DNB Working Paper
No. 507 / March 2016

Improving model-based near-term GDP forecasts by subjective forecasts: A real-time exercise for the G7 countries

Jos Jansen and Jasper de Winter

DeNederlandscheBank
EUROSYSTEEM

Improving model-based near-term GDP forecasts by subjective forecasts: A real-time exercise for the G7 countries

Jos Jansen and Jasper de Winter *

* Views expressed are those of the authors and do not necessarily reflect official positions of De Nederlandsche Bank.

Working Paper No. 507

March 2016

De Nederlandsche Bank NV
P.O. Box 98
1000 AB AMSTERDAM
The Netherlands

Improving model-based near-term GDP forecasts by subjective forecasts: A real-time exercise for the G7 countries

Jos Jansen[a] and Jasper de Winter[b]

[a] *Financial and Economic Policy Department, Ministry of Finance.*
E-mail: w.j.jansen@minfin.nl
[b] *Economics and Research Division, De Nederlandsche Bank.*
E-mail: j.m.de.winter@dnb.nl

March 2016

Abstract

We investigate to what extent it is feasible to improve model-based near-term GDP forecasts by combining them with judgmental (quarterly) forecasts by professional analysts (Consensus survey) in a real-time setting. Our analysis covers the G7 countries over the years 1999-2013. We consider as combination schemes the weighted average and the linear combination. Incorporating subjective information delivers sizable gains in forecasting ability of statistical models for all countries except Japan in 1999-2013, even when subjective forecasts are somewhat dated. Accuracy gains are much more pronounced in the volatile period after 2008 due to a marked improvement in predictive power of Consensus forecasts. Since 2008, Consensus forecasts are superior at the moment of publication for most countries. For some countries Consensus forecasts can be enhanced by model-based forecasts in between the quarterly release dates of the Consensus survey, as the latter embody more recent monthly information.

Keywords: Forecast combination; encompassing test; nowcasting; factor models; judgment.
JEL classifications: C33, C53, E37.

1 Introduction

Policy makers and economic agents have to make decisions in real time on the basis of incomplete and inaccurate information on current economic conditions. For example, data on real GDP, which is the broadest measure of aggregate economic activity, are released on a quarterly basis with a substantial time lag (six weeks in many advanced countries); they are also subject to revisions. However, a wealth of statistical information that is directly and indirectly related to economic activity is nowadays available from public and private sources. Policy makers, firms and financial market participants may exploit this vast body of statistical information to form expectations on the current state of the economy and its near-term development. This requires solving the practical problem of handling a large-scale information set of potentially hundreds of time series that are observed at different frequencies and with different publication lags (the so-called ragged edge problem). The recent nowcasting literature has developed several statistical methodologies for generating near-term GDP forecasts based on large mixed-frequency data sets with ragged edges. Examples are bridge models, factor models, mixed-data sampling models (MIDAS), mixed-frequency vector-autoregressive (MFVAR) models and Bayesian VARs and mixed-frequency models.[1]

Apart from model-based predictions, policy makers and economic agents may also take advantage of published forecasts made by professional analysts. From a practical point of view, such forecasts are cheap and easy to use. Currently, several surveys on the economic outlook are available on a regular basis. The Federal Reserve Bank of Philadelphia and the European Central Bank (ECB) both maintain a regular Survey of Professional Forecasters. Moreover, the survey firm Consensus Economics publishes a well-known compilation of macroeconomic forecasts by professional forecasters for many countries. Model-based forecasts are the result of purely mechanical recipes using statistical data and do not incorporate subjective elements. By contrast, forecasts by professional analysts reflect much more information than statistical data, which are inevitably limited. For example, Meyler and Rubene (2009) report that the participants of the ECB Survey of Professional Forecasters consider forty percent of their short-term GDP forecasts to be judgment-based. Based on in-sample encompassing tests, Jansen et al. (2016) find in a pseudo-real time set-up that subjective predictions by private sector analysts often embody valuable information that sophisticated mechanical forecasting procedures fail to pick up. Liebermann (2014) presents a similar result for the US in the period 2000–2010 in a real-time setting.

The empirical evidence in Jansen et al. (2016) and Liebermann (2014) suggests that publicly available subjective forecasts offer the potential of enhancing real-time model-based GDP forecasts and thus a better assessment of the current state of the economy. The main purpose of this paper is to investigate whether predictions by analysts are actually able to improve GDP forecasts generated by purely statistical procedures in real time. We also pay attention to the reverse question whether model-based forecasts can be used to enhance subjective forecasts. Our procedure takes into account the information availability constraints facing practitioners when running forecasting models and forming expectations. As our benchmark, we use predictions

[1] See among others Baffigi et al. (2004), Stock and Watson (2011), Kuzin et al. (2011), Ghysels et al. (2007), Foroni and Marcellino (2014), Bańbura et al. (2010), Carriero et al. (2015) and Jansen et al. (2016).

produced by a dynamic factor model estimated on real-time data vintages using a rolling estimation window of 15 years. We employ the quarterly forecasts published by Consensus Economics as our measure of judgmental forecasts. Our analysis covers the G7 countries in the period 1999–2013, contrasting the experience in the volatile post-crisis period of 2008–2013 with that in the more tranquil period of 1999–2007.

The remainder of the paper is structured as follows. Section 2 describes the real-time data set and the Consensus forecasts, the benchmark forecasting model and the way predictions are generated in real time. Section 3 presents the empirical results and Section 4 concludes.

2 Data, benchmark model and forecast design

2.1 Data

Data on real GDP and monthly indicators are released with different publication lags and are possibly subject to revisions at a later stage, which at least for GDP may be sizable. Many papers on nowcasting employ a pseudo real-time design, which takes publication delays into account and applies recursive estimation, but disregards data revisions of GDP and monthly indicators, such as industrial production. Such an approach is unsuitable in our case for two reasons. First, our aim is to investigate whether it is in practice, hence in real time, possible to enhance mechanical forecasts by judgmental forecasts. Second, as noted by Croushore (2011), data revisions could affect the results of forecast evaluations and comparisons of different (statistical) approaches. This criticism is even more relevant in our case, which involves comparing model-based forecasts and forecasts by professional analysts. The expectations of analysts at a certain point in time necessarily reflect the then available information, including inaccurate initial estimates of GDP's recent past and that of key monthly indicators. This puts them at a disadvantage vis-a-vis statistical approaches in a pseudo-real time setting, where the latter can take data revisions on board for model estimation and projections. For this reason we have compiled real-time data sets. This means that the model-based forecasts used in our forecast evaluation exercise only incorporate information that was available at the time of forecasting.[2]

We have compiled real-time monthly data sets that consist of similar variables across countries and that cover the broad range of information that is readily available to economic agents. We have selected the variables in the spirit of Babura et al. (2013), who focus on "headline" macroeconomic variables that financial market participants and the media primarily pay attention to. Bańbura et al. (2011) and Bańbura and Modugno (2014) provide evidence that the marginal impact of disaggregated data on forecast accuracy is very small. Accordingly, the selected indicators refer to the aggregate level. Possibly available disaggregated information by sector, subcategory, region, etc. is not included. The data set for a country consists of three parts. The first part concerns information about the domestic economy, the second part refers to global variables related to global economic activity, and the third part contains key data on the two most important trade partners of the country.[3]

[2] Studies using real-time data include Schumacher and Breitung (2008), Camacho and Perez-Quiros (2010), Lahiri and Monokroussos (2013) and Liebermann (2014).

[3] We determine the two most important trade partners on the basis of the OECD trade-in-value-added database, which focuses on the value added contribution of (bilateral) exports. The most important trade partners are the US and UK for Canada, the US and France for Germany, the US and Germany for France, France and

The indicator variables that refer to the domestic economy fall into four categories. The first category is hard, quantitative information on production and expenditures, such as industrial production, car sales, retail sales, exports, imports and unemployment. The second category refers to consumer and producer prices. The third category contains financial variables, both quantities (monetary aggregates) and prices (interest rates, stock prices and exchange rates). These determine the 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. We also included a composite leading indicator compiled by the OECD. Following Golinelli and Parigi (2014), our global variables, which are common to all countries, include oil and commodity prices, semiconductor sales, the Baltic Freight Index, the Standard and Poors Exchange Volatility Index (VIX) and world trade. The key data on a country's two most important trade partners comprise three variables: imports, industrial production and the composite leading indicator compiled by the OECD. Finally, for all European countries we also included four closely-watched confidence indices from Germany (Ifo), France (INSEE), Italy (ISEA) and Belgium (BNB).

The statistical monthly information set reflects real-time public knowledge in the middle of the second week of the month. The number of monthly indicators varies from around 30 for Japan and Canada to around 36 for the other countries. All monthly indicator series start in January 1985, while the quarterly GDP series start in 1985.I. Table A.I in Appendix A.1 provides details on the exact composition of each country's statistical data set and the data sources. Monthly data are seasonally (and calendar effects) adjusted at the source, except for prices and financial variables. All monthly series are made stationary by taking three-month differences, log-differences (in case of trending data, such as industrial production) or double log-differences (in case of prices). Moreover, all variables are standardized by subtracting the mean and dividing by the standard deviation. This normalization is necessary to avoid overweighting of large variance series in the extraction of common factors.

The subjective quarterly GDP forecasts by professional analysts were collected from paper copies of the monthly publication *Consensus Forecasts*, published by Consensus Economics. Consensus Forecasts offers a survey 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, which have been analyzed in several papers (e.g. Ager et al., 2009; Batchelor, 2001; Loungani and Rodriguez, 2008; Lahiri et al., 2006). Once a quarter, this publication also provides averaged forecasts for quarterly GDP over a horizon of six quarters, starting with the nearest quarter for which no officially released figure is available. The number of respondents varies somewhat over time and across countries.[4] Fresh quarterly Consensus forecasts become available in the second week of the last month of

Germany for Italy, the US and China for Japan, the US and Germany for the UK, and Canada and Mexico for the US.

[4] The average number of participants is about 15 for Canada and Italy, about 20 for France and Japan and about 28 for Germany, the UK and the US. Consensus Forecasts publishes the simple average of the forecasts by all respondents, but no individual forecasts. This is not a serious limitation, as Genre et al. (2013) show that the potential gains of combining expert forecasts with alternative schemes are very limited. Moreover, a potential advantage is that analysts may be more likely to submit their true expectations, as this procedure guarantees their anonymity, reducing motives for possible strategic behaviour (Lamont, 2002; Laster et al., 1999).

the quarter. The survey date (deadline for respondents) is typically the second Monday of the third month of a quarter; publication is usually 3 days later on Thursday. The timing of the survey is therefore in line with the riming of the monthly data vintages we collected. For our information set this means that Consensus forecasts are not updated in the first and second months in a quarter, while monthly indicators are updated every month. Moreover, at the time analysts form their expectations they have official information on GDP growth in the preceding quarter. Quarterly Consensus forecasts for the G7 countries are available from the early 1990s onwards.

2.2 Benchmark model: Dynamic Factor Model

The model used to generate model-based forecasts is a dynamic factor model (DFM). Dynamic factor models summarize the information of the data set in a limited number of factors, whose dynamic behavior is specified as a vector-autoregressive process. A key feature of this approach is the use of the Kalman filter, which allows for an efficient handling of the unbalancedness of the data set and the different frequencies of the data. The Kalman filter replaces any missing monthly indicator observations with optimal predictions and also generates estimates of unobserved monthly real GDP subject to a temporal aggregation constraint for the quarterly observation. Jansen et al. (2016) found in their comparative study that the DFM was the best statistical procedure overall, in particular for nowcasting and backcasting. Dynamic factor models have been applied to many countries, generally delivering relatively accurate macroeconomic forecasts.[5]

In this paper we employ 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

$$x_m = \Lambda f_m + \xi_m, \qquad \xi_m \sim N(0, \Sigma_\xi) \tag{1}$$

which relates the n monthly indicators $x_m = (x_{1,m}, \ldots, x_{n,m})'$ to r monthly static factors $f_m = (f_{1,m}, \ldots, f_{r,m})'$ via an $n \times r$ matrix of factor loadings Λ and an idiosyncratic component $\xi_m = (\xi_{1,m}, \ldots, \xi_{n,m})'$, where $r << n$. m is a monthly time index. As explained above, the monthly indicators $x_{i,m}$ are normalized three-month growth rates or differences. 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:

$$f_m = \sum_{s=1}^{p} A_s f_{m-s} + \zeta_m, , \qquad \zeta_m \sim (0, \Sigma_\zeta) \tag{2}$$

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:

$$\zeta_m = B\eta_m, \qquad \eta_m \sim N(0, I_q) \tag{3}$$

[5] Examples are Giannone et al. (2008) and Liebermann (2014) for the United States; Bańbura et al. (2011), Camacho and Perez-Quiros (2010), Rünstler et al. (2009) and Bańbura and Modugno (2014) 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) and de Winter (2011) for the Netherlands.

