used for estimation.

A.2. State space representations

A.2.1. Mixed frequency VAR

This section describes the state space representation of the mixed frequency VAR described in Section 2.3.2. Let $p^* = \max(p, 3)$, and the transition equation of the state vector is as follows:

$$\begin{bmatrix} z_{i,t+1} - \mu_{z_i} \\ z_{i,t} - \mu_{z_i} \\ \cdots \\ z_{i,t-p^*+2} - \mu_{z_i} \end{bmatrix} = \begin{bmatrix} A_1 & A_2 & \cdots & A_p & 0_{2\times 2(3-p^*)} \\ I_{2(p^*-1)} & & & & 0_{2(p^*-1)\times 2} \end{bmatrix} \times \begin{bmatrix} z_{i,t} - \mu_{z_i} \\ z_{i,t-1} - \mu_{z_i} \\ \cdots \\ z_{i,t-p^*+1} - \mu_{z_i} \end{bmatrix} + \begin{bmatrix} \Sigma_\varepsilon^{1/2} \\ 0_{2(p^*-1)\times 2} \end{bmatrix} \upsilon_t, \tag{20}$$

where $\upsilon_t \sim N(0, I_2)$. The measurement equation is:

$$z_{i,t}^Q - \mu_{z_i^Q} = \begin{bmatrix} \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0_{1\times(p^*-6)} \\ 0 & 1 & 0 & 0 & 0 & 0 & 0_{1\times(p^*-6)} \end{bmatrix} \times \begin{bmatrix} z_{i,t} - \mu_{z_i} \\ z_{i,t-1} - \mu_{z_i} \\ \cdots \\ z_{i,t-p^*+1} - \mu_{z_i} \end{bmatrix}. \tag{21}$$

Since y_t^Q is assigned to the third month of the quarter, we replace the missing observations in months 1 and 2 with a random draw from the standard normal distribution $N(0, 1)$, as in Mariano and Murasawa (2010). We modify the measurement equation of months 1 and 2 in accordance with the missing observation treatment. For months for which y_t^Q is unavailable, the upper row of the matrix on the right hand side of Eq. (21) is set equal to zero and white noise is added.

A.2.2. Dynamic factor model

The DFM equations, Eqs. (5)–(8), can be cast in state space form, as is illustrated below for the case of $p = 1$. The aggregation rule is implemented in a recursive way in Eq. (23) by introducing a latent cumulator variable Ξ, for which $\Xi_t = 0$ for t corresponding to the first month of the quarter and $\Xi_t = 1$ otherwise. The monthly state space representation is given by the following observation equation:

$$\begin{bmatrix} x_t \\ y_t^Q \end{bmatrix} = \begin{bmatrix} \Lambda & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_t \\ y_t \\ \hat{y}_t^Q \end{bmatrix} + \begin{bmatrix} \xi_t \\ \varepsilon_t^Q \end{bmatrix} \tag{22}$$

and the transition equation:

$$\begin{bmatrix} I_r & 0 & 0 \\ -\beta' & 1 & 0 \\ 0 & -\frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} f_{t+1} \\ y_{t+1} \\ \hat{y}_{t+1}^Q \end{bmatrix} = \begin{bmatrix} A_{r1} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Xi_{t+1} \end{bmatrix} \begin{bmatrix} f_t \\ y_t \\ \hat{y}_t^Q \end{bmatrix} + \begin{bmatrix} \zeta_{t+1} \\ \varepsilon_t \\ 0 \end{bmatrix}. \tag{23}$$

The application of the Kalman filter and smoother provides the minimum mean square linear estimates (MMSLE) of the state vector $\alpha_t = (f_t, y_t, \hat{y}_t^Q)$ and enables the quarterly GDP growth y_t^Q to be forecast. Moreover, it enables efficient handling of an unbalanced dataset with missing observations at the end of the series by replacing the missing data with optimal predictions. Moreover, relative to using the principal components technique alone, the two-step estimator allows for the dynamics of the common factors and the cross-sectional heteroskedasticity of the idiosyncratic component.

A.3. Additional results

Tables A.7 and A.8 present a sensitivity analysis regarding the numbers of factors in the factor-augmented versions of the MIDAS-AR, MFVAR and QVAR models. In the case of the F-VAR model, we derive the factors by applying simple principal component analysis. For the mixed-frequency models, we use the Kalman-filtered factors generated by the dynamic factor model, averaged over all possible parameterizations. The maximum number of factors is set at six. Table A.7 treats each factor as a separate indicator and reports the RMSFEs of a weighted average of factor-specific forecasts relative to those of the one-factor versions reported in Table 2. For all countries except Spain, the incorporation of additional factors tends to push up the RSMFEs for all horizons. For Spain, using two extra factors improves the forecast accuracy of the F-MIDAS-AR model by 20%. For the F-MFVAR and F-VAR models, the effect is fairly small. Table A.8 looks into the marginal value of the factors when they are added sequentially to the encompassing test regression in Eq. (19). This analysis leads to the same conclusion.

Table A.9 presents a sensitivity analysis regarding the three weighting schemes for indicator-specific indicators used in the BEQ-AR, QVAR, MIDAS-AR and MFVAR models. In addition, the table also considers the weighting scheme for the different varieties of the dynamic factor model in terms of the numbers of static factors, dynamic factors and lags (126 parametrizations in total). All of the weighting schemes produce similar results.

