Nowcasting GDP in real time: a density combination approach

Knut Are Aastveit
Norges Bank

Karsten R. Gerdrup
Norges Bank

Anne Sofie Jore
Norges Bank

Leif Anders Thorsrud
BI Norwegian Business School
Norges Bank

This is the authors' accepted and refereed manuscript to the article published in

Journal of Business and Economic Statistics, **32(2014)1: 48-68**

DOI: http://dx.doi.org/10.1080/07350015.2013.844155

Nowcasting GDP in Real-Time: A Density Combination Approach *

Knut Are Aastveit† Karsten R. Gerdrup‡ Anne Sofie Jore§ Leif Anders Thorsrud¶

August 15, 2013

Abstract

In this paper, we use U.S. real-time data to produce combined density nowcasts of quarterly GDP growth, using a system of three commonly used model classes. We update the density nowcast for every new data release throughout the quarter, and highlight the importance of new information for nowcasting. Our results show that the logarithmic score of the predictive densities for U.S. GDP growth increase almost monotonically, as new information arrives during the quarter. While the ranking of the model classes changes during the quarter, the combined density nowcasts always perform well relative to the model classes in terms of both logarithmic scores and calibration tests. The density combination approach is superior to a simple model selection strategy and also performs better in terms of point forecast evaluation than standard point forecast combinations.

JEL-codes: C32, C52, C53, E37, E52.

Keywords: Density combination; Forecast densities; Forecast evaluation; Monetary policy; Nowcasting; Real-time data

---

*We thank John Geweke, Ed Herbst, Francesco Ravazzolo, Shaun Vahey, Simon van Norden and Ken Wallis as well as seminar and conference participants at the Allied Social Science Associations Annual Meeting in Chicago, the Federal Reserve Bank of Cleveland, the Erasmus University Rotterdam, the Workshop on Central Bank Forecasting at the Federal Reserve Bank of Kansas City, the 19th Symposium of the Society of Nonlinear Dynamics and Econometrics in Washington D.C., the 31st International Symposium on Forecasting in Prague, the 65th European Meeting of the Econometric Society in Oslo, the 7th Annual CIRANO Workshop on Data Revision in Macroeconomic Forecasting and Policy in Montreal and the 5th CSDA International Conference on Computational and Financial Econometrics in London for helpful comments. The views expressed in this paper are those of the authors and should not be attributed to Norges Bank.

† *Corresponding author*: Norges Bank, Email: Knut-Are.Aastveit@norges-bank.no

‡ Norges Bank, Email: Karsten.Gerdrup@norges-bank.no

§ Norges Bank, Email: Anne-Sofie.Jore@norges-bank.no

¶ BI Norwegian Business School and Norges Bank, Email: Leif.A.Thorsrud@bi.no

1 Introduction

Economic decision making in real time is based on assessments of the recent past and current economic conditions, under a high degree of uncertainty. Many key statistics are released with a long delay, are subsequently revised and are available at different frequencies. As a consequence, there has been substantial interest in developing a framework for forecasting the present and recent past, i.e., nowcasting (see Banbura et al. (2011) for a survey of nowcasting).

Until recently, the academic literature on nowcasting has focused on developing single models that increase forecast accuracy in terms of point nowcast (see, among others, Evans (2005), Giannone et al. (2008) and Kuzin et al. (2011)). This differs in two important ways from economic decision making in practice.

First, as the data generating process is unknown and likely to change over time, decision makers are often given several different models that may produce different forecasts. This naturally leads to the question of what forecast or combination of forecasts should be used. The idea of combining forecasts of different models was first introduced by Bates and Granger (1969). Timmermann (2006) provides an extensive survey of different combination methods.

Second, if the decision maker's loss function is not quadratic, then it no longer suffices to focus solely on first moments of possible outcomes (point forecasts). To ensure appropriate decision making, the decision maker should be given suitable characterizations of forecast uncertainty. Density forecasts provide an estimate of the probability distribution of forecasts. Gneiting (2011) discusses in detail the difference between point forecasting and density forecasting, while Mitchell and Hall (2005) and Hall and Mitchell (2007) provide a justification for density combination.

In this paper, we combine density nowcasts of U.S. GDP growth from three different model classes: bridge equation models, factor models and mixed-frequency vector autoregressive (VAR) models, all widely used for short-term macroeconomic forecasting. More precisely, we extend the use of bridge equation models, as in Angelini et al. (2011), factor models, as in Giannone et al. (2008) and mixed-frequency VAR models, as in Kuzin et al. (2011), to produce density nowcasts for a wide range of different model specifications within each model class. Our recursive nowcasting exercise is applied to U.S. real-time data. We update the density nowcasts for every new data release during a quarter and highlight the importance of new data releases for the evaluation period 1990Q2-2010Q3.