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

$$y_m = \beta' f_m + \varepsilon_m, \qquad \varepsilon_m \sim N(0, \sigma_\varepsilon^2) \tag{4}$$

where y_m denotes the (unobserved) three-month growth rate of monthly real GDP, i.e. the growth rate vis-à-vis the same month of the previous quarter. Quarterly real GDP growth in quarter t, y_t^Q, is assigned to month $3t$ on the monthly time scale. The relation between the quarterly and monthly GDP growth rates is given by $y_t^Q = \frac{1}{3}(y_{3t} + y_{3t-1} + y_{3t-2})$.

The model is estimated in four steps. In the first step we obtain the factors loadings Λ and the estimated static factors $\hat{f}_m$. In the second step we estimate the coefficient matrices A_s in eq. (2) and β in eq. (4) 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 form and use the Kalman filter and smoother to re-estimate the estimated factors ($\hat{f}_m$) and monthly GDP growth.[6] Finally, we calculate forecasts of quarterly GDP growth by applying eq. (4) to forecasts of monthly factors generated by eq. (2), and then aggregating to quarterly values.

To estimate the model we need to specify the number of static and dynamic common factors, denoted by r and q respectively. We set the largest possible value of r at 6, based on the scree test of Cattell (1966). Moreover, $q \leq r$ by definition. In view of potential misspecification and instabilities we follow Kuzin et al. (2013) and Jansen et al. (2016), 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 number of static and dynamic factors and the number of lags p in eq. (2), with $p \leq 2$. The total number of model specifications is $p(r+1)r/2 = 42$. This strategy avoids any hindsight bias.[7]

2.3 Forecast design

The forecast design entails the construction of six consecutive forecasts for real GDP growth for each quarter in the period 1999.I–2013.IV. Table I explains the timing of the forecasting exercise, taking the forecast for a second quarter as an example. We make the first forecast in mid-March, just after the release of a new Consensus survey. We subsequently produce a monthly forecast for the next five months. The last forecast is made in mid-August. Following the conventional terminology, *forecasts* (F) refer to predictions made prior to the start of the quarter of interest, *nowcasts* (N) refer to current quarter forecasts and *backcasts* (B) refer to forecasts for the preceding quarter, as long as official (target) GDP figures are not yet available. We refer to these forecasts (or horizons) as F3, N1, N2, N3, B1 and B2, respectively. A Consensus survey consists of new F3 and N3 forecasts. There is no genuine Consensus backcast, since analysts have official information on GDP growth in the preceding quarter when they form their expectations.

[6] The state-space setup of the dynamic factor model is outlined in Appendix A.2. See Bańbura and Rünstler (2011) and Stock and Watson (2011) for a more detailed description of the dynamic factor model and the estimation procedure. See Durbin and Koopman (2012) for a comprehensive treatment of state space models and the use of the Kalman filter and smoother.

[7] A different approach is to choose the number of factors r and q on the basis of in-sample criteria, as described in Bai and Ng (2002, 2007). Bańbura and Rünstler (2011) and Jansen et al. (2016) report that these criteria tend to indicate a relatively large number of factors, leading to volatile and less accurate forecasts.

Table I: Timing of forecast exercise for second quarter GDP growth

Nr.	Forecast type	Month	DFM	Consensus
F3	Forecast	3	March	new
N1	Nowcast	1	April	1 month old
N2		2	May	2 months old
N3		3	June	new
B1	Backcast	1	July	1 month old
B2		2	August	2 months old

In the first and second months of a quarter we use as Consensus forecast the (non-updated) forecast dating from the third month of the previous quarter. For example, the Consensus backcasts made for quarter t are equal to the (non-updated) Consensus nowcast published in the last month of quarter t. We do not analyze forecasts further ahead than one quarter, as earlier research shows that the forecastability of GDP growth is very low beyond this horizon. This holds for statistical models as well as professional forecasters; see for instance Stark (2010) and Jansen et al. (2016). Estimation of the DFM and the subsequent calculation of forecasts happens recursively on the basis of the most recent 15 years of data on monthly indicators and quarterly real GDP. The data set is real-time; all data used for estimation and prediction were actually available at the time of estimation.

An important issue in a forecast evaluation exercise is how to measure the realized outcome, "actual GDP". The latest-available GDP data represent the current state-of-thinking about the "true" history of real GDP. However, these data partly reflect benchmark revisions, which both analysts and forecasting models cannot foresee. Using the latest-available data thus introduces noise in a comparison of forecasting performance between professional forecasters and mechanical statistical procedures, making this data concept unsuitable for our purpose. Statistical agencies publish a sequence of preliminary GDP estimates, with the first release (flash estimate) receiving by far the most attention in the financial press and the media. It is therefore reasonable to assume that analysts are primarily focused on predicting the flash or early estimates rather than a number that will be released far into the future. Our real-time measure of actual GDP is determined by the latest officially released information for the preceding quarter that analysts know at the moment when they formulate predictions for the current and subsequent quarters. For five countries (and the UK before 2008) this is the flash release, which is published six weeks after the reference quarter has ended. For the US and the UK (as from 2008) it is the first revision to the flash, as in these cases the flash is already released four weeks after the quarter has ended. This choice allows us to include two backcasts in the analysis for all countries.

3 Empirical results

3.1 Forecasting performance dynamic factor model and subjective forecasts

Table II presents data on the forecast performance of the dynamic factor model for the full sample period 1999.I–2013.IV, the relatively tranquil pre-crisis period 1999.I–2007.IV and the volatile post-crisis period 2008.I–2013.IV. Throughout the paper we will present the empirical results in this way on account of the large difference in the level of volatility before and after 2008. We measure forecast performance by the root mean square forecast error (RMSFE). As usual in the literature, we use the random walk (RW) with drift as a purely statistical benchmark model. The first column of the table reports the RMSFE of the random walk to gauge the overall level of volatility.[8] The other columns report the DFM's RMSFE relative to that of the RW benchmark in order to improve the comparability of the results across countries and horizons. Bold faced entries indicate that the RMSFE differs by more than $\pm$ 10% from the RW's RMSFE. The 10 per cent threshold is meant as a rough informal assessment of the economic importance of the gain in forecasting accuracy from using auxiliary monthly information. In addition, we performed (one-sided) Diebold and Mariano (1995) (DM) tests as a formal test of statistical significance at the conventional levels (denoted by asterisks).[9] Non-starred, normal-type entries thus indicate models that are equal in terms of forecasting accuracy, both statistically and economically. We will follow the same two-way approach to statistical/economic significance in all tables that feature RMSFEs in this paper.

Table II demonstrates that incorporating monthly information pays off in terms of forecasting accuracy. In the period 1999–2013, virtually all relative RMSFEs are smaller than one and tend to decline if the horizon shortens and more monthly information has been absorbed. However, there is a marked contrast between the pre-crisis and post-crisis periods. The scope for exploiting monthly information for prediction is significantly smaller in the tranquil pre-crisis period, when real GDP growth was more predictable in general, as indicated by the much lower RMSFE of the RW benchmark compared to that of the post-crisis period. The DFM performs, on average, around 11% better than the RW for nowcasts and around 18% for backcasts before 2008; after 2008 the gains are twice as large: 20% and 39%, respectively. Accuracy gains vis-à-vis the random walk are less than 10% for all horizons for Italy and Japan before 2008. Moreover, the DFM's predictive ability develops unevenly over the six months of the forecasting exercise for several countries, Germany and the UK in particular, before the financial crisis. By contrast, the DFM delivers improvements in forecasting accuracy that are large and steadily increasing with the forecast horizon for all countries in the volatile post-crisis period 2008–2013. The dynamic factor model thus helps forecasters to compensate for the generalized deterioration in predictability that characterize volatile times, in particular for nowcasting and backcasting. The evidence also suggests that the DFM's relative strength is to exploit information pertaining to the quarter under consideration and to improve the assessment of the current state of the economy. The findings in Table II are broadly consistent with the empirical literature on dynamic factor models referred to in Section 2.2.

In Table A.II in Appendix A.3, we look into the impact of two features of our forecasting design, namely (i) its real-time nature; and (ii) the use of initial estimates as the measure of target GDP. The left-hand side of the table reports the DFM's forecasting performance obtained

[8] To save space we only report the RW's RMSFE for the third nowcast (N3), as it hardly varies with the horizon. See the online appendix for a full set of results for the RW model. The drift parameter is recursively estimated in real time on the most recent 15 years of GDP data. The online appendix also reports a full set of results for both the DFM and RW model if the estimation window is 10 years or recursively expanding.

[9] The DM test broadly paints the same picture as the informal 10% improvement criterion, although the two do not always match. In some cases, large differences in accuracy are not statistically significant, whilst the reverse also happens. This suggest that statistical significance and economic importance are different concepts. Moreover, the power of the DM test may be low due to the small number of observations.

Table II: Forecasting performance dynamic factor model, 1999.I-2013.IV

	RW	B2	B1	N3	N2	N1	F3
	RMSFE	Relative RMSFE DFM vs. Random Walk					
				Full sample 1999.I-2013.IV			
Canada	*0.57*	**0.61****	**0.66****	**0.69***	**0.76***	**0.81***	**0.86**
France	*0.46*	**0.67***	**0.77***	**0.81***	**0.85**	**0.87**	0.93
Germany	*0.82*	**0.72***	**0.76***	**0.84**	**0.88**	**0.88***	0.92*
Italy	*0.64*	**0.62***	**0.71***	**0.76***	**0.81***	**0.80***	**0.88***
Japan	*1.08*	**0.67***	**0.72***	**0.78***	0.93	1.04	1.06
UK	*0.63*	**0.76***	**0.78***	**0.83***	0.90	0.92	0.91*
US	*0.61*	**0.57***	**0.62***	**0.70***	**0.80**	**0.86**	**0.89**
				Pre-crisis period 1999.I-2007.IV			
Canada	*0.39*	**0.69*****	**0.73*****	**0.82****	0.94	0.92**	0.94
France	*0.33*	**0.79****	**0.88***	**0.83*****	**0.80*****	**0.83*****	0.95
Germany	*0.42*	**0.83**	**0.87**	0.99	0.97	0.90	**0.87****
Italy	*0.39*	0.91*	0.95	0.97	0.97	0.92*	0.95
Japan	*0.86*	0.92	0.94	0.94	0.97	0.97	1.01
UK	*0.25*	**0.77****	**0.76****	**0.79***	**0.78***	**0.66****	**0.67*****
US	*0.49*	**0.71*****	**0.75*****	**0.82*****	0.91*	0.96	0.98
				Post-crisis period 2008.I-2013.IV			
Canada	*0.77*	**0.57***	**0.62***	**0.63***	**0.68***	**0.76**	**0.82**
France	*0.60*	**0.62***	**0.72**	**0.80**	**0.87**	**0.89**	0.93
Germany	*1.19*	**0.70***	**0.73***	**0.81**	**0.86**	**0.88**	0.92
Italy	*0.89*	**0.50***	**0.63***	**0.70***	**0.75***	**0.76***	**0.86***
Japan	*1.35*	**0.45***	**0.53***	**0.67**	0.90	1.08	1.09*
UK	*0.94*	**0.76***	**0.78***	**0.84**	0.91	0.94	0.93
US	*0.75*	**0.46**	**0.51**	**0.61**	**0.72**	**0.78**	**0.83**

Notes: Entries refer to the RMSFE for RW for the third nowcast (in italics) and the RMSFE relative to random walk's RMSFE in all other cases. Figures in boldface indicate a deviation by more than ± 10% from 1. Starred entries (*,**,***) denote that the one-sided Diebold-Mariano test is significant at the 10%, 5% or 1% level. The alternative hypothesis of the DM test is MFSE > 1 or MFSE < 1, depending on whether the RMSFE is greater or smaller than 1, respectively.

in the corresponding pseudo-real-time set-up, which also implies that prediction errors have been computed using the last vintage for GDP. A comparison of the RMSFEs of the random walk model reveals that forecasting errors tend to be larger for the pseudo-real-time design on account of the extra noise generated by the data revisions. At the same time, the DFM's relative RMSFE tend to decline more steeply in the pseudo-real-time procedure. This phenomenon is especially pronounced for nowcasts in the pre-crisis period. A pseudo-real-time set-up may thus overestimate the potential gains from exploiting monthly disaggregated information when nowcasting GDP, especially in stable environments. The right-hand side of Table A.II presents the outcomes for the real-time procedure, when RMSFEs are calculated using the final GDP vintage as measure of actual GDP. The first column of this half now reports the RMSFE of the first estimate as a predictor of the final vintage. These results suggest that revisions are a substantial source of unpredictability as the RMSFE of the first estimate varies from 0.2 and 0.4 for six countries, while for Japan it is even larger. Moreover, we find that using the final GDP vintage as the target rather than the first estimate tends to lead to a more optimistic view of the potential gains from using monthly information before the crisis, but this does not hold after the crisis.