Finally, Tables A.10–A.13 are the counterparts of Tables 2 and 3 in the main text, focusing on the forecast accuracy and marginal value of models during the Great Moderation (1996.I–2007.IV) and the financial crisis (2008.I–2011.III).

Table A.7
Relative RMSFEs of multiple single-factor models versus a one-factor model, 1996.I–2011.III.

♯ factors	F-MIDAS-AR					F-MFVAR					F-VAR				
	2	3	4	5	6	2	3	4	5	6	2	3	4	5	6
Euro area															
2Q ahead	1.03	1.04	1.04	1.05	1.05	0.97	0.97	0.97	0.97	0.97	1.00	0.99	1.00	1.00	1.00
1Q ahead	1.04	1.07	1.07	1.09	1.11	0.97	0.97	0.97	0.97	0.97	1.01	1.01	1.02	1.02	1.02
Nowcast	1.04	1.06	1.05	1.07	1.08	1.00	0.99	0.99	0.98	0.99	1.02	1.03	1.04	1.05	1.05
Backcast	1.02	1.02	1.02	1.03	1.05	0.98	0.97	0.97	0.97	0.96	1.01	1.02	1.03	1.04	1.04
Germany															
2Q ahead	1.02	1.03	1.04	1.04	1.04	0.99	0.98	0.98	0.97	0.97	0.99	0.99	0.99	0.99	0.99
1Q ahead	1.01	1.02	1.02	1.03	1.04	0.99	0.98	0.98	0.98	0.98	1.00	1.01	1.01	1.01	1.01
Nowcast	1.02	1.05	1.06	1.08	1.10	0.98	0.98	0.98	0.98	0.99	1.02	1.04	1.04	1.04	1.04
Backcast	0.98	0.99	1.00	1.01	1.03	0.97	0.98	0.97	0.97	0.97	1.04	1.06	1.06	1.07	1.07
France															
2Q ahead	1.00	1.01	1.01	1.02	1.03	0.99	1.00	1.00	1.00	0.99	0.99	0.98	0.99	1.00	1.00
1Q ahead	1.03	1.05	1.06	1.07	1.08	0.98	0.99	1.00	0.99	0.99	1.00	0.98	1.00	1.00	1.00
Nowcast	1.03	1.07	1.07	1.10	1.12	0.98	0.97	0.97	0.97	0.97	1.01	0.99	1.00	1.01	1.00
Backcast	1.00	1.00	0.99	0.99	1.01	0.98	0.96	0.96	0.96	0.96	1.02	1.01	1.03	1.04	1.04
Italy															
2Q ahead	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	1.00	1.00	1.01	1.01	1.01
1Q ahead	1.02	1.05	1.05	1.06	1.07	0.99	0.99	0.99	0.99	0.98	1.00	1.01	1.01	1.01	1.02
Nowcast	1.04	1.06	1.07	1.09	1.12	0.99	0.98	0.98	0.98	0.97	1.01	1.02	1.03	1.04	1.04
Backcast	1.00	1.00	1.00	1.00	1.01	1.01	0.99	0.99	0.98	0.98	1.03	1.05	1.06	1.07	1.07
Spain															
2Q ahead	0.93	0.91	0.91	0.92	0.92	0.98	0.97	0.97	0.97	0.97	1.00	1.02	1.04	1.04	1.04
1Q ahead	0.92	0.87	0.87	0.86	0.84	1.01	0.98	0.99	1.00	1.00	0.99	1.04	1.06	1.08	1.08
Nowcast	0.87	0.80	0.80	0.78	0.78	1.02	0.99	1.00	1.00	1.00	0.96	1.02	1.07	1.09	1.10
Backcast	0.89	0.80	0.81	0.80	0.79	1.02	1.00	1.01	1.01	1.01	0.97	1.02	1.07	1.10	1.12
Netherlands															
2Q ahead	1.01	1.02	1.03	1.02	1.02	0.97	0.97	0.97	0.97	0.97	1.01	1.01	1.02	1.02	1.02
1Q ahead	1.00	1.01	1.03	1.04	1.05	0.98	0.99	0.99	0.99	0.99	1.01	1.02	1.03	1.03	1.04
Nowcast	0.99	0.99	1.01	1.02	1.03	1.00	1.00	1.01	1.00	1.00	1.01	1.03	1.04	1.05	1.05
Backcast	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.99	1.02	1.04	1.06	1.06	1.07

Notes: RMSFE (multiple single-factor models)/RMSFE (one-factor model). The factor-specific forecasts are weighted via a four-quarter moving window RMSFE scheme.
F-VAR: factor augmented quarterly VAR model; F-MFVAR: factor augmented mixed frequency VAR model; F-MIDAS-AR: factor augmented mixed data sampling model with an AR term.

Table A.8
Marginal value of additional factors (evaluation period 1996.I–2011.III).