The density nowcasts are combined in a two-step procedure. In the first step, nowcasts from all individual models within a model class are combined, using their logarithmic scores (log score) to compute their weights (see, among others, Jore et al. (2010)). This yields a combined density nowcast for each of the three model classes. In a second step, these three predictive densities are combined into a single density nowcast, again using log score weights. The advantages of this approach are that it explicitly accounts for uncertainty of model specification and instabilities within each model class, and that it implicitly gives a priori equal weight to each model class. We evaluate our density nowcasts both in terms of scoring rules and the probability integral transforms, to check whether predictive densities are accurate and well-calibrated.

Our novel approach of combining density nowcasts from different model classes extends the findings of earlier nowcasting and forecast combination literature in several ways.

First, we show that the log scores of the final combined predictive densities, as well as the predictive densities of the three model classes, increase almost monotonically as new information arrives during the quarter. The final combined densities seem well-calibrated. Our exercise is close to that of Giannone et al. (2008), who apply a dynamic factor model, showing that the root mean square forecasting error decreases monotonically with each data release. The importance of using non-synchronous data releases (the jagged edge problem) for point nowcasting have also been highlighted by among others Evans (2005) and Banbura and Rünstler (2011).

Second, we show that while the ranking of the model classes changes during the quarter, as new data are released, the final combined density nowcast always performs well relative to the model classes. Furthermore, our combined density nowcasts outperform nowcasts based on a simple selection strategy. This result extends the findings reported in, e.g., Rünstler et al. (2009), who study point forecasts and model selection strategies.

Third, the density combination framework also performs better than standard point forecast combination methods in terms of point forecast evaluation (see e.g. Faust and Wright (2009), for a real-time application of combining point forecasts). We show that, as new information arrives throughout the quarter, the log score weights change more rapidly than standard point forecast weights (e.g., inverse MSE weights and equal weights).

The two papers most closely related to ours are Kuzin et al. (2013) and Mitchell et al. (2013). Kuzin et al. (2013) study pooling versus model selection for nowcasting, finding that pooling provides more stable and, in most cases, better point nowcasts than model selection. Our analysis confirms these results, when evaluating density nowcasts for GDP growth utilizing 120 variables, grouped into 15 data block releases, during each month of a quarter. Mitchell et al. (2013) combine a small set of leading indicator models to construct density nowcasts for Euro-area GDP growth. They focus particularly on the ability to probabilistically anticipate the 2008-2009 Euro area recession. Our approach differs from theirs, as we combine a wide set of models and study in detail the importance of monthly data releases over a 20-year period.

Finally, our study also has similarities with, and supplement the findings of, e.g., Bache et al. (2011) and Geweke and Amisano (2011), all of whom combine density forecasts from different types of models, but do not study nowcasting.

Our results are robust to various robustness checks. Computing model weights and evaluating final densities using different real-time data vintages do not alter the qualitative results. While the nowcasting performance of different model classes may vary according to benchmark vintage, the combined density nowcast always performs well. Changing the weighting scheme by using a one-step procedure, and/or equal weights, has no effect on our conclusions: performance improves almost monotonically throughout the quarter as new information becomes available, and the combination approach is still superior to the selection strategy.

The remainder of the paper is organized as follows. In the next section, we describe the real-time data set. In the third section, we describe the modeling framework and discuss the rationale for combining densities of different model classes, while in the fourth section we describe the recursive forecasting exercise. In the fifth section, we present the results of our out-of-sample nowcasting experiment. Finally, we conclude in the sixth section.

2 Data

Our aim is to evaluate the current quarter density nowcast of the quarterly growth rate of GDP. Accordingly, in our forecasting experiment, we consider 120 monthly leading indicators, x_{i,t_m}, for $i = 1, \ldots, 120$, to nowcast quarterly growth in U.S. GDP, y_{t_q}.

The monthly data are mainly collected from the ALFRED (ArchivaL Federal Reserve Economic Data) database maintained by the Federal Reserve Bank of St. Louis. This database consists of collections of real-time vintages of data for each variable. Vintages vary across time as either new data are released or existing data are revised by the relevant statistical agency. Using data from this database ensures that we are using only data that were available on the date of the forecast origin. In addition, several real-time data series are collected from the Federal Reserve Bank of Philadelphia's Real-Time Data Set for Macroeconomists. Only quarterly vintages are available for these series, where each vintage reflects the information available around the middle of the respective quarter. Croushore and Stark (2001) provide a description of the database.