Turning to the Consensus forecasts, panel A of Table III presents the RMSFE of the Consensus forecasts, relative to the random walk. Note that only the entries for the third one-quarter ahead forecast and the third nowcast (columns F3 and N3) are based on a fresh Consensus forecast. The other entries refer to Consensus forecasts that are one or two months old. In the 1999–2013, Consensus forecasts outperform the random walk model for all countries and all horizons, often by large margins, with the exception of the F3 forecast in the Japanese case. Moreover, Consensus forecast display clear learning behaviour: the relative RMSFEs of consecutive fresh quarterly Consensus forecasts (N3 versus F3) decline steeply in most cases, varying from 12% for the UK to 31% for Canada.[10] Again, forecast performance differs markedly between the pre-crisis and post-crisis periods. Before 2008, the Consensus F3 forecast is mostly unable to beat the random walk. It does 10% better in the case of the UK, but does 18% and 9% worse in the cases of Canada and Japan, respectively. The Consensus third nowcast (column N3) has a better record, reducing forecasting inaccuracy markedly for most countries compared to the RW. After 2008, and similar to the experience with the DFM, Consensus forecasts perform much stronger for almost all countries and all horizons. The RMSFE of the third nowcast is 29 to 65% smaller than the RW RMSFE.The most remarkable case concerns Canada: Consensus forecasts predict badly before 2008, but very well after 2008 (and especially during the crisis episode in 2008–2009). Comparing the 1999–2007 and 2008–2013 episodes, Consensus forecasts show only modest and uneven improvements for Germany and Japan.

To gain further insights into the relative strenghts of strictly model-based and judgmental predictions, panel B of Table III presents the RMSFE of the Consensus forecasts relative to that of the DFM. Newly released Consensus nowcasts (column N3) do better than the DFM for almost all countries in the evaluation period 1999–2013, and often by a significant margin.

[10] The generally small changes in relative RMSFEs in the months without a Consensus survey (columns B2, B1, N2 and N1) are due to revisions of GDP data, which may change the RW forecasts, and the fact that the Consensus survey was released one month earlier than normal in 2001.IV.

Table III: Forecasting performance Consensus forecasts, 1999.I-2013.IV

	B2	B1	N3	N2	N1	F3	B2	B1	N3	N2	N1	F3
	A. Relative RMSFE CF versus RW						B. Relative RMSFE CF versus DFM					
	Full sample 1999.i-2013.iv						Full sample 1999.i-2013.iv					
Canada	**0.56***	**0.57***	**0.57***	**0.84**	**0.82**	**0.82**	0.92	**0.86**	**0.82**	**1.11**	1.02	0.96
France	**0.68****	**0.68****	**0.68****	**0.84****	**0.86***	**0.86***	1.01*	**0.88****	**0.84*****	0.98**	0.98*	0.92*
Germany	**0.68****	**0.68****	**0.68****	**0.89***	0.90	0.90	0.94	0.90**	**0.81*****	1.01	1.02	0.98
Italy	**0.69****	**0.69****	**0.69****	**0.81****	**0.84***	**0.84***	**1.12**	0.97	0.90	1.01	1.05	0.95
Japan	**0.81***	**0.81***	**0.81***	0.98	0.98	0.98	**1.22***	**1.14**	1.04	1.06	0.94	0.93
UK	**0.57***	**0.57***	**0.57***	**0.64***	**0.65***	**0.65***	**0.75***	**0.73***	**0.69***	**0.71***	**0.71***	**0.71***
US	**0.57***	**0.57****	**0.57****	**0.82**	**0.80**	**0.80**	1.00	0.92	**0.81*****	1.03	0.94	0.90*
	Pre-crisis period 1999.I-2007.IV						Pre-crisis period 1999.I-2007.IV					
Canada	**0.89****	0.91*	0.91*	**1.23****	**1.18****	**1.18****	**1.29****	**1.24***	**1.11**	**1.31****	**1.29****	**1.26****
France	**0.87***	**0.87***	**0.87***	0.94	1.01	1.00	**1.11**	0.99*	1.05	**1.18**	**1.21**	1.06
Germany	**0.66****	**0.68****	**0.68*****	0.94	1.00	1.00	**0.80**	**0.78**	**0.69****	0.97	**1.12**	**1.15**
Italy	0.92	0.93	0.92	0.98	1.09**	1.09**	1.01	0.98	0.96	1.01	**1.19***	**1.15***
Japan	0.96	0.96	0.97	0.94	0.96	0.96	1.04	1.03	1.03	0.98	0.99	0.96
UK	**0.81***	**0.81***	**0.80***	**0.85**	**0.89**	0.90	1.05	1.06	1.01	1.09	**1.35*****	**1.35*****
US	**0.74*****	**0.73*****	**0.73*****	1.00	0.96	0.96	1.04	0.98	0.90**	1.10	1.00	0.98
	Post-crisis period 2008.I-2013.IV						Post-crisis period 2008.I-2013.IV					
Canada	**0.35***	**0.35***	**0.35***	**0.63***	**0.63***	**0.63***	**0.61*****	**0.56****	**0.55****	0.93	**0.83*****	**0.76****
France	**0.58***	**0.58***	**0.58***	**0.78****	**0.78****	**0.78****	0.94***	**0.81*****	**0.72*****	0.90***	**0.88****	**0.84*****
Germany	**0.68****	**0.68****	**0.68****	**0.88***	**0.88***	**0.88***	0.97	0.93*	**0.84*****	1.02	1.00	0.95
Italy	**0.61*****	**0.61*****	**0.61*****	**0.76*****	**0.76*****	**0.76*****	**1.21**	0.96	**0.87**	1.00	0.99	**0.88****
Japan	**0.71***	**0.71***	**0.71***	1.00	0.99	0.99	**1.59**	**1.32**	1.06	**1.11**	0.92	0.91
UK	**0.54***	**0.54***	**0.54***	**0.61****	**0.61****	**0.62****	**0.71****	**0.69****	**0.65****	**0.67****	**0.65****	**0.66****
US	**0.43**	**0.43**	**0.43**	**0.68**	**0.68**	**0.68**	0.93	**0.84***	**0.70*****	0.95	**0.88***	**0.82***

Notes: Figures in boldface indicate a deviation by more than ± 10% from 1. Starred entries (*,**,***) denote that the one-sided Diebold-Mariano test is significant at the 10%, 5% or 1% level. The alternative hypothesis of the DM test is MFSE > 1 or MFSE < 1, depending on whether the RMSFE is greater or smaller than 1, respectively.

The advantage is much less pronounced for fresh one quarter ahead forecasts (column F3), which clearly beat the DFM for only the UK and the US. Looking at the results across subperiods, the relatively positive score of the Consensus F3 and N3 forecasts is largely driven by their performance after the financial crisis. In the stable pre-crisis period, Consensus forecasts basically do worse or at most marginally better than the DFM across the board; the N3 nowcast for Germany is the one clear favourable exception. However, new Consensus forecasts (N3 and F3) have the edge over the DFM after the crisis in all cases but Japan. The relative RMSFE versus the DFM in general falls between horizons F3 and N3, which suggests that the value added of subjective insights may be greater when analysts know at least some hard and soft data on the quarter of interest itself. In between the quarterly release dates, the performance of Consensus forecasts versus DFM forecasts deteriorates, as the latter are updated. This catching-up by the DFM is stronger after 2008. However, at the B1 horizon, a two-month old Consensus nowcast still outperforms the DFM by more than 10% for the UK over the sample period, while this holds for Canada and Germany in one of the subperiods.

Whether judgmental forecasts by analysts embody valuable additional information can be more formally investigated by running encompassing tests versus the DFM forecasts. Even if Consensus forecasts are a poor predictor on their own, they may still possess a positive marginal value provided they are able to pick up specific useful information. In that case taking a combination of the Consensus and DFM forecasts may improve forecast accuracy. In the empirical literature encompassing tests may take on slightly different forms. In this paper we employ two versions of the encompassing test, which feature either a pure weighted average or an unconstrained linear combination of both forecasts. The respective test regressions are:

$$y^Q_{t+h|t} = \beta \hat{y}^Q_{\mathrm{CF}(t+h|t)} + (1-\beta)\hat{y}^Q_{\mathrm{DFM}(t+h|t)} + \varepsilon_t \tag{5}$$

$$y^Q_{t+h|t} = \alpha + \beta \hat{y}^Q_{\mathrm{CF}(t+h|t)} + \gamma \hat{y}^Q_{\mathrm{DFM}(t+h|t)} + \varepsilon_t \tag{6}$$

where y^Q_t is GDP growth in quarter t, $\hat{y}^Q_{\mathrm{CF}(t+h|t)}$ and $\hat{y}^Q_{\mathrm{DFM}(t+h|t)}$ are the predictions for quarter $t+h$ on time t by the Consensus survey and the DFM respectively. These equations can be seen as extreme forms of a linear combination. Eq. (6) is fully unconstrained, while eq. (5) is the most constrained version.[11] Eq. (5) goes back to Bates and Granger (1969) and has been applied by Stekler (1991) and Jansen et al. (2016). Eq. (6) was proposed by Granger and Ramanathan (1984) and has been used by Fair and Shiller (1990), Liebermann (2014) and Hubert (2014). The main advantage of eq. (6) is its ability to neutralize any biases in the underlying forecasts. Negative estimates of β or γ are theoretically possible in case of biases (Timmermann, 2006). Its disadvantage is that it may deliver imprecise estimates in small samples if the underlying forecasts are (highly) correlated. By construction, eq. (6) will obtain a better fit of observed GDP than eq. (5) for a given estimation sample. Consensus forecasts contain additional information in relation to the dynamic factor model if $\beta > 0$ in eq. (5) or $\beta \neq 0$ in eq. (6). Note that the encompassing test is an in-sample backward-looking test, which

[11] Eq. (5) imposes the following restrictions on eq. (6): $\alpha = 0$, $0 \leq \beta \leq 1$, $0 \leq \gamma \leq 1$ and $\beta + \gamma = 1$. Empirical studies have also applied intermediate forms of the encompassing test. For example, Rünstler et al. (2009) impose $\beta + \gamma = 1$, but neither $0 \leq \beta \leq 1$ nor $0 \leq \gamma \leq 1$. There is a large literature on the combination of forecasts; see Timmermann (2006) for an overview.

is just meant to signal the potential of benefitting from combining model-based forecasts with judgmental forecasts.

Table IV reports the estimated weight of Consensus forecasts in eq. (5).[12] We estimate β and its standard error on the interval [0,1] by Maximum Likelihood (ML) and perform a one-sided (asymptotically valid) tests of the hypotheses $\beta = 0$ and $\beta = 1$. The main message of Table IV is that both Consensus forecasts and DFM forecasts contain useful information. The point estimates lie mostly between 0 and 1; corner solutions are rare, except for Canada and the UK. The relative value added of Consensus forecasts differs across time periods. In the pre-crisis period the estimated weight of Consensus forecasts is typically lower than 0.5, and often not statistically significant for F3 forecasts and early nowcasts. After the crisis, they are often greater than 0.5 for all horizons. For the UK the estimated weight of the Consensus Forecasts even equals one for all horizons, suggesting that DFM forecasts do not possess extra information compared to forecasts by professional analysts; the same holds for the backcasts for Canada. For Japan, the added value of the Consensus nowcast is low. We also find that Consensus forecasts that are one or two months old often still offer the potential of improving DFM forecasts that incorporate more recent monthly information. This finding is another piece of evidence that analysts' forecasts contain information that is fundamentally different from the information that statistical models are able to pick up.

3.2 Enhancing model-based forecasts in real time

Our preliminary analysis suggests that there is room for improving mechanical model-based nowcasts and backcasts, by combining them with judgmental predictions by professional forecasters. This section investigates what benefits can be realistically expected from such a strategy in practice by simulating the forecasting procedure on the basis of real time data. Our procedure consists of two steps. In the first step we obtain the DFM forecast in real time, in the second step we determine the combination formula and compute the forecast combination. We use both the weighted average and the linear combination as combination schemes, as it is difficult to choose between them on theoretical or methodological grounds. As argued by Clemen (1986), which of the two methods of combining forecasts offers the best out-of-sample performance in real time is an empirical issue. Although eq. (6) is able to remove the effect of possible biases in the DFM and Consensus forecasts on the combined forecast, the simpler eq. (5) may still prove, despite possible bias, to be a more efficient combination rule. The combination schemes (5) and (6) are re-estimated every month according to the schedule in Table I.[13] This requires a choice on the length of the estimation period. As it is difficult to motivate a specific number on a priori grounds, we estimate the prediction rules using a moving window of 4–8 years, and then average the resulting five predictions.[14]

[12] Table A.V in Appendix A.3 reports the results of eq. (6). These results display the same pattern as the ones for eq. (5). As expected, in some cases the estimates of β or γ are larger than one or less than zero.