♯ factors	F-MIDAS-AR					F-MFVAR					F-VAR				
	2	3	4	5	6	2	3	4	5	6	2	3	4	5	6
Euro area															
2Q ahead	–	–	–	–	–	**0.95**	–	–	–	–	–	0.99	–	–	–
1Q ahead	–	–	–	–	–	**0.95**	–	–	–	–	–	–	–	–	–
Nowcast	–	–	–	–	–	0.99	**0.99**	–	–	–	–	–	–	–	–
Backcast	–	0.99	–	–	–	**0.97**	–	–	–	**0.98**	–	–	–	–	–
Germany															
2Q ahead	–	–	–	–	–	**0.98**	–	**0.99**	–	–	0.99	–	–	–	–
1Q ahead	–	–	–	–	–	**0.97**	–	–	–	–	–	–	–	–	–
Nowcast	–	–	–	–	–	**0.96**	–	–	–	–	–	–	–	–	–
Backcast	**0.98**	–	–	–	–	**0.96**	–	–	0.99	–	–	–	–	–	–
France															
2Q ahead	–	–	–	–	–	**0.98**	–	–	–	0.99	0.99	**0.99**	–	–	–
1Q ahead	–	–	–	–	–	**0.97**	–	–	–	–	–	**0.97**	–	–	–
Nowcast	–	–	–	–	–	**0.98**	**0.99**	–	–	–	–	**0.97**	–	–	–
Backcast	0.99	–	**0.98**	–	–	**0.97**	**0.98**	–	–	–	–	0.99	–	–	–
Italy															
2Q ahead	–	–	–	–	–	–	–	0.99	–	0.99	–	–	–	–	–
1Q ahead	–	–	–	–	–	**0.98**	–	–	–	0.99	–	–	–	–	–
Nowcast	–	–	–	–	–	**0.99**	**0.99**	–	0.99	–	–	–	–	–	–
Backcast	–	0.99	–	–	–	–	**0.96**	–	–	–	–	–	–	–	–
Spain															
2Q ahead	**0.93**	**0.98**	0.99	–	–	**0.98**	**0.97**	–	–	–	0.99	–	–	–	–
1Q ahead	**0.92**	**0.92**	–	–	**0.97**	–	**0.95**	–	–	–	**0.99**	–	–	–	–
Nowcast	**0.83**	**0.88**	–	–	–	–	**0.95**	–	–	–	**0.97**	–	–	–	–

(continued on next page)

Table A.8 (*continued*)

♯ factors	F-MIDAS-AR					F-MFVAR					F-VAR				
	2	3	4	5	6	2	3	4	5	6	2	3	4	5	6
Backcast	**0.88**	**0.81**	–	–	–	–	**0.98**	–	–	–	**0.97**	–	–	–	–
Netherlands															
2Q ahead	–	–	–	–	–	**0.95**	–	–	–	–	–	–	–	–	–
1Q ahead	–	–	–	–	–	**0.98**	–	–	–	–	–	–	–	–	–
Nowcast	0.99	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Backcast	–	–	0.99	–	–	**0.98**	–	–	–	–	–	–	–	–	–

Notes: Figures in boldface indicate that the encompassing test is statistically significant at the 5% level. Dots indicate the corner solution of zero weight. F-VAR: factor augmented quarterly VAR model; F-MFVAR: factor augmented mixed frequency VAR model; F-MIDAS-AR: factor augmented mixed data sampling model with an AR term.

Table A.9
Pooling schemes (evaluation period 1996.I–2011.III).