Some of the series we use, for example financial market data, are not revised. Other variables, such as consumer prices and most survey data, only undergo revisions due to changes in seasonal factors. When real-time vintage data are not available for these variables, we use the last available data vintage as their real-time observations. These data series are collected from Reuters EcoWin. Series such as equity prices, dividend yields, currency rates, interest rates and commodity prices are constructed as monthly averages of daily observations. Finally, for some series, such as disaggregated measures of industrial production, real-time vintage data exist only for parts of the evaluation period. For such variables, we use the first available real-time vintage and truncate these series backwards recursively. A more detailed description of the data series and the availability of real-time vintages are given in the appendix, section C.

The full forecast evaluation period runs from 1990Q2 to 2010Q3. We use monthly real-time data with quarterly vintages from 1990Q3 to 2010Q4, i.e., we abstract from data revisions in the monthly variables within a quarter. Hence, the quarterly vintages reflect information available just before the first release of the GDP estimate. The starting point of the estimation period is 1982M1. A key issue in this exercise is the choice of a benchmark for the "actual" measure of GDP. Stark and Croushore (2002) discuss three alternative benchmark data vintages: the most recent data vintage, the last vintage before a structural revision (called a benchmark vintage) and finally the estimate that is released a fixed period of time after the first release. We follow Romer and Romer (2000) in using the second available estimate of GDP as the actual measure. We have also computed results using the fifth release and last available vintage of GDP, finding that qualitatively there are no major changes (see section 5.4). The nowcasting exercise is described in more detail in section 4.

3 Forecast framework

Combining density forecasts is a rather new field of study in economics. The novel aspect of our study is that we combine predictive densities for nowcasting. As we nowcast quarterly U.S. GDP growth on the basis of the flow of information that becomes available during the quarter, the individual models in the forecast framework must accommodate both missing observations and time aggregations from monthly to quarterly frequencies. We use a system of three different model classes suitable to this task: bridge equation models (Bridge), mixed-frequency VARs (MF-VAR), and factor models (FM). Lately, increased interest has also been given to mixed data sampling (MIDAS) models (see, among others, Clements and Galvão (2008, 2009), Ghysels and Wright (2009) and Kuzin et al. (2011)). We abstract from this type of model in our combination framework, since the scope of models is already fairly exhaustive, and the MIDAS approach has not yet been extended to density forecasting.

For each model class, there is considerable uncertainty regarding specification, for example, choice of lag length, which variables to include, number of factors, etc. Recent work by Clark and McCracken (2009, 2010) shows that VARs may be prone to instabilities. The authors suggest combining forecasts from a wide range of VAR specifications to circumvent these problems. In our application, we include a wide range of specifications for each of the three model classes.

In total, we include 244 individual models, distributed unevenly among the three model classes. Importantly, each individual model must produce density forecasts. We do this

Table 1. A summary of all models and model classes

Model class	Description	Models
Bridge	Bivariate bridge equation models with GDP growth and different monthly indicators	120
	Lag length: 1	
	Transformation of monthly indicators: First differences or log differences	
	Estimation period: Recursive sample	
	Combination method: Linear opinion pool and log score weights	
FM	Dynamic Factor Models	4
	Number of factors: $1-4$	
	Transformation of monthly indicators: First differences or log differences	
	Estimation period: Recursive sample	
	Combination method: Linear opinion pool and log score weights	
MF-VAR	Bivariate mixed-frequency VARs with GDP growth and different monthly indicators	120
	Lag-length: 1	
	Transformation of monthly indicators: First differences or log differences	
	Estimation Period: Recursive sample	
	Combination method: Linear opinion pool and log score weights	
Combination	Combination method: Linear opinion pool and log score weights	244

Note: Each model class is described in more detail in the appendix A. The estimation period begins in 1982M1, for all models.

using bootstrapping techniques that account for both parameter and forecast uncertainty. Table 1 provides a short overview of the different specifications within each model class, while appendix A summarizes the estimation and simulation procedures. Details about the different model classes can be found in the appendix and in Angelini et al. (2011) (Bridge), Giannone et al. (2008) (FM), and Kuzin et al. (2011) (MF-VAR).