[13] In preliminary calculations we found that the linear combination scheme occasionally produced implausible forecasts. To deal with this problem we have set a lower and upper bound to the forecast. The lower bound is the minimum of the DFM and Consensus forecasts minus 0.3 percentage points; the upper bound is the maximum of the DFM and Consensus forecasts plus 0.3 percentage points.

[14] As a robustness check, we have investigated the effect on forecasting performance of employing shorter or longer estimation windows. We found this effect to be minor. See the online appendix for alternative versions of Table V for moving windows 3–6 years and 5–10 years and a recursively expanding estimation window.

Table IV: Encompassing test: weighted average of DFM and Consensus forecasts, 1999.I-2013.IV

	B2	B1	N3	N2	N1	F3
	Weight(β) of Consensus Forecast					
	Full sample 1999.I-2013.IV					
Canada	**0.66***	**0.74***	0.84*	**0.29***	**0.45***	**0.63***
France	**0.48***	**0.64***	**0.70***	**0.52***	**0.52***	**0.65***
Germany	0.67*	0.77*	1.00*	**0.48***	**0.42***	0.59*
Italy	**0.35***	**0.56***	0.72*	**0.49***	**0.35***	0.65*
Japan	**0.18**	**0.26**	**0.39***	**0.35***	0.68*	0.81*
UK	1.00*	1.00*	1.00*	1.00*	1.00*	1.00*
US	**0.51***	0.72*	1.00*	**0.42***	0.73*	0.89*
	Pre-crisis period 1999.I-2007.IV					
Canada	**0.00**	**0.09**	**0.28**	**0.02**	**0.06**	**0.01**
France	**0.35***	**0.52***	**0.43***	**0.26**	**0.19**	**0.35**
Germany	0.82*	0.80*	0.95*	**0.55***	**0.25**	**0.12**
Italy	**0.46***	0.56	0.63*	**0.45**	**0.00**	**0.00**
Japan	**0.41***	**0.45**	**0.43***	**0.55***	**0.52***	0.61*
UK	**0.47***	**0.46***	**0.49***	**0.44***	**0.27***	**0.23***
US	**0.32**	0.67	1.00*	**0.00**	0.52	0.72
	Post-crisis period 2008.I-2013.IV					
Canada	1.00*	1.00*	1.00*	0.67*	0.95*	1.00*
France	**0.55***	**0.70***	0.82*	**0.61***	**0.65***	0.76*
Germany	0.59	0.76	1.00*	0.44	0.49	0.78
Italy	**0.32***	**0.56***	0.76*	**0.50***	0.52	0.89*
Japan	**0.00**	**0.07**	**0.32**	**0.18**	0.82*	1.00*
UK	1.00*	1.00*	1.00*	1.00*	1.00*	1.00*
US	**0.58***	0.73*	0.97*	0.59*	0.78*	0.92*

Notes: Entries refer to the Maximum Likelihood estimate of the weight of the Consensus forecast β on the interval $[0, 1]$. The weight of the DFM forecast is $1 - \beta$. Starred entries (*) denote that the estimate is statistically different from 0 at the 5% level. Bold-type entries denote that the estimate is statistically different from 1 at the 5% level.

Table V reports the RMSFEs of both forecast combinations, relative to RMSFE of the DFM. The main conclusion of our exercise is that both combination methods are suitable to enhance DFM forecasts in economically meaningful ways. Japan is the only case for which combining forecasts does not work, which is consistent with the poor performance of the Consensus forecast vis-à-vis the DFM (see panel B of Table III). The scope for improvement is in general considerably greater in the volatile post-crisis period than in the tranquil times before 2008. In the post-crisis period gains in prediction accuracy of 20 to 35% are feasible for at least some horizons for many countries. Gains are on average more limited (often less than 10%) before 2008, but still reach 33% for Germany and 16% for the UK for the third month nowcast using the weighted average scheme (30% and 9% for the linear combination scheme). After 2008, the forecasting accuracy of the combined forecast is better than the DFM forecast in the majority of countries and forecasting horizons. The greatest advantage of combining DFM and Consensus forecasts generally occurs in the third month of the quarter when just-released Consensus forecasts are available. Although the relative RMSFE of the combined forecast versus the DFM tends to increase in the subsequent two months, it remains below one in most cases. This is evidence that Consensus forecasts incorporates information that goes beyond the information represented by the statistical information set and that this type of information is still valuable for predicting even if it is somewhat dated. Moreover, enhancing model-based forecasts with subjective forecasts may offer some insurance against a weak performance of mechanical models. The experience of the UK before 2008 is a case in point. As Table II shows, the DFM loses predictive power between horizons F3 and N3. This loss is roughly neutralized by utilizing additional, subjective information.

Comparing the two combination methods, our results indicate that the weighted average scheme tends to work better than the linear combination scheme before 2008, although the relative difference in RMSFE typically does not exceed 10%. After 2008, the linear combination scheme appears to have a slight edge on average. It performs extremely well in case of all of Italy's predictions and Germany's backcasts, but does a generally poor job at all horizons for the US. The empirical evidence thus points to quite some country heterogeneity in the relative performance of both schemes, which may even differ across horizons, as the Canadian and British experience illustrates. Looking at the evidence over the whole sample period across all forecasting horizons, the weighted average scheme performs better for France, Japan and the US, the linear combination scheme is better for Germany and Italy, while there is no clear winner for Canada and the UK. Another pattern in Table V is that the linear combination scheme is riskier than the weighted average scheme in the sense that combining forecasts may actually hurt forecasting accuracy appreciably; see the US results. Our findings thus suggest that the linear combination scheme seems to be less robust to overfitting in small samples, especially in stable macroeconomic environments.

As combining forecasts is a symmetric operation, it is also interesting to look at the marginal value of the DFM forecast. Table A.III offers this alternative perspective, presenting the RMSFE of both forecast combinations in terms of the Consensus forecast. We obtain mostly modest gains in accuracy by this measure in most cases in 1999–2013. The linear combination scheme works well for Canada and Italy, but not for the US.

Table V: Forecasting performance forecast combination schemes, 1999.I-2013.IV

	B2	B1	N3	N2	N1	F3	B2	B1	N3	N2	N1	F3
	Relative RMSFE combination DFM and CF versus DFM											
	A. Weighted average scheme						B. Linear combination scheme					
	Full sample 1999.I-2013.IV						Full sample 1999.I-2013.IV					
Canada	**0.82***	**0.78***	**0.77****	0.98	0.94*	0.90*	**0.83***	**0.80***	**0.80***	0.92	0.90*	**0.86***
France	0.91**	**0.86*****	**0.84*****	0.90***	0.91***	**0.89****	0.95**	**0.89****	**0.87*****	0.98***	0.98***	0.91*
Germany	0.96	0.91**	**0.81*****	1.01	1.00	0.98	**0.87****	**0.82*****	**0.79*****	1.00	1.02	1.01
Italy	1.00	0.96	0.97	0.98	0.98	0.96*	**0.83****	**0.83****	**0.80****	0.92	0.90*	0.94
Japan	1.01	1.00	1.00	1.02	1.03	0.94	1.04	1.05	1.05	1.07*	1.04	1.05
UK	**0.76****	**0.74***	**0.70***	**0.71***	**0.70***	**0.71***	**0.79***	**0.78***	**0.74***	**0.69***	**0.66***	**0.65***
US	0.94*	0.92*	**0.85****	0.97	0.94*	0.92*	1.01	1.01	0.95	1.08	1.07	1.05
	Pre-crisis period 1999.I-2007.IV						Pre-crisis period 1999.I-2007.IV					
Canada	1.00	1.00	0.98	1.01	1.02	1.01	0.95	0.99	1.02	0.99	1.01	0.99
France	0.98**	0.92**	0.95**	0.99**	1.01***	0.99	1.03*	0.95**	0.99**	1.05*	1.09**	1.06
Germany	**0.78****	**0.76****	**0.67****	0.96	1.05	1.06	**0.80****	**0.76****	**0.70****	0.90	0.98	1.00
Italy	0.98	0.96*	0.94**	1.00	1.02*	1.02*	0.98	0.96	0.96	0.97	1.02	1.02
Japan	0.99	0.98	0.99	0.97	0.98	0.96	1.03	1.04	1.04	1.03	1.04	1.02
UK	**0.88**	**0.85**	**0.84**	**0.85**	0.96	0.99	0.94	0.95	0.91	**0.88**	0.99	1.00
US	0.98*	0.98*	0.95*	1.03*	0.99	0.99	1.04	1.07	1.04	**1.16***	**1.14***	**1.15***
	Post-crisis period 2008.I-2013.IV						Post-crisis period 2008.I-2013.IV					
Canada	**0.70*****	**0.62****	**0.59*****	0.95*	**0.89*****	**0.85****	**0.74****	**0.67****	**0.61****	**0.85**	**0.84****	**0.78***
France	**0.86*****	**0.81*****	**0.78*****	**0.87*****	**0.87*****	**0.85*****	**0.88*****	**0.85*****	**0.81*****	0.96***	0.93***	**0.83*****
Germany	1.00	0.95	**0.84*****	1.02	1.00	0.96*	**0.88****	**0.83*****	**0.82*****	1.02	1.03*	1.01
Italy	1.01	0.96	0.98	0.97	0.96*	0.93***	**0.65*****	**0.73*****	**0.70*****	0.90	**0.85****	0.91**
Japan	1.06	1.04	1.01	1.05**	1.05**	0.93	1.06**	1.07***	1.05*	1.10*	1.04*	1.06**
UK	**0.74****	**0.73****	**0.68****	**0.69****	**0.68****	**0.69****	**0.77***	**0.76***	**0.72***	**0.68****	**0.64****	**0.63****
US	**0.88***	**0.84****	**0.72*****	0.91**	0.90*	**0.84***	0.96	0.93	**0.83*****	1.00	1.00	0.95

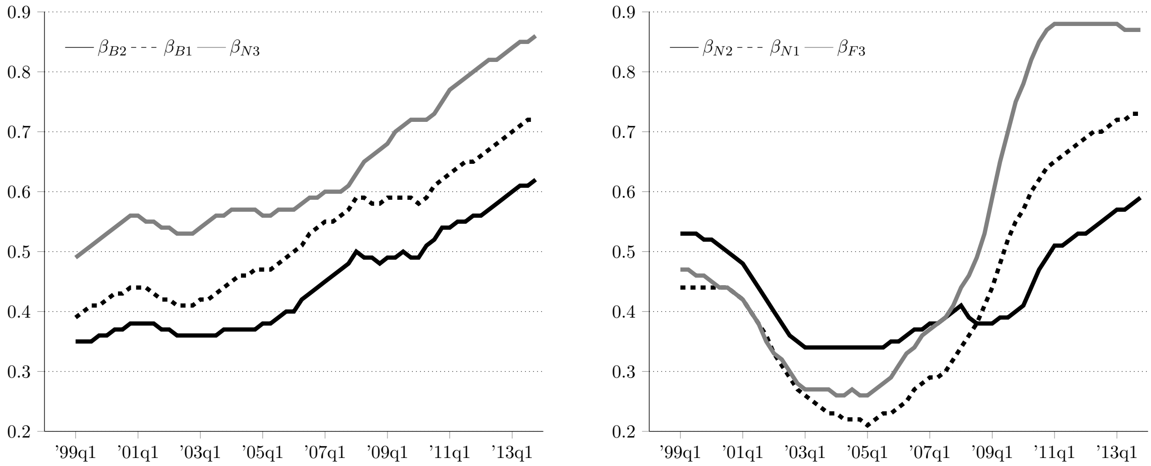
Notes: Figures in boldface indicate a deviation by more than ± 10% from 1. Starred entries (*,**,***) denote that the one-sided Diebold-Mariano test is significant at the 10%, 5% or 1% level. The alternative hypothesis of the DM test is MFSE > 1 or MFSE < 1, depending on whether the RMSFE is greater or smaller than 1, respectively.

Moreover, the marginal value of DFM forecasts is generally larger in the pre-crisis period than in the post-crisis period, which is consistent with the improvement in prediction capabilities of Consensus forecasts in the latter period. In fact, it is even significantly negative in a number of cases after 2008, reducing forecasting accuracy. An important empirical result is that it is difficult to beat a fresh Consensus forecast (columns N3 and F3) in the post-crisis period. With some exceptions, both combination rules perform worse than the Consensus forecast for both types of forecast after 2008. This finding suggests that during that period the Consensus forecast, which reflects the aggregate of both model-based predictions and subjective insights of all survey respondents, may incorporate all available information at the moment of release. Hence, the additional value of the DFM appears to be quite limited in the post-crisis period, and the optimal forecast is the Consensus forecast or something very close to it. The DFM may still serve as a means to enhance (non-updated) Consensus forecasts in the two months after the quarterly release. Looking at backcasts (B1 and B2), substantial accuracy gains are possible for Japan and Italy, and modest gains for France and Germany. There is also scope for improvement for the nowcasts N2 and N1, but it is smaller than for the corresponding backcasts.