	BEQ-AR			QVAR			DFM			MIDAS-AR			MFVAR		
	AV	RC	4Q	AV	RC	4Q	AV	RC	4Q	AV	RC	4Q	AV	RC	4Q
Euro area															
2Q ahead	0.62	0.62	0.62	0.65	0.65	0.64	0.63	0.63	0.63	0.62	0.62	0.62	0.62	0.62	0.62
1Q ahead	0.58	0.59	0.57	0.63	0.64	0.63	0.57	0.57	0.57	0.59	0.59	0.59	0.60	0.59	0.59
Nowcast	0.51	0.52	0.50	0.58	0.58	0.58	0.44	0.44	0.43	0.51	0.53	0.50	0.53	0.50	0.52
Backcast	0.44	0.45	0.43	0.50	0.50	0.49	0.32	0.32	0.31	0.45	0.45	0.39	0.46	0.47	0.44
Germany															
2Q ahead	0.90	0.90	0.90	0.92	0.93	0.92	0.90	0.90	0.90	0.93	0.93	0.93	0.91	0.90	0.91
1Q ahead	0.88	0.86	0.88	0.91	0.92	0.91	0.84	0.84	0.84	0.89	0.91	0.89	0.89	0.88	0.89
Nowcast	0.84	0.82	0.82	0.89	0.90	0.90	0.70	0.70	0.70	0.85	0.86	0.84	0.85	0.81	0.82
Backcast	0.81	0.77	0.78	0.87	0.87	0.88	0.61	0.61	0.60	0.82	0.78	0.75	0.81	0.74	0.72
France															
2Q ahead	0.51	0.51	0.51	0.53	0.53	0.54	0.50	0.50	0.50	0.51	0.51	0.51	0.52	0.51	0.51
1Q ahead	0.48	0.46	0.48	0.51	0.51	0.53	0.44	0.44	0.44	0.49	0.49	0.48	0.48	0.48	0.48
Nowcast	0.42	0.41	0.41	0.46	0.46	0.46	0.34	0.34	0.34	0.43	0.42	0.42	0.43	0.41	0.42
Backcast	0.39	0.39	0.38	0.42	0.42	0.41	0.28	0.28	0.28	0.39	0.38	0.37	0.38	0.37	0.38
Italy															
2Q ahead	0.73	0.73	0.74	0.75	0.75	0.76	0.74	0.74	0.74	0.74	0.73	0.73	0.74	0.74	0.74
1Q ahead	0.71	0.70	0.71	0.73	0.73	0.74	0.68	0.68	0.68	0.72	0.72	0.71	0.71	0.71	0.71
Nowcast	0.65	0.65	0.64	0.70	0.68	0.70	0.55	0.55	0.55	0.65	0.65	0.65	0.65	0.65	0.65
Backcast	0.60	0.59	0.59	0.63	0.62	0.63	0.47	0.47	0.47	0.60	0.58	0.59	0.58	0.58	0.58
Spain															
2Q ahead	0.59	0.57	0.58	0.59	0.58	0.59	0.56	0.56	0.56	0.59	0.59	0.57	0.61	0.59	0.59
1Q ahead	0.56	0.54	0.53	0.55	0.54	0.54	0.47	0.47	0.46	0.51	0.49	0.49	0.58	0.56	0.56
Nowcast	0.54	0.54	0.49	0.49	0.46	0.47	0.40	0.39	0.39	0.47	0.50	0.42	0.56	0.55	0.52
Backcast	0.56	0.65	0.49	0.48	0.45	0.45	0.36	0.35	0.35	0.55	0.57	0.47	0.57	0.56	0.52
Netherlands															
2Q ahead	0.69	0.69	0.69	0.71	0.71	0.71	0.71	0.71	0.71	0.69	0.70	0.69	0.71	0.71	0.70
1Q ahead	0.66	0.65	0.64	0.70	0.70	0.70	0.64	0.64	0.64	0.66	0.66	0.65	0.68	0.68	0.69
Nowcast	0.61	0.60	0.58	0.66	0.66	0.66	0.54	0.54	0.53	0.62	0.60	0.60	0.64	0.63	0.62
Backcast	0.57	0.55	0.55	0.59	0.59	0.59	0.48	0.48	0.47	0.57	0.55	0.54	0.62	0.61	0.60

Notes: BEQ-AR: bridge equation with an AR term; QVAR: quarterly VAR model; DFM: dynamic factor model; MIDAS-AR: mixed data sampling model with an AR term; MFVAR: mixed frequency VAR model.
AV: equal weights (simple average); RC: weights inversely proportional to the RMSFE; recursively calculated; 4Q: weights inversely proportional to the RMSFE, calculated over the last four quarters. Shaded cells indicate the lowest RMSFE.

Table A.10
Forecasting performances of statistical models (RMSFE), 1996.I–2007.IV.

Frequency model	Benchmark		BVAR	Pooling forecasts				Pooling information			
	RW	AR		BEQ-AR	QVAR	MIDAS-AR	MFVAR	DFM	F-VAR	F-MIDAS-AR	F-MFVAR
Euro area											
2Q ahead	*0.34*	**1.00**	1.08	**0.96**	**0.99**	0.96	**0.98**	**0.99**	1.02	1.00	1.01
1Q ahead	*0.34*	**0.99**	1.03	**0.92**	**0.97**	**0.91**	**0.94**	**0.92**	**1.00**	**0.91**	0.91
Nowcast	*0.34*	0.96	0.93	0.87	0.93	**0.85**	0.89	**0.82**	0.93	0.86	0.78
Backcast	*0.34*	0.91	0.83	0.86	0.88	0.82	0.84	0.72	0.85	0.79	0.88
Germany											
2Q ahead	*0.62*	1.04	1.01	0.97	**0.97**	**0.98**	**0.99**	**0.99**	1.00	1.07	1.01
1Q ahead	*0.62*	1.05	**0.97**	**0.98**	0.97	**0.99**	**0.98**	**0.97**	**1.00**	1.04	**0.99**
Nowcast	*0.62*	1.05	**0.92**	**0.94**	**0.99**	**0.95**	0.91	**0.93**	**0.97**	**1.00**	**0.97**
Backcast	*0.62*	1.06	0.90	**0.91**	**0.99**	**0.93**	**0.91**	**0.92**	**0.96**	**0.97**	1.06
France											
2Q ahead	*0.35*	1.00	1.05	**0.95**	**0.98**	**0.95**	**0.96**	0.90	1.01	**0.99**	**0.96**
1Q ahead	*0.35*	0.98	1.00	**0.90**	0.95	0.92	0.92	0.82	0.99	**0.89**	0.91
Nowcast	*0.35*	0.94	0.92	0.83	0.89	0.86	0.85	0.70	0.91	**0.73**	**0.72**
Backcast	*0.35*	0.93	0.85	0.86	0.88	0.84	0.84	0.65	0.89	0.76	0.78
Italy											
2Q ahead	*0.50*	1.03	1.00	**0.99**	0.99	**0.99**	**0.99**	1.01	1.00	1.10	1.01
1Q ahead	*0.50*	1.02	**1.00**	**0.99**	**0.98**	**0.98**	**0.97**	**0.93**	1.00	**0.98**	0.91
Nowcast	*0.50*	1.00	0.96	0.91	0.96	0.92	0.92	**0.85**	0.95	**0.86**	0.80
Backcast	*0.50*	0.97	0.91	0.88	0.90	0.89	0.88	**0.82**	**0.84**	**0.86**	0.79
Spain											
2Q ahead	*0.32*	**0.96**	0.93	**0.95**	1.02	**0.99**	**0.96**	**0.96**	1.00	**0.98**	1.02
1Q ahead	*0.31*	**0.91**	0.88	**0.95**	**0.93**	**0.92**	0.97	**0.93**	0.97	1.10	1.03
Nowcast	*0.31*	**0.87**	0.82	1.00	0.95	0.92	0.99	0.96	1.02	1.02	1.06
Backcast	*0.31*	0.91	0.78	1.12	1.02	1.05	1.24	1.00	1.16	1.05	1.19
Netherlands											
2Q ahead	*0.53*	1.00	**0.93**	**0.97**	**0.99**	**0.95**	**0.98**	**0.97**	**0.96**	0.91	**0.94**
1Q ahead	*0.53*	**1.00**	**0.89**	**0.92**	0.98	**0.91**	**0.97**	**0.94**	**0.93**	0.89	**0.90**
Nowcast	*0.53*	0.99	0.85	**0.92**	0.97	**0.92**	0.94	**0.90**	**0.90**	**0.88**	0.94
Backcast	*0.52*	0.97	0.82	0.91	0.95	**0.88**	0.96	**0.87**	**0.87**	**0.88**	1.00