We combine the forecasts in two steps (see Garratt et al. (2009) and Bache et al. (2011) for a similar procedure). In the first step, nowcasts from all individual models within a model class are combined. This yields one combined predictive density for each model class. In the second step, we combine density nowcasts from the three model classes to obtain a single combined density nowcast. An advantage of this approach is that it explicitly accounts for uncertainty about model specification and instabilities within each model class. Hence, our predictive densities for each model class will be more robust to mis-specification and instabilities than if we were to follow the common approach in which only one model from each model class is used. Further, the two-step procedure ensures that, a priori, we put equal weight on each model class. Our approach is close to Aiolfi and

Timmermann (2006) in the sense that we combine forecasts in more than one step. They find that forecasting performance can be improved by first sorting models into clusters based on their past performance, then pooling forecasts within each cluster, and finally estimating weights for the clusters.

3.1 Combining predictive densities

To combine density forecasts, we employ the linear opinion pool:

$$p(y_{\tau,h}) = \sum_{i=1}^{N} w_{i,\tau,h}\, g(y_{\tau,h}|I_{i,\tau}), \quad \tau = \underline{\tau}, ..., \overline{\tau} \tag{1}$$

where N denotes the number of models to combine, $I_{i,\tau}$ is the information set used by model i at time τ to produce the density forecast $g(y_{\tau,h}|I_{i,\tau})$ for variable y at forecasting horizon h. $\underline{\tau}$ and $\overline{\tau}$ are the periods over which the individual densities are evaluated, and $w_{i,\tau,h}$ are a set of time-varying non-negative weights that sum to unity.

Combining N density forecasts according to equation 1 can potentially produce a combined density forecast with characteristics quite different from those of the individual densities. As Hall and Mitchell (2007) note, if all the individual densities are normal, but have different mean and variance, the combined density forecast using the linear opinion pool will be mixture normal. This distribution can accommodate both skewness and kurtosis and be multimodal (see Kascha and Ravazzolo (2010)). If the true unknown density is non-normal, this is an appealing feature. As the combined density is a linear combination of all the individual densities, the variance of the combined density forecast will generally, and more realistically, be higher than that of the individual models. The reason for this is that the variance of the combination is a weighted sum of a measure of model uncertainty and dispersion of (or disagreement about) the point forecast (see Wallis (2005)).

We follow Jore et al. (2010) in constructing weights, $w_{i,\tau,h}$, based on the logarithmic scores (log scores) of the individual models' predictive densities. A log score is the logarithm of a probability density function evaluated at the outturn of the forecast, providing an intuitive measure of density fit. Hoeting et al. (1999) also argue that the log score can be seen as a combined measure of bias and calibration. More specifically, the weights for the h-step ahead densities can be expressed as:

$$w_{i,\tau,h} = \frac{exp[\sum_{\underline{\tau}-1}^{\tau-h} ln\ g^*(y_{\tau,h}|I_{i,\tau})]}{\sum_{i=1}^{N} exp[\sum_{\underline{\tau}-1}^{\tau-h} ln\ g^*(y_{\tau,h}|I_{i,\tau})]}, \quad \tau = \underline{\tau}, ..., \overline{\tau}, \quad \text{and} \quad \underline{\tau} > h \tag{2}$$

where τ, h, y, N, i are defined above. $g^*(y_{\tau,h}|I_{i,\tau})$ is the probability density function evaluated at the outturn, $y_{\tau,h}$, of the density forecast, $g(y_{\tau,h}|I_{i,\tau})$, and $\underline{\tau} - 1$ to $\underline{\tau}$ comprises the training period used to initialize the weights, i.e., we use only the first period as a training sample. Two points are worth emphasizing: the weights are derived based on out-of-sample performance, and the weights are horizon specific.

Weighting schemes based on the log score have frequently been discussed and employed in the density combination literature (see Amisano and Giacomini (2007), Geweke and Amisano (2011), Kascha and Ravazzolo (2010), Bjørnland et al. (2011) and Mitchell and Wallis (2011)). Hall and Mitchell (2007) show that by maximizing the log score, the weights in equation 2 will minimize the Kullback-Leibler divergence between the combined density forecast and the "true," but unobserved density. As our focus is on density combination, this is an appealing feature. However, we also consider equally-weighted combinations and weights derived from the sum of squared forecast errors (SSE). For point forecast combinations, these weighting schemes have been found to work well, both empirically and theoretically (see, e.g., Clemen (1989), Stock and Watson (2004), and Bates and Granger (