A constant refrain in our results is that subjective Consensus forecasts have improved in predictive power relative to predictions from the dynamic factor model over time. This development is visualized in Figure 1, which depicts the weight of the Consensus forecast β in eq. (5) over the years 1999–2013.[15] The weights of the Consensus nowcast all show a clear upward trend (left-hand graph). Moreover, as expected, they shift downwards as they become older and the DFM exploits new information. The weight for the F3 forecast steeply increases after the financial crisis (right-hand panel). In the last years of the sample, the N3 and F3 forecast combinations are determined for 80–90% by the Consensus forecast and for only 10–20% by the DFM forecast. The observed trend may reflect several phenomena. First, our procedure may flatter the performance by the DFM in the early part of the sample, because analysts may in reality have used statistical models that were less sophisticated than the DFM, such as pure time series models, bridge models and VAR models. Second, analysts may put more effort in making their predictions in recessions and times of high volatility. This is consistent with the findings of the literature that forecasters adjust their forecast more frequently during recessions or when the information set changes a lot. Our findings support the findings 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. Third, the DFM forecasts are derived by a mechanical procedure that may be ill-suited to deal with large shocks, such as the financial crisis, although taking averages over all possible specifications and using a rolling estimation window offer some protection. By contrast, analysts base their subjective assessments on potentially a multitude of relevant time-varying factors and they may adjust their models, data and estimation modalities in response to large shocks. Recent work by Castle, Clements, and Hendry (2015) and Castle, Doornik, Hendry, and Pretis (2015) demonstrates the value of employing a statistical procedure that fully incorporates the possibility of location shifts. Finding out how such a forecasting procedure would alter the relative performance of statistical models and subjective forecasts is an interesting topic for future research.

[15] For presentational reasons we have averaged β across countries and estimation windows and smoothed the resulting time series by applying a centered eight-quarter moving average.

Figure 1: Weight of Consensus forecast in weighted average combination scheme, 1999.I-2013.IV



Finally, given the variation in relative performance of both combination schemes across countries and forecasting horizons, we investigate to what extent it is feasible for forecasters to identify in real time what combination scheme they should apply at a particular moment in time. We look into two strategies. The first strategy does not involve a choice, but consists of simply averaging the predictions generated by the two schemes (for all estimation windows). The empirical literature has found that such a strategy leads to less volatile predictions and possibly an improvement in accuracy (e.g. Kuzin et al., 2013; Timmermann, 2006). The second strategy tries to select the best forecasting rule among ten candidates (two combination schemes estimated using estimation windows of 4–8 years) on the basis of their recently observed (out-of-sample) forecasting ability. The latter is measured over a moving evaluation window that varies from one to four years. In view of the difficulty to optimally choose the length of the evaluation window on a priori grounds, we again apply averaging. The strategy first selects the pair of combination scheme and estimation window that delivers the best forecasts for each evaluation window. We then average the four resulting predictions.

The relative RMSFEs (versus the DFM) of the two strategies are presented in Table VI.[16] Our main finding is that the differences between them are often quite small and do not show a clear pattern. This is true for both subperiods and the whole sample. In many cases, the difference in forecasting performance between the weighted average and linear combination schemes does not appear to be large and persistent enough to be exploitable by a forward-looking selection strategy in real time. Moreover, a comparison of Tables V and VI shows that averaging tends leads to somewhat lower RMSFEs than either of the combination schemes. Using a simple average of combination schemes may thus provide a valuable hedge against misspecification and instability of the combination schemes for practitioners. However, as the Italian case illustrates, it may pay off to monitor the prediction performance of the selection strategy at the same time, and possibly switch strategies.

[16] Table A.IV in Appendix A.3 reports the relative RMSFEs versus the Consensus forecast.

Table VI: Forecasting performance of averaged or selected forecast combinations, 1999.I-2013.IV

	B2	B1	N3	N2	N1	F3	B2	B1	N3	N2	N1	F3
	Relative RMSFE combination DFM and CF versus DFM											
	A. Average of forecast combinations						B. Selection of best forecast combination in the recent past					
	Full sample 1999.I-2013.IV						Full sample 1999.I-2013.IV					
Canada	**0.80****	**0.77***	**0.77***	0.92*	0.90**	**0.86***	**0.76****	**0.75***	**0.77***	0.92	**0.89**	**0.84***
France	0.90**	**0.86*****	**0.83*****	0.92***	0.93***	**0.89****	0.95**	**0.89****	**0.87*****	0.94***	0.95***	0.92*
Germany	**0.89*****	**0.84*****	**0.79*****	0.99	1.00	0.98	**0.89*****	0.93	**0.85*****	1.04	1.01	1.01
Italy	**0.87****	**0.85****	**0.85****	0.92*	0.91**	0.92*	**0.81****	**0.86****	**0.81****	0.91	**0.89***	0.94
Japan	1.02	1.02	1.02	1.04	1.03	0.98	1.06	1.05	1.03	1.08	1.04	1.03
UK	**0.76***	**0.74***	**0.70***	**0.68***	**0.67***	**0.67***	**0.77***	**0.74***	**0.71***	**0.69***	**0.65***	**0.66***
US	0.96	0.94	**0.88****	1.01	1.00	0.97	0.97	0.95	0.91	1.00	0.97	0.95
	Pre-crisis period 1999.I-2007.IV						Pre-crisis period 1999.I-2007.IV					
Canada	0.93	0.96	0.98	0.99	1.00	0.98	**0.89**	0.95	1.01	1.02	1.04*	1.00
France	1.00**	0.93**	0.96**	1.01**	1.04**	1.02	1.01**	0.97**	1.00**	1.03*	1.06**	1.04
Germany	**0.76****	**0.74****	**0.67****	0.90	0.97	0.97	**0.78****	**0.76****	**0.69****	**0.89**	0.96	0.96
Italy	0.96	0.94	0.93*	0.96	0.99	0.99	0.95	0.97	0.97	0.98	1.02	1.01
Japan	1.00	1.00	1.01	1.00	1.00	0.99	1.03	1.03	1.01	0.99	1.00	0.98
UK	0.90	**0.89**	**0.86**	**0.86**	0.97	0.98	0.91	**0.89**	**0.86**	**0.88**	0.99	0.99
US	0.98	0.99	0.97	1.07*	1.05	1.06	1.01	1.04	1.02	1.06	1.04	1.05
	Post-crisis period 2008.I-2013.IV						Post-crisis period 2008.I-2013.IV					
Canada	**0.71*****	**0.63****	**0.59****	**0.86****	**0.85*****	**0.80****	**0.67*****	**0.61****	**0.56****	**0.84****	**0.78****	**0.74****
France	**0.83*****	**0.81*****	**0.76*****	**0.89*****	**0.88*****	**0.83*****	0.90***	**0.84*****	**0.80*****	0.90***	0.91***	**0.87*****
Germany	0.92***	**0.87*****	**0.81*****	1.01	1.00	0.98	0.92**	0.97	**0.88*****	1.08	1.03	1.02
Italy	**0.77*****	**0.79*****	**0.80*****	0.90*	**0.87*****	0.90***	**0.65*****	**0.78*****	**0.71*****	**0.88***	**0.84*****	0.92*
Japan	1.05**	1.05**	1.03	1.06*	1.04**	0.98	**1.13***	1.10*	1.05	**1.13****	1.07**	1.05**
UK	**0.74****	**0.72****	**0.69****	**0.66****	**0.65****	**0.65****	**0.75****	**0.72***	**0.70****	**0.67****	**0.63****	**0.64****
US	0.91	**0.87***	**0.77*****	0.94**	0.94	**0.89**	0.91	**0.83****	**0.76*****	0.94	0.90*	**0.86***

Notes: Figures in boldface indicate a deviation by more than $\pm$ 10% from 1. Starred entries (*,**,***) denote that the one-sided Diebold-Mariano test is significant at the 10%, 5% or 1% level. The alternative hypothesis of the DM test is MFSE $>$ 1 or MFSE $<$ 1, depending on whether the RMSFE is greater or smaller than 1, respectively.

4 Conclusion

This paper investigates to what extent subjective information incorporated in forecasts by professional analysts may enrich mechanical forecasting procedures exploiting monthly statistical data in a truly real-time context, in which both statistical models and analysts have to deal with possibly inaccurate initial GDP estimates. We take our judgmental forecasts from the quarterly Consensus survey (averaged over the panelists). Our model-based forecasts are generated by a dynamic factor model that is estimated using real-time monthly vintages. For the sake of robustness, our analysis covers seven large countries (G7 countries) over the years 1999–2013, allowing a systematic comparison of the tranquil period 1999–2007 before the financial crisis and the volatile post-crisis period after 2008. Moreover, we analyze two different schemes to combine model-based and subjective forecasts: (i) the weighted average and (ii) the linear combination. Our research addresses a new question within the empirical literature on forecasting real GDP in the near term.

Our findings can be summarized in five points. First, in keeping with other work, we find that monthly statistical indicators contain valuable information that can be extracted by the dynamic factor model in real-time, in particular as the horizon shortens and more monthly information is processed. The largest gains in predictive accuracy are for late nowcasts and backcasts, when the model is able to use statistical data that pertain to the quarter of interest. Its relative strength is thus to improve the assessment of the current state of the economy. Moreover, the dynamic factor model is generally more efficient in extracting monthly information in volatile times.

Second, the forecasting abilities of Consensus forecasts (averaged over the panelists) remarkably improve after the financial crisis, making them a tough competitor for the mechanical dynamic factor model since 2008. In the stable pre-crisis period, the dynamic factor model tends to outperform professional analysts. But in the volatile post-crisis period, Consensus nowcasts and forecasts constitute superior predictions at the moment of their release. This pattern suggests that analysts pay more attention and devote more effort to forecasting in volatile times. This is also consistent with earlier findings that analysts quickly adjust their forecasts during recessions or when the information set changes a lot. It also suggests that strictly mechanical procedures may be more susceptible to extreme observations in the estimation sample, even when taking averages across specifications and using rolling windows in estimation.

Third, the difference in forecasting performance between professional analysts and the DFM tends to be greater for a fresh Consensus nowcast (N3) than for a fresh Consensus one-quarter ahead forecast (F3). This suggests that the value added of subjective insights may be greater when analysts know at least some hard and soft data relating to the quarter of interest itself. In between the quarterly release dates, the performance of Consensus forecasts versus DFM forecasts deteriorates, as the latter are able to benefit from newly released monthly data.

Fourth, enhancing model-based forecasts with subjective information via simple combination schemes delivers sizable gains in forecasting ability of statistical models for all countries except Japan in the years 1999–2013, even when the Consensus forecasts are somewhat dated. Accuracy gains are often modest in the rather stable pre-crisis years 1999-2007, with DFM and Consensus forecasts contributing about equally to the combined forecast. The advantages of adding judgmental information are much more pronounced in the volatile period after 2008 due to a marked improvement in predictive power of Consensus forecasts. Towards the end of the sample, we find that Consensus forecasts are the main determinant of the forecast combination, suggesting that the marginal information content of the DFM forecasts has become rather low for many countries. Consequently, the benefits from using a combination scheme when measured against the Consensus forecasts' performance have generally diminished over time, and are mostly small or absent for all countries except for Italy and Japan since 2008.

Fifth, both the weighted average scheme and the linear combination scheme are suitable to enhance DFM forecasts in economically meaningful ways. Although we find some heterogeneity in forecasting capabilities of the weighting schemes, both across countries as well as time, our results suggest that in many cases forecasters are unlikely to be able to reliably identify in real time what combination scheme they should apply at any point in time. In practical applications, using a simple average of combination schemes may therefore offer a valuable hedge against misspecification and instability of combination schemes.

The results of our analysis may be useful to policy makers, financial analysts and economic agents, as information on where the economy stands and where it is heading to in the short run is particularly valuable in times of great uncertainty. Our analysis demonstrates that judgmental forecasts by professional analysts contain valuable additional information that can be extracted in real time. This paper does so in a mechanical fashion, but forecasters may in practice follow more flexible approaches to take this information on board. Moreover, forecasters may use Consensus forecasts, which represent the view of their peers, as a cross-check on their own model predictions and judgmental views on the near-term prospects of the economy.