Notes: The entries for RW (in italics) refer to the RMSFE; those for the other models refer to the RMSFEs relative to the RW model's RMSFEs. Grey cells indicate the models with the lowest RMSFEs. Figures in boldface indicate models with RMSFEs that are no more than 10% larger than those of the best model.
RW: random walk; AR: autoregressive model; BEQ-AR: bridge equation with an AR term; QVAR: quarterly vector autoregressive model; BVAR: Bayesian QVAR model; F-VAR: factor augmented QVAR model; DFM: dynamic factor model; MFVAR: mixed frequency vector autoregressive model; F-MFVAR: factor augmented MFVAR model; MIDAS-AR: Mixed data sampling model with AR-term; F-MIDAS-AR: factor augmented MIDAS-AR.

Table A.11
Forecasting performances of the statistical models (RMSFE), 2008.I–2011.III.

Frequency model	Benchmark		BVAR	Pooling forecasts				Pooling information			
	RW	AR		BEQ-AR	QVAR	MIDAS-AR	MFVAR	DFM	F-VAR	F-MIDAS-AR	F-MFVAR
Euro area											
2Q ahead	*1.15*	1.00	0.94	**0.97**	1.02	**0.97**	**0.97**	**0.99**	**0.99**	1.02	1.01
1Q ahead	*1.15*	1.00	0.90	0.90	1.01	0.93	0.93	0.89	0.95	0.80	**0.85**
Nowcast	*1.14*	0.95	0.80	0.77	0.92	0.78	0.80	**0.65**	0.87	0.61	**0.63**
Backcast	*1.14*	0.83	0.68	0.63	0.76	0.55	0.65	0.43	0.78	0.57	0.58
Germany											
2Q ahead	*1.50*	1.05	1.03	**0.99**	1.02	1.04	**0.99**	**0.99**	1.01	0.97	1.02
1Q ahead	*1.50*	1.03	1.01	**0.96**	1.01	0.96	0.97	**0.89**	0.97	**0.89**	0.87
Nowcast	*1.50*	1.01	0.95	0.89	0.99	0.90	0.90	**0.66**	0.92	**0.67**	0.63
Backcast	*1.49*	1.01	0.86	0.82	0.96	0.76	0.72	0.48	0.88	0.58	0.61
France											
2Q ahead	*0.88*	1.03	0.87	0.97	1.03	0.97	0.97	0.96	0.99	1.00	0.98
1Q ahead	*0.88*	0.99	**0.84**	0.91	1.03	0.91	0.91	**0.84**	0.97	**0.82**	0.81
Nowcast	*0.87*	0.88	0.76	0.75	0.88	0.75	0.77	0.61	0.88	0.55	**0.58**
Backcast	*0.87*	0.79	0.67	0.65	0.75	0.64	0.65	0.45	0.75	0.52	0.54
Italy											
2Q ahead	*1.24*	1.04	0.92	**0.98**	1.02	**0.98**	**0.99**	**0.98**	**0.99**	1.05	**0.98**
1Q ahead	*1.24*	0.95	**0.90**	**0.94**	0.99	**0.94**	**0.94**	**0.91**	0.95	0.86	**0.87**
Nowcast	*1.23*	0.92	0.85	0.84	0.94	0.84	0.84	**0.68**	0.88	0.63	**0.67**
Backcast	*1.23*	0.89	0.76	0.75	0.82	0.74	0.72	0.52	0.78	0.60	0.60
Spain											
2Q ahead	*1.17*	**0.91**	**0.85**	**0.91**	**0.90**	**0.88**	0.92	**0.86**	**0.87**	0.83	**0.91**
1Q ahead	*1.16*	0.81	**0.74**	0.82	0.83	**0.75**	0.87	0.70	0.77	0.78	0.79
Nowcast	*1.15*	0.71	**0.57**	0.72	0.70	0.61	0.79	0.53	0.67	0.62	0.75
Backcast	*1.15*	0.76	0.40	0.68	0.63	0.67	0.71	**0.41**	0.62	0.52	0.75
Netherlands											
2Q ahead	*1.12*	**0.99**	0.90	**0.96**	**0.99**	**0.98**	**0.98**	1.01	**0.97**	**0.97**	1.08
1Q ahead	*1.11*	0.99	0.89	0.88	0.98	0.92	0.96	0.88	0.94	0.78	0.88
Nowcast	*1.11*	0.94	0.82	0.73	0.90	0.80	0.83	**0.64**	0.88	0.61	**0.65**
Backcast	*1.10*	0.85	0.70	0.66	0.75	0.67	0.75	0.51	0.76	0.58	0.60