Our finding that since 2008 for many countries (freshly released) Consensus forecasts are hard to improve upon does not imply that practitioners are better off without making their own model-based forecasts. Model-based forecasts are still valuable for a number of reasons. First, Consensus forecasts are released just once a quarter. Especially in volatile times, when it really counts, they run the risk of becoming out-of-date rather quickly. Second, an important disadvantage of forecast surveys, such as the Consensus survey, is their black box nature. If forecasts change, it is unclear for what reasons. As argued by Bańbura et al. (2011), a crucial aim of nowcasting is the structured way of updating forecasts on the basis of the continuous flow of new statistical data. The forecasting process in this way generates information on the relative marginal information content of the various statistical indicators. Moreover, several types of models, including the dynamic factor model, allow the decomposition of forecasts into the contributions of the various (types of) statistical variables in the data set (e.g. Bańbura and Rünstler, 2011), which may also assist analysts in their reading of economic conditions and their near-term development. These analytical by-products of statistical procedures may also serve to detect implausible aspects of model-based forecasts, which constitute crucial ingredients for the judgmental component of the forecast. The story behind the forecast is for many forecasters at least as important as the number itself. Third, the expectations by the Consensus survey participants reflect both model-based and subjective information. As different forecasters will employ different models, estimation procedures and data sets, the model-based component differs across analysts. The strong performance of the mean Consensus forecast is partly attributable to the fact that different statistical approaches have been used to filter the available statistical data.

Finally, our findings suggest that model predictions obtained by mechanical estimation procedures may be vulnerable to extreme observations in estimation samples in real time, which may signal structural breaks in data generating processes. Consequently, an important topic for future research is to explore how statistical forecasting methods can be made more robust to this phenomenon in real time, either by heuristic methods or by fully incorporating the identification and estimation of location shifts into the empirical forecasting strategy. Moreover, an in-depth investigation into the prediction strategies that professional analysts employ in practice when dealing with large shocks, both when these have just occurred and after economic conditions have settled down, could prove to be very insightful.

Acknowledgements

The opinions expressed in this paper are the personal views of the authors and do not necessarily reflect positions of De Nederlandsche Bank or the Ministry of Finance. We are grateful to Peter van Els, Jakob de Haan and seminar and conference participants at the ECB and the International Symposium on Forecasting (2015, Riverside, CA) for their helpful comments. We are particularly indebted to Job Swank for his valuable comments and stimulating suggestions.

References

Ager, P., M. Kappler, and S. Osterloh (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.

Babura, M., D. Giannone, M. Modugno, and L. Reichlin (2013). Now-casting and the real-time data flow. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2, Part A of *Handbook of Economic Forecasting*, pp. 195 – 237. Elsevier.

Baffigi, A., R. Golinelli, and G. Parigi (2004). Bridge models to forecast the euro area GDP. *International Journal of Forecasting 20*(3), 447–460.

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

Bai, J. and S. Ng (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics 25*(1), 52–60.

Bańbura, M., D. Giannone, and L. Reichlin (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics 25*(1), 71–92.

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

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

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

Batchelor, R. (2001). How useful are the forecasts of intergovernmental agencies? the IMF and OECD versus the Consensus. *Applied Economics 33*(2), 225–235.

Bates, J. M. and C. W. J. Granger (1969). The combination of forecasts. *Operations Research Quarterly 20*(4), 451–468.

Camacho, M. and G. Perez-Quiros (2010). Introducing the euro-sting: short-term indicator of euro area growth. *Journal of Applied Econometrics 25*(4), 663–694.

Camacho, M. and G. P. Quiros (2011). Spain-sting: Spain short term indicator of growth. *Manchester School 79*(1), 594–616.

Carriero, A., T. E. Clark, and M. Marcellino (2015). Realtime nowcasting with a bayesian mixed frequency model with stochastic volatility. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 178*(4), 837–862.

Castle, J. L., M. P. Clements, and D. F. Hendry (2015). Robust approaches to forecasting. *International Journal of Forecasting 31*(1), 99–112.

Castle, J. L., J. A. Doornik, D. F. Hendry, and F. Pretis (2015). Detecting location shifts during model selection by step-indicator saturation. *Econometrics 3*(2), 240–264.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research 1*(2), 245–276.

Cheung, C. and F. Demers (2007). Evaluating forecasts from factor models for Canadian GDP growth and core inflation. Working Papers 07-8, Bank of Canada.

Clemen, R. T. (1986). Linear constraints and the efficiency of combined forecasts. *Journal of Forecasting 5*(1), 31–38.

Croushore, D. (2011). Frontiers of real-time data analysis. *Journal of Economic Literature 49*(1), 72–100.

de Winter, J. (2011). Forecasting GDP growth in times of crisis: private sector forecasts versus statistical models. DNB Working Papers 320, De Nederlandsche Bank.

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.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 253–63.

Durbin, J. and S. J. Koopman (2012). *Time series analysis by state space methods.* Oxford University Press.

Fair, R. C. and R. J. Shiller (1990). Comparing information in forecasts from econometric models. *American Economic Review 80*(3), 375–389.

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

Genre, V., G. Kenny, A. Meyler, and A. Timmermann (2013). Combining expert forecasts: Can anything beat the simple average? *International Journal of Forecasting 29*(1), 108–121.

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

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

Golinelli, R. and G. Parigi (2014). Tracking world trade and GDP in real time. *International Journal of Forecasting 30*(4), 847–862.

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

Hubert, P. (2014). Revisiting the greenbook's relative forecasting performance. *Revue de l'OFCE-Débats et politiques 137*, 151–179.

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

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

Kuzin, V., M. Marcellino, and C. Schumacher (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., G. Isiklar, and P. Loungani (2006). How quickly do forecasters incorporate news? evidence from cross-country surveys. *Journal of Applied Econometrics 21*(6), 703–725.

Lahiri, K. and G. Monokroussos (2013). Nowcasting US GDP: The role of ISM business surveys. *International Journal of Forecasting 29*(4), 644 – 658.

Lamont, O. A. (2002). Macroeconomic forecasts and microeconomic forecasters. *Journal of Economic Behavior & Organization 48*(3), 265–280.

Laster, D., P. Bennett, and I. S. Geoum (1999). Rational bias in macroeconomic forecasts. *The Quarterly Journal of Economics 114*(1), 293–318.

Liebermann, J. (2014). Real-time nowcasting of GDP: A factor model vs. professional forecasters. *Oxford Bulletin of Economics and Statistics 76*(6), 783–811.

Loungani, P. and J. Rodriguez (2008). Economic forecasts. *World Economics 9*(2), 1–12.

Lundquist, K. and H. O. Stekler (2012). Interpreting the performance of business economists during the great recession. *Business Economics 47*(2), 148–154.

Meyler, A. and I. Rubene (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., K. Barhoumi, S. Benk, R. Cristadoro, A. den Reijer, A. Jakaitiene, P. Jelonek, A. Rua, K. Ruth, and C. V. Nieuwenhuyze (2009). Short-term forecasting of GDP using large datasets: a pseudo real-time forecast evaluation exercise. *Journal of Forecasting 28*(7), 595–611.

Schneider, M. and M. Spitzer (2004). Forecasting Austrian GDP using the generalized dynamic factor model. Working Papers 89, Oesterreichische Nationalbank.

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

Stark, T. (2010). Realistic evaluation of real-time forecasts in the survey of professional forecasters. *Research Rap Special Report*.

Stekler, H. O. (1991). Macroeconomic forecast evaluation techniques. *International Journal of Forecasting 7*(3), 375–384.

Stock, J. H. and M. W. Watson (2011). Dynamic factor models. *Oxford Handbook of Economic Forecasting 1*, 35–59.

Timmermann, A. (2006). Forecast combinations. In C. G. G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1 of *Handbook of Economic Forecasting*, pp. 135 – 196. Elsevier.

A Appendix

A.1 Data set

Table A.I provides the list of the monthly indicators that have been used for the estimation of the dynamic factor models. As discussed in the main text, the data can be split-up into three parts: the domestic economy, global economic activity and the main trading partners; see the headings in the table. Furthermore we classify our data into four categories: hard, quantitative information (hard), consumer and producer prices (price), financial variables (financial) and soft, qualitative information (soft). We collected real-time vintages for all time series, which start in January 1985 if possible.[17]

The main data source for our real-time database for the United States was the ALFRED database, the US real-time database maintained by the Federal Reserve Bank of St. Louis. We used real-time data vintages from January 1992 up until September 2014. The variables in the ALFRED database are updated with each subsequent release of one of the series. On the basis of these release dates we constructed a database that mimics the release pattern of the main data source for the other G7-countries, the OECD Main Economic Indicators Original Data Release and Revisions Database (OECD RTDB).

We augmented the real-time data for each country with indicators for global economic activity, financial variables and qualitative information on expectations derived from surveys among consumers, retailers and firms. Concerning the indicators for global economic activity we used real-time vintages on world trade from the CPB World Trade Monitor.[18] The other indicators of global activity and the financial variables are not subject to revisions. For these indicators we used the latest data vintage to construct backdated vintages based on the release pattern. The main source for survey data was the European Commission. Moreover, we collected country-specific business survey data for Germany, France, Italy and Belgium: the Ifo business climate index, the INSEE business cycle indicator, the ISAE consumer confidence indicator and the BNB business survey, respectively.

Quarterly GDP data for the US were taken from the ALFRED database. For the other G7 countries, the OECD RTDB was the source. As the latter contains no German GDP data before 1999.I, we took the German GDP data before 1999.I from the Deutsche Bundesbank.

[17] There are a few exceptions: the 10-year treasury bill rate (January 1989) and the 3-month treasury bill rate (July 1985) of Japan, negotiated wages (January 1990) for Germany, the Baltic freight index (May 1985), the VIX Standard and Poor's 500 index (January 1986) and the crude oil price, west Texas intermediate (January 1986).

[18] The World Trade Monitor series start in January 1991. We backdated the world trade data for the period January 1985 until December 1990, using monthly import and export volume series from the IMF.

Table A.I: Description monthly database

Nr.	Variable name	Type	Source	CA	DE	FR	IT	JA	UK	US
Domestic Economy Variables										
1	Industrial production index	hard	RTDB, ALFRED	X	X	X	X	X	X	X
2	Construction production index	hard	RTDB	X	X	X	.	.	.	.
3	Personal consumption expenditure	hard	ALFRED	.	.	.	.	.	.	X
4	Permits issued for dwellings	hard	ALFRED	X	.	.	.	.	.	.
5	Housing starts	hard	ALFRED	.	.	.	.	.	.	X
6	New residential sales	hard	ALFRED	.	.	.	.	.	.	X
7	Real disposable personal income	hard	ALFRED	.	.	.	.	.	.	X
8	Unemployment rate	hard	RTDB, ALFRED	X	X	X	X	X	X	X
9	Employment	hard	RTDB, ALFRED	X	.	.	.	X	.	X
10	Machinery orders, private sector	hard	COJ	.	.	.	.	X	.	.
11	Manufacturers' new orders, durable goods	hard	ALFRED	.	.	.	.	.	.	X
12	Imports	hard	RTDB, ALFRED	X	X	X	X	X	X	X
13	Exports	hard	RTDB, ALFRED	X	X	X	X	X	X	X
14	Retail and food service sales	hard	RTDB, ALFRED	X	X	X	X	X	X	X
15	Passenger car registration	hard	MEI	X	X	X	X	X	X	X
16	Producer price index	price	MEI, ALFRED	X	X	.	.	X	.	X
17	Consumer price index, all urban consumers	price	RTDB, ALFRED	X	X	X	X	X	X	X
18	Hourly earnings manufacturing	price	RTDB	X	.	.	X	X	X	.
19	Negotiated wages and salaries per hour	price	Buba	.	X	.	.	.	.	.
20	Share price index	financial	MEI, ALFRED	X	X	X	X	X	X	X
21	10-year treasury bill rate	financial	RTDB, ALFRED	X	X	X	.	X	X	X
22	3-month treasury bill rate	financial	RTDB, ALFRED	X	X	X	X	X	X	X
23	Real effective exchange rate (CPI weights)	financial	MEI	X	X	X	X	X	X	.
24	Trade weighted exchange index, maj. curr	financial	ALFRED	.	.	.	.	.	.	X
25	M1	financial	MEI	X	X	X	X	X	X	.
26	Ifo business climate index	soft	Ifo	.	X	X	X	.	X	.
27	BNB business survey	soft	BNB	.	X	X	X	.	X	.
28	INSEE business cycle indicator	soft	INSEE	.	X	X	X	.	X	.
29	ISAE consumer confidence indicator	soft	ISAE	.	X	X	X	.	X	.
30	Composite leading indicator	soft	RTDB	X	X	X	X	X	X	X
31	Economic sentiment indicator	soft	EC	.	X	X	X	.	X	.
32	Industrial confidence indicator	soft	EC	.	X	X	X	.	X	.
33	Consumer confidence indicator	soft	EC	.	X	X	X	.	X	.
34	Retail trade confidence indicator	soft	EC	.	X	X	X	.	X	.
35	Construction confidence indicator	soft	EC	.	X	X	X	.	X	.
36	Index of business conditions	soft	COJ	.	.	.	.	X	.	.

Continued on next page...

Table A.I – Continued

Nr.	Variable name	Type	Source	CA	DE	FR	IT	JA	UK	US
37	General business conditions US	soft	Philadelphia FED	.	.	.	.	.	.	X
38	Conf. board consumer confidence index	soft	Conference Board	.	.	.	.	.	.	X
39	Bloomberg consumer comfort index	soft	Bloomberg	.	.	.	.	.	.	X
40	Purchasing manager index, manufact.	soft	Markit	.	.	.	.	.	.	X
Global Variables										
41	World trade	hard	CPB	X	X	X	X	X	X	X
42	Worldwide semiconductor sales	hard	SIA	X	X	X	X	X	X	X
43	Crude oil, west Texas intermediate	price	ALFRED	X	X	X	X	X	X	X
44	World market ind. mat. price	price	HWWA	X	X	X	X	X	X	X
45	World market metals price, metals	price	HWWA	X	X	X	X	X	X	X
46	Baltic freight index	price	BALTEX	X	X	X	X	X	X	X
47	VIX Standard and Poor's 500 index	financial	CBOE	X	X	X	X	X	X	X
Trading Partners										
48	CA: Industrial production index	hard	RTDB	.	.	.	.	.	.	X
49	CA: Imports	hard	RTDB	.	.	.	.	.	.	X
50	CA: OECD Composite leading indicator	soft	RTDB	.	.	.	.	.	.	X
51	CH: OECD Composite leading indicator	soft	RTDB	.	.	.	.	X	.	.
52	DE: Industrial production index	hard	RTDB	.	.	X	X	.	X	.
53	DE: Imports	hard	RTDB	.	.	X	X	.	X	.
54	DE: OECD Composite leading indicator	soft	RTDB	.	.	X	X	.	X	.
55	FR: Industrial production index	hard	RTDB	.	X	.	X	.	.	.
56	FR: Imports	hard	RTDB	.	X	.	X	.	.	.
57	FR: OECD Composite leading indicator	soft	RTDB	.	X	.	X	.	.	.
58	MX: Industrial production	hard	RTDB	.	.	.	.	.	.	X
59	MX: Imports	hard	RTDB	.	.	.	.	.	.	X
60	MX: OECD Composite leading indicator	soft	RTDB	.	.	.	.	.	.	X
61	US: Industrial production index	hard	ALFRED	X	X	X	.	X	X	.
62	US: Imports	hard	ALFRED	X	X	X	.	X	X	.
63	US: General business conditions	soft	Philadelphia FED	X	X	X	.	X	X	.
64	US: Conference board cons. conf.	soft	ALFRED	X	X	X	.	X	X	.
65	UK: Industrial production	hard	RTDB	X	.	.	.	.	.	.
66	UK: Imports	hard	RTDB	X	.	.	.	.	.	.
67	UK: OECD Composite Leading Indicator	soft	RTDB	X	.	.	.	.	.	.
N				32	39	37	35	30	37	36

Abbreviations: ALFRED: Archival federal reserve economic data, BNB: Banque nationale de Belgique, Buba: Bundesbank, CBOE: Chicago board options exchange, Baltex: Baltic exchange London, CPB: Cpb economic policy analysis, COJ: Cabinet office Japan, EC: European commission, HWWA: Hamburgisches welt-wirtschafts-archiv, Ifo: Ifo institute for economic research, INSEE: National institute of statistics and economic studies, ISAE: Italian institute for studies and economic analysis, RTDB: OECD real-time data and revisions database, MEI: OECD main economic indicators, SIA: Semiconductor industry association.

A.2 State space representation dynamic factor model

The equations of the DFM, eqs. 1-4, can be cast in state space form as illustrated below for the case of $p = 1$. The aggregation rule is implemented in a recursive way in eq. 8 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:

$$\begin{bmatrix} x_t \\ y_t^Q \end{bmatrix} = \begin{bmatrix} \Lambda & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_t \\ y_t \\ \hat{y}_t^Q \end{bmatrix} + \begin{bmatrix} \xi_t \\ \varepsilon_t^Q \end{bmatrix} \tag{7}$$

and the transition equation:

$$\begin{bmatrix} I_r & 0 & 0 \\ -\beta' & 1 & 0 \\ 0 & -\frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} f_{t+1} \\ y_{t+1} \\ \hat{y}_{t+1}^Q \end{bmatrix} = \begin{bmatrix} A_{r1} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Xi_{t+1} \end{bmatrix} \begin{bmatrix} f_t \\ y_t \\ \hat{y}_t^Q \end{bmatrix} + \begin{bmatrix} \zeta_{t+1} \\ \varepsilon_t \\ 0 \end{bmatrix} \tag{8}$$

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 forecasting of quarterly GDP growth y_t^Q and dealing efficiently with an unbalanced data set of missing observations at the beginning and at the end of the series by replacing the missing data with optimal predictions. Moreover, when compared with using principal components technique alone, the two-step estimator allows for dynamics of the common factors and cross-sectional heteroskedasticity of the idiosyncratic component.

A.3 Additional results

Table A.II is a counterpart of Table II in the main text. It presents relative RMSFEs versus the RW for two different set-ups. Panel B reports results for the real-time procedure when the prediction errors are computed using the last vintage for GDP as measure of actual GDP. Panel A reports results for the corresponding pseudo-real-time procedure, which also implies that the last vintage for GDP serves as the measure of actual GDP. Tables A.III and A.IV are the respective counterparts of Tables V and VI in the main text, where the relative RMSFE is expressed in terms of the RMSFE of the Consensus forecast. Finally, Table A.V reports the estimated coefficients of eq. (6), which display the same pattern as the ones for eq. (5).

Table A.II: Forecasting performance of Dynamic Factor Model in alternative set-ups, 1999.I-2013.IV

	RW	B2	B1	N3	N2	N1	F3	Flash	B2	B1	N3	N2	N1	F3
	RMSFE	Relative RMSFE DFM versus Random Walk						RMSFE	Relative RMSFE DFM versus Random Walk					
		A. Pseudo real time							B. Final vintage is target					
		Full sample 1999.I-2013.IV							Full sample 1999.I-2013.IV					
Canada	*0.66*	**0.60***	**0.64***	**0.72***	**0.76***	**0.82**	**0.88**	*0.27*	**0.63***	**0.66***	**0.72***	**0.78***	**0.82**	**0.88**
France	*0.55*	**0.57****	**0.61****	**0.67****	**0.73***	**0.77***	**0.84***	*0.22*	**0.60****	**0.66****	**0.73****	**0.77***	**0.80***	**0.86***
Germany	*0.90*	**0.63***	**0.67***	**0.75***	**0.79***	**0.84***	**0.88***	*0.39*	**0.70***	**0.72***	**0.79***	**0.84**	**0.84***	**0.89***
Italy	*0.78*	**0.67***	**0.68***	**0.77***	**0.79***	**0.74***	**0.81***	*0.33*	**0.62***	**0.66***	**0.74***	**0.77***	**0.78****	**0.86****
Japan	*1.12*	**0.69***	**0.73***	**0.80***	0.96	1.05	1.07*	*0.74*	**0.70***	**0.74***	**0.81**	0.96	1.09	1.08
UK	*0.82*	**0.74**	**0.77**	**0.80**	**0.85**	0.91	0.92	*0.43*	**0.82**	**0.83**	**0.86**	0.91	0.93	0.94
US	*0.70*	**0.63***	**0.62***	**0.69***	**0.78**	**0.85**	0.90	*0.36*	**0.67***	**0.70***	**0.75***	**0.83**	**0.88**	0.92
		Pre-crisis period 1999.I-2007.IV							Pre-crisis period 1999.I-2007.IV					
Canada	*0.45*	**0.78***	**0.81***	**0.87**	**0.89**	0.97	0.98	*0.24*	**0.81****	**0.84***	**0.86***	0.92	0.96	0.97
France	*0.36*	**0.63*****	**0.68*****	**0.66*****	**0.69*****	**0.73*****	**0.84****	*0.21*	**0.66*****	**0.70*****	**0.68*****	**0.69*****	**0.73*****	**0.86****
Germany	*0.59*	**0.79*****	**0.81*****	**0.85*****	**0.88*****	**0.87****	0.90**	*0.44*	**0.84*****	**0.85*****	0.91**	0.94	0.90**	0.93*
Italy	*0.45*	**0.86*****	**0.87****	0.90*	0.90*	**0.85*****	**0.89****	*0.32*	**0.88*****	**0.87****	**0.89****	**0.88****	**0.85*****	**0.89****
Japan	*0.68*	0.99	0.97	0.94	0.96	0.96	0.96*	*0.85*	1.01	1.00	0.96	0.95	0.94	0.96
UK	*0.46*	1.09	1.06	1.01	1.01	1.00	0.96	*0.47*	0.97	0.96	0.96	0.96	0.95	0.96
US	*0.52*	**0.79****	**0.79****	**0.83****	0.90*	0.91**	0.95	*0.35*	**0.80****	**0.81*****	**0.84****	0.92*	0.93*	0.96
		Post-crisis period 2008.I-2013.IV							Post-crisis period 2008.I-2013.IV					
Canada	*0.89*	**0.52***	**0.57***	**0.65**	**0.71**	**0.75**	**0.83**	*0.31*	**0.54**	**0.57**	**0.66**	**0.71**	**0.76**	**0.84**
France	*0.75*	**0.54***	**0.58***	**0.67***	**0.75**	**0.78**	**0.84**	*0.22*	**0.58***	**0.65***	**0.74**	**0.80**	**0.83**	**0.86**
Germany	*1.22*	**0.56***	**0.61***	**0.71***	**0.76***	**0.83***	**0.87***	*0.29*	**0.64***	**0.66***	**0.74***	**0.80***	**0.82***	**0.87***
Italy	*1.11*	**0.61***	**0.63***	**0.73***	**0.76***	**0.72***	**0.79***	*0.33*	**0.53***	**0.60***	**0.69***	**0.74***	**0.76****	**0.85****
Japan	*1.57*	**0.58****	**0.64****	**0.75***	0.96	1.07*	1.10***	*0.53*	**0.58****	**0.65***	**0.77**	0.96	**1.12***	**1.11****
UK	*1.16*	**0.63**	**0.68**	**0.74**	**0.81**	**0.89**	0.92	*0.37*	**0.77**	**0.80**	**0.84**	0.90	0.92	0.94
US	*0.91*	**0.53**	**0.52**	**0.60**	**0.72**	**0.83**	**0.87**	*0.38*	**0.60**	**0.63**	**0.70**	**0.79**	**0.85**	0.91

Notes: Entries in italics report RMSFEs; column RW refers to the third nowcast of the RW and column Flash shows the RMSFE of the flash as predictor of the final vintage. All other entries report the RMSFE relative to random walk's RMSFE. Figures in boldface indicate a deviation by more than $\pm$ 10% from 1. Starred entries (*,**,***) denote that the one-sided Diebold-Mariano test is significant at the 10%, 5% or 1% level. The alternative hypothesis of the DM test is MFSE $>$ 1 or MFSE $<$ 1, depending on whether the RMSFE is greater or smaller than 1, respectively.