Notes: The entries for RW (in italics) refer to the RMSFE; those for the other models refer to RMSFEs relative to the RW model's RMSFEs. Grey cells indicate models with the lowest RMSFEs. Figures in boldface indicate models with RMSFEs that are no more than 10% larger than those of the best model.
RW: random walk; AR: autoregressive model, BEQ-AR: bridge equation with an AR term; QVAR: quarterly vector autoregressive model; BVAR: Bayesian QVAR model; F-VAR: factor augmented QVAR model; DFM: dynamic factor model; MFVAR: mixed frequency vector autoregressive model; F-MFVAR: factor augmented MFVAR model; MIDAS-AR: Mixed data sampling model with AR-term; F-MIDAS-AR: factor augmented MIDAS-AR.

Table A.12
Marginal value of the statistical models, 1996.I–2007.IV.

Frequency model	AR	BVAR	Pooling forecasts				Aggregating information			
			BEQ-AR	QVAR	MIDAS-AR	MFVAR	DFM	F-VAR	F-MIDAS-AR	F-MFVAR
Euro area										
2Q ahead	–	–	–	–		–	–	–	–	–
1Q ahead	**0.98**	**0.99**	**0.97**	**0.99**	**0.97**	**0.98**	0.99	**0.99**	0.99	
Nowcast	–	**0.98**	–	–	0.99	–	0.99	–	–	
Backcast	–	0.99	–	–	–	–		–	0.99	–
Germany										
2Q ahead	–	–		–	**0.99**	–	–	–	–	–
1Q ahead	–	**0.98**	–		–	–	**0.98**	–	–	**0.98**
Nowcast	–	**0.97**	–	–	–		**0.99**	–	0.99	0.99
Backcast	–		**0.95**	0.99	**0.97**	**0.95**	**0.96**	**0.98**	**0.97**	**0.98**
France										
2Q ahead	–	–	–	–	–	–		–	–	–
1Q ahead	–	–	–	–	–	–		–	–	–
Nowcast	–	–	–	–	–	–		–	**0.98**	**0.98**
Backcast	–	–	–	–	–	–		–	–	–
Italy										
2Q ahead	–	–	–		–	–	–	–	–	–
1Q ahead	–	–	–	–	–	–	–	–	–	
Nowcast	–	–	–	–	–	–	–	–	–	
Backcast	–	–	–	–	–	–	0.99	–	–	
Spain										
2Q ahead	–		**0.99**	–	–	–	**0.98**	–	**0.99**	**0.98**
1Q ahead	–		0.99	–	0.99	–	**0.97**	–	–	**0.98**
Nowcast	**0.99**		–	–	–	–	**0.97**	–	**0.98**	**0.98**
Backcast	–		–	–	–	–	**0.98**	–	**0.98**	0.99
Netherlands										
2Q ahead	–	**0.98**	–	–	–	–	–	–		–
1Q ahead	–	**0.97**	0.99	–	0.99	–	–	–		–
Nowcast	–		**0.99**	–	0.99	**0.99**	**0.98**	**0.99**	**0.98**	**0.98**
Backcast	–		0.99	–	0.99	0.99	**0.98**	**0.98**	0.99	0.99

Notes: Grey cells indicate the models with the lowest RMSFEs. Figures in boldface indicate that the encompassing test is statistically significant at the 5% level.
AR: autoregressive model; BEQ-AR: bridge equation with an AR term; QVAR: quarterly vector autoregressive model; BVAR: Bayesian QVAR model; F-VAR: factor augmented QVAR model; DFM: dynamic factor model; MFVAR: mixed frequency vector autoregressive model; F-MFVAR: factor augmented MFVAR model; MIDAS-AR: mixed data sampling model with AR-term; F-MIDAS-AR: factor augmented MIDAS-AR.

Table A.13
Marginal value of the statistical models, 2008.I–2011.III.

Frequency model	AR	BVAR	Pooling forecasts				Pooling information			
			BEQ-AR	QVAR	MIDAS-AR	MFVAR	DFM	F-VAR	F-MIDAS-AR	F-MFVAR
Euro area										
2Q ahead	–		–	–	–	–	0.99	–	–	0.99
1Q ahead	–	–	–	–	–	–	–	–		–
Nowcast	–	0.98	–	–	–	–	–	–		–
Backcast	–	–	–	–	–	–		–	–	–
Germany										
2Q ahead	–	–	–	–	–	–	–	–		–
1Q ahead	–	–	–	–	–	–	–	0.99	–	
Nowcast	–	–	–	–	–	–	–	–	–	
Backcast	–	–	–	–	–	–		–	–	–
France										
2Q ahead	–		–	–	–	–	–	–	–	–
1Q ahead	0.98	**0.95**	0.99	0.99	0.98	–	–	0.99	**0.96**	
Nowcast	–	0.99	–	–	–	–	–	–		–
Backcast	–	0.98	–	–	–	–		–	–	–
Italy										
2Q ahead	–		–	–	–	–	–	–	–	–
1Q ahead	–	–	–	–	–	–	–	–		0.98
Nowcast	–	–	–	–	–	–	–	–		–
Backcast	–	–	–	–	–	–		–	–	–
Spain										
2Q ahead	0.98	–	–	–	0.99	–	–	0.99		–
1Q ahead	**0.95**	0.98	–	–	**0.97**	–		0.98	–	–
Nowcast	**0.90**	**0.92**	–	–	**0.97**	–		**0.97**	–	–
Backcast	–		–	–	–	–	**0.81**	–	**0.95**	0.99
Netherlands										
2Q ahead	–		–	–	–	–	–	–	–	–
1Q ahead	–	–	–	–	–	–	–	–		–
Nowcast	–	–	0.99	0.99	–	–	0.99	–		0.98
Backcast	–	–	–	–	–	–		–	–	–

Notes: Grey cells indicate the models with the lowest RMSFEs. Figures in boldface indicate that the encompassing test is statistically significant at the 5% level.
AR: autoregressive model; BEQ-AR: bridge equation with an AR term; QVAR: quarterly vector autoregressive model; BVAR: Bayesian QVAR model; F-VAR: factor augmented QVAR model; DFM: dynamic factor model; MFVAR: mixed frequency vector autoregressive model; F-MFVAR: factor augmented MFVAR model; MIDAS-AR: mixed data sampling model with AR-term; F-MIDAS-AR: factor augmented MIDAS-AR.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.ijforecast.2015.05.008.

References

Ager, P., Kappler, M., & Osterloh, S. (2009). The accuracy and efficiency of the Consensus forecasts: a further application and extension of the pooled approach. *International Journal of Forecasting*, *25*(1), 167–181.
Baffigi, A., Golinelli, R., & Parigi, G. (2004). Bridge models to forecast the euro area GDP. *International Journal of Forecasting*, *20*(3), 447–460.
Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*(1), 191–221.
Bai, J., & Ng, S. (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics*, *25*(1), 52–60.
Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, *25*(1), 71–92.
Bańbura, M., Giannone, D., & Reichlin, L. (2011). Nowcasting. In M. P. Clements, & D. F. Hendry (Eds.), *Oxford handbook on economic forecasting* (pp. 63–90). Oxford University Press.
Bańbura, M., & Modugno, M. (2014). Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, *29*(1), 133–160.
Bańbura, M., & Rünstler, G. (2011). A look into the factor model black box: publication lags and the role of hard and soft data in forecasting GDP. *International Journal of Forecasting*, *27*(2), 333–346.
Batchelor, R. (2001). How useful are the forecasts of intergovernmental agencies? The IMF and OECD versus the Consensus. *Applied Economics*, *33*(2), 225–235.
Bernanke, B. S., & Boivin, J. (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics*, *50*(3), 525–546.
Bloor, C., & Matheson, T. (2011). Real-time conditional forecasts with Bayesian VARs: An application to New Zealand. *The North-American Journal of Economics and Finance*, *22*(1), 26–42.
Camacho, M., & Perez-Quiros, G. (2010). Introducing the euro-sting: short-term indicator of euro area growth. *Journal of Applied Econometrics*, *25*(4), 663–694.
Camacho, M., & Quiros, G. P. (2011). Spain-sting: Spain short term indicator of growth. *Manchester School*, *79*(1), 594–616.
Camba-Mendez, G., Kapetanios, G., Smith, R. J., & Weale, M. R. (2001). An automatic leading indicator of economic activity: forecasting GDP growth for European countries. *Econometrics Journal*, *4*(1), 37.
Carriero, A., Clark, T. E., & Marcellino, M. (2012). Real-time nowcasting with a Bayesian mixed frequency model with stochastic volatility. Working paper 1227, Federal Reserve Bank of Cleveland.
Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245–276.
Cheung, C., & Demers, F. (2007). Evaluating forecasts from factor models for Canadian GDP growth and core inflation. Working papers 07-8, Bank of Canada.
Clark, T. E., & McCracken, M. W. (2010). Averaging forecasts from VARs with uncertain instabilities. *Journal of Applied Econometrics*, *25*(1), 5–29.
Clements, M. P., & Galvão, A. B. (2008). Macroeconomic forecasting with mixed-frequency data: forecasting output growth in the United States. *Journal of Business & Economic Statistics*, *26*(4), 546–554.
D'Agostino, A., & Giannone, D. (2012). Comparing alternative predictors based on large-panel factor models. *Oxford Bulletin of Economics and Statistics*, *74*(2), 306–326.
den Reijer, A. (2013). Forecasting Dutch GDP and inflation using alternative factor model specifications based on large and small datasets. *Empirical Economics*, *44*(2), 435–453.
de Winter, J. (2011). Forecasting GDP growth in times of crisis: private sector forecasts versus statistical models. DNB Working papers 320, De Nederlandsche Bank.
Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263.
Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford University Press.
Foroni, C., & Marcellino, M. (2014). A comparison of mixed frequency approaches for nowcasting Euro area macroeconomic aggregates. *International Journal of Forecasting*, *30*(3), 554–568.
Ghysels, E., Sinko, A., & Valkanov, R. (2007). MIDAS regressions: further results and new directions. *Econometric Reviews*, *26*(1), 53–90.
Giannone, D., Lenza, M., & Primiceri, G. E. (2012). Prior selection for vector autoregressions. CEPR discussion papers 8755, C.E.P.R. Discussion Papers.
Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: the real-time informational content of macroeconomic data. *Journal of Monetary Economics*, *55*(4), 665–676.
Kitchen, J., & Monaco, R. (2003). Real-time forecasting in practice: The US treasury staff's real-time GDP forecast system. *Business Economics*, *38*(4), 10–28.
Kuzin, V., Marcellino, M., & Schumacher, C. (2011). MIDAS vs. mixed-frequency VAR: nowcasting GDP in the euro area. *International Journal of Forecasting*, *27*(2), 529–542.
Kuzin, V., Marcellino, M., & Schumacher, C. (2013). Pooling versus model selection for nowcasting GDP with many predictors: empirical evidence for six industrialized countries. *Journal of Applied Econometrics*, *28*(3), 392–411.
Lahiri, K., Isiklar, G., & Loungani, P. (2006). How quickly do forecasters incorporate news? Evidence from cross-country surveys. *Journal of Applied Econometrics*, *21*(6), 703–725.
Liebermann, J. (2012). *Real-time forecasting in a data-rich environment. Research Technical Papers 07/RT/12*. Central Bank of Ireland.
Loungani, P., & Rodriguez, J. (2008). Economic forecasts. *World Economics*, *9*(2), 1–12.
Lundquist, K., & Stekler, H. O. (2012). Interpreting the performance of business economists during the great recession. *Business Economics*, *47*(2), 148–154.
Marcellino, M., & Schumacher, C. (2010). Factor MIDAS for nowcasting and forecasting with ragged-edge data: a model comparison for German GDP. *Oxford Bulletin of Economics and Statistics*, *72*(4), 518–550.
Mariano, R. S., & Murasawa, Y. (2010). A coincident index, common factors, and monthly real GDP. *Oxford Bulletin of Economics and Statistics*, *72*(1), 27–46.
Meyler, A., & Rubene, I. (2009). Results of a special questionnaire for participants in the ECB survey of professional forecasters (SPF). MPRA paper 20751, University Library of Munich, Germany.
Rünstler, G., Barhoumi, K., Benk, S., Cristadoro, R., den Reijer, A., Jakaitiene, A., Jelonek, P., Rua, A., Ruth, K., & Nieuwenhuyze, C. V. (2009). Short-term forecasting of GDP using large datasets: a pseudo real-time forecast evaluation exercise. *Journal of Forecasting*, *28*(7), 595–611.
Rünstler, G., & Sédillot, F. (2003). Short-term estimates of euro area real GDP by means of monthly data. Working paper series 0276, European Central Bank.
Schneider, M., & Spitzer, M. (2004). Forecasting Austrian GDP using the generalized dynamic factor model. Working papers 89, Oesterreichische Nationalbank.
Schorfheide, F., & Song, D. (2013). Real-time forecasting with a mixed-frequency VAR. NBER working papers 19712, National Bureau of Economic Research, Inc.
Schumacher, C. (2014). MIDAS and bridge equations. DB discussion papers 26/2014, Deutsche Bundesbank.
Schumacher, C., & Breitung, J. (2008). Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. *International Journal of Forecasting*, *24*(3), 386–398.
Stekler, H. O. (1991). Macroeconomic forecast evaluation techniques. *International Journal of Forecasting*, *7*(3), 375–384.
Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, *23*(6), 405–430.
Stock, J. H., & Watson, M. W. (2011). Dynamic factor models. In *Oxford handbook of economic forecasting, Vol. 1* (pp. 35–59).
Timmermann, A. (2006). *Handbook of economic forecasting*: *Vol. 1. Forecast combinations* (pp. 135–196). Elsevier, (Chapter 4).
Wallis, K. F. (1986). Forecasting with an econometric model: The 'ragged edge' problem. *Journal of Forecasting*, *5*(1), 1–13.

W. Jos Jansen works as a policy advisor at the Financial and Economic Policy Department, Ministry of Finance, The Hague. His main research interests are business cycle synchronization, economic and financial integration, and short-term forecasting.

Xioawen Jin works as a reseacher at the Department of Economics at Ludwig Maximilians University, Munich. Her main research interest is short-term forecasting.

Jasper M. de Winter works as a reseacher and policy advisor at the Policy and Research Department of De Nederlandsche Bank, Amsterdam. His main research interest is short-term forecasting.