Table A.III: Forecasting performance forecast combination schemes (relative to Consensus forecast), 1999.I-2013.IV

	B2	B1	N3	N2	N1	F3	B2	B1	N3	N2	N1	F3
	Relative RMSFE combination DFM and CF versus CF											
	A. Weighted average scheme						B. Linear combination scheme					
	Full sample 1999.I-2013.IV						Full sample 1999.I-2013.IV					
Canada	0.90	0.90	0.94	**0.88***	0.92	0.94	0.90	0.92	0.98	**0.83****	**0.89***	0.90
France	0.90**	0.97	1.00	0.92	0.92*	0.97*	0.93*	1.01	1.03	1.00	0.99	0.99
Germany	1.02	1.02	1.00	1.00	0.98	1.00	0.92	0.91	0.98	0.99	1.00	1.03
Italy	**0.89**	0.99	1.07*	0.97	0.93	1.00	**0.74**	**0.85**	**0.88**	0.92	**0.86****	0.98
Japan	**0.83***	**0.88**	0.96	0.96	1.09	1.01	**0.85**	0.92	1.00	1.02	1.10	**1.13***
UK	1.01	1.01	1.02	0.99	0.99	0.99	1.05	1.06	1.08	0.98	0.93	0.92*
US	0.95	1.00	1.05**	0.95	1.01	1.02	1.01	1.10***	**1.17****	1.05*	**1.14*****	**1.17*****
	Pre-crisis period 1999.I-2007.IV						Pre-crisis period 1999.I-2007.IV					
Canada	**0.77****	**0.80***	**0.88**	**0.78****	**0.79****	**0.80****	**0.74**	**0.80**	0.92	**0.76****	**0.78****	**0.79***
France	**0.89*****	0.93	0.91**	**0.84*****	**0.83*****	0.93**	0.93*	0.96	0.95*	0.90***	0.90**	1.00
Germany	0.99	0.97	0.98	0.99	0.94	0.92	1.01	0.97	1.02	0.93	**0.88**	**0.87**
Italy	0.96	0.98	0.99	0.99	**0.85***	**0.89**	0.97	0.98	1.00	0.96	**0.86***	**0.89**
Japan	0.95*	0.96*	0.96*	0.99	0.99	1.00	0.99	1.01	1.01	1.06*	1.05	1.06*
UK	**0.84****	**0.80*****	**0.82*****	**0.78*****	**0.71*****	**0.74*****	0.90	**0.89**	0.90	**0.81*****	**0.73*****	**0.74*****
US	0.95	1.00	1.06**	0.94	0.99	1.01	1.00	1.10**	**1.16***	1.06*	**1.15****	**1.17****
	Post-crisis period 2008.I-2013.IV						Post-crisis period 2008.I-2013.IV					
Canada	**1.16**	**1.11***	1.08*	1.03	1.07*	**1.11***	**1.23***	**1.20****	**1.12****	0.92	1.01	1.03
France	0.91	1.01	1.08	0.96	0.98	1.00	0.94	1.05	**1.11*****	1.06	1.06	0.98
Germany	1.03*	1.03***	1.00	1.00	1.00	1.01	0.91	0.90	0.97	1.00	1.03	1.06*
Italy	**0.83**	1.00	**1.12*****	0.97	0.97	1.06	**0.54**	**0.76**	**0.80**	0.90	**0.86**	1.03
Japan	**0.67***	**0.78**	0.96	0.95	**1.15**	1.02	**0.67**	**0.81**	1.00	0.99	**1.14**	**1.16***
UK	1.05	1.06	1.06	1.03	1.05	1.04	1.09*	1.10*	**1.12****	1.01	0.98	0.95
US	0.95	1.00	1.03*	0.96	1.02	1.03*	1.04	**1.11****	**1.18***	1.05	**1.14*****	**1.16*****

Notes: Figures in boldface indicate a deviation by more than ± 10% from 1. Starred entries (*,**,***) denote that the one-sided Diebold-Mariano test is significant at the 10%, 5% or 1% level. The alternative hypothesis of the DM test is MFSE $>$ 1 or MFSE $<$ 1, depending on whether the RMSFE is greater or smaller than 1, respectively.

Table A.IV: Forecasting performance of averaged or selected forecast combinations (relative to Consensus forecast), 1999.I-2013.IV

	B2	B1	N3	N2	N1	F3	B2	B1	N3	N2	N1	F3
	Relative RMSFE combination DFM and CF versus CF											
	A. Average of combined forecasts						B. Selection of best combined forecast in the recent past					
	Full sample 1999.I-2013.IV						Full sample 1999.I-2013.IV					
Canada	**0.87**	**0.89**	0.94	**0.83****	**0.89***	0.90	**0.82**	**0.87**	0.94	**0.83****	**0.87****	**0.87****
France	**0.89****	0.97	0.99	0.94	0.94	0.97	0.93*	1.01	1.04	0.95	0.96	1.01
Germany	0.95	0.94	0.97	0.98	0.98	1.00	0.95	1.04	1.04	1.04	0.99	1.03
Italy	**0.77**	**0.88**	0.94	0.91	**0.86****	0.97	**0.72**	**0.88**	0.90	0.91	**0.85****	0.99
Japan	**0.83***	**0.89**	0.98	0.98	1.09	1.06*	**0.87**	0.92	0.99	1.02	**1.11**	**1.11**
UK	1.01	1.01	1.02	0.95**	0.94**	0.94***	1.03	1.01	1.04	0.97	0.92*	0.92*
US	0.96	1.03	1.08**	0.98	1.06***	1.08***	0.98	1.04	**1.12****	0.98	1.03**	1.06***
	Pre-crisis period 1999.I-2007.IV						Pre-crisis period 1999.I-2007.IV					
Canada	**0.72***	**0.78**	**0.88**	**0.76****	**0.77****	**0.78****	**0.69****	**0.77**	0.91	**0.78****	**0.81****	**0.79****
France	0.90**	0.94	0.92**	**0.86*****	**0.86*****	0.96	0.91*	0.98	0.95	**0.88*****	**0.87*****	0.98
Germany	0.96	0.94	0.98	0.93	**0.87***	**0.84***	0.99	0.97	1.01	0.91	**0.86**	**0.84***
Italy	0.95**	0.96**	0.97	0.95**	**0.83****	**0.86***	0.94**	0.98	1.01	0.97	**0.86***	**0.88***
Japan	0.96	0.98	0.98	1.02	1.01	1.03	0.99	1.00	0.98	1.02	1.00	1.03
UK	**0.86****	**0.84****	**0.85****	**0.79*****	**0.71*****	**0.73*****	**0.87****	**0.84****	**0.84****	**0.81*****	**0.73*****	**0.74*****
US	0.95	1.01	1.07*	0.98	1.06**	1.08***	0.98	1.06**	**1.13****	0.96	1.04*	1.07**
	Post-crisis period 2008.I-2013.IV						Post-crisis period 2008.I-2013.IV					
Canada	**1.18**	**1.13***	1.09*	0.93	1.02	1.05*	1.10***	1.10*	1.02	0.90	0.94	0.97
France	**0.88**	1.00	1.05*	0.99	1.00	0.98	0.95	1.04	**1.11****	1.00	1.03	1.03
Germany	0.94	0.94	0.97*	0.99	1.00	1.03	0.94	1.05*	1.05	1.06*	1.02	1.07*
Italy	**0.64**	**0.82**	0.91	0.90	**0.88**	1.03	**0.54**	**0.81**	**0.81**	**0.88**	**0.84***	1.05
Japan	**0.66**	**0.79**	0.97	0.96	**1.14**	1.07	**0.71**	**0.83**	0.99	1.02	**1.16***	**1.16***
UK	1.04*	1.05*	1.06**	0.99	0.99	0.98	1.06*	1.04	1.08	1.00	0.96	0.96
US	0.98	1.05*	1.10*	0.98	1.07**	1.08**	0.98	0.99	1.09	0.99	1.02	1.05***

Notes: Figures in boldface indicate a deviation by more than ± 10% from 1. Starred entries (*,**,***) denote that the one-sided Diebold-Mariano test is significant at the 10%, 5% or 1% level. The alternative hypothesis of the DM test is MFSE > 1 or MFSE < 1, depending on whether the RMSFE is greater or smaller than 1, respectively.

Table A.V: Encompassing test: Coefficients of linear combination of Consensus forecasts (β) and DFM (γ), 1999.I-2013.IV

		B2	B1	N3	N2	N1	F3
				Full sample 1999.I-2013.IV			
Canada	β	0.65***	0.74***	0.83***	0.38	0.56**	0.65**
	γ	0.35*	0.27*	0.22	0.97***	0.71***	0.77**
France	β	0.69***	0.76***	0.82***	0.66***	0.59***	0.59***
	γ	0.17	0.09	0.02	0.25**	0.32**	0.27*
Germany	β	1.15***	1.26***	1.47***	0.90*	0.92	1.01
	γ	0.49*	0.40	0.17	0.63**	0.71**	0.53*
Italy	β	0.49***	0.71***	0.71***	0.63***	0.50***	0.69***
	γ	0.77***	0.63***	0.75***	0.84***	1.05***	0.73**
Japan	β	0.32	0.33	0.52**	0.36	0.40	0.51
	γ	0.89***	0.76***	0.69**	0.67	0.24	0.04
UK	β	1.06***	1.05***	1.12***	1.37***	1.40***	1.40***
	γ	0.27	0.30	0.20	0.49**	0.42**	0.41**
US	β	0.57***	0.79***	1.10***	0.46**	0.77***	0.90**
	γ	0.49***	0.28	-0.11	0.70***	0.37	0.21
				Pre-crisis period 1999.I-2007.IV			
Canada	β	-0.25	-0.02	0.26	-0.28	-0.43	-0.51
	γ	0.94***	0.79***	0.53**	0.64**	0.69**	0.75*
France	β	0.58***	0.73***	0.77***	0.62*	0.45	0.33
	γ	0.28	0.12	0.05	0.13	0.19	0.36
Germany	β	0.72***	0.70***	0.80***	0.33	0.17	0.07
	γ	0.20	0.29	0.03	0.36	0.60**	0.78**
Italy	β	0.33	0.44*	0.50*	0.50*	0.08	0.15
	γ	0.67	0.53	0.43	0.34	0.92**	0.67
Japan	β	0.60*	0.64*	0.60*	0.79*	0.75	0.65
	γ	0.75*	0.71	0.70	0.52	0.64	-0.12
UK	β	0.41***	0.44***	0.46***	0.53***	0.40**	0.35***
	γ	0.51***	0.54***	0.54***	0.67***	0.85***	0.91***
US	β	0.35	0.63	1.09**	-0.02	0.38	0.65
	γ	0.74*	0.42	-0.19	0.82**	0.30	-0.18
				Post-crisis period 2008.I-2013.IV			
Canada	β	1.13***	1.18***	1.13***	0.85**	1.19***	1.27***
	γ	-0.07	-0.13	-0.10	0.99**	0.38	0.49
France	β	0.88*	1.00**	1.21***	0.40	0.30	0.32
	γ	0.09	0.01	-0.13	0.30*	0.38*	0.31*
Germany	β	2.19***	2.38***	2.65***	2.43*	2.56*	3.03*
	γ	0.11	-0.04	-0.28	0.45	0.40	0.08
Italy	β	0.35	0.74**	0.55*	0.24	0.31	0.57
	γ	0.79***	0.60**	0.84**	0.98***	1.08***	0.74*
Japan	β	0.05	-0.02	0.51	-0.14	0.28	0.68
	γ	1.02***	0.94**	0.65	0.73	0.14	-0.22
UK	β	1.16***	1.13***	1.24***	1.54***	1.62***	1.63***
	γ	0.16	0.20	0.06	0.44*	0.27	0.23
US	β	0.41*	0.66***	1.09***	0.54**	0.72*	0.83*
	γ	0.58***	0.38*	-0.06	0.69***	0.43	0.36

Notes: Entries refer to (unconstrained) OLS estimates. Starred entries (*,**,***) denote that the estimate is statistically different from 0 at the 10%, 5% or 1% level on the basis of the two-sided t-test.

Previous DNB Working Papers in 2016

No. 493 **Jacob Bikker, Dirk Gerritsen and Steffie Schwillens,** Competing for savings: how important is creditworthiness during the crisis?
No. 494 **Jon Danielsson and Chen Zhou**, Why risk is so hard to measure
No. 495 **Gabriele Galati, Irma Hindrayanto, Siem Jan Koopman and Marente Vlekke**, Measuring financial cycles with a model-based filter: Empirical evidence for the United States and the euro area
No. 496 **Dimitris Christelis, Dimitris Georgarakos, Tullio Jappelli and Maarten van Rooij**, Consumption uncertainty and precautionary saving
No. 497 **Marco Hoeberichts and Ad Stokman**, Price level convergence within the euro area: How Europe caught up with the US and lost terrain again
No. 498 **Janko Cizel, Jon Frost, Aerdt Houben and Peter Wierts**, Effective macroprudential policy: Cross-sector substitution from price and quantity measures
No. 499 **Frank van der Horst, Martina Eschelbach, Susann Sieber and Jelle Miedema**, Does banknote quality affect counterfeit detection? Experimental evidence from Germany and the Netherlands
No. 500 **Jochen Mierau and Mark Mink**, A descriptive model of banking and aggregate demand
No. 501 **Clemens Bonner, Daniel Streitz and Michael Wedow**, On the differential impact of securitization on bank lending during the financial crisis
No. 502 **Mijntje Lückerath-Rovers and Margriet Stavast-Groothuis**, The changing composition of the supervisory boards of the eight largest banks and insurers during 2008-2014 and the impact of the "4+4 suitability screenings"
No. 503 **Dirk Broeders, Damiaan Chen, Peter Minderhoud and Willem Schudel**, Pension funds' herding
No. 504 **Ronald Heijmans, Richard Heuver and Zion Gorgi**, How to monitor the exit from the Eurosystem's unconventional monetary policy: Is EONIA dead and gone?
No. 505 **Steven Ongena, Alexander Popov and Neeltje Van Horen**, The invisible hand of the government: "Moral suasion" during the European sovereign debt crisis
No. 506 **Wändi Bruine de Bruin, Wilbert van der Klaauw, Maarten van Rooij, Federica Teppa and Klaas de Vos**, Measuring expectations of inflation: Effects of survey mode, wording, and opportunities to revise

DeNederlandscheBank

EUROSYSTEEM

De Nederlandsche Bank N.V.
Postbus 98, 1000 AB Amsterdam
020 524 91 11
dnb.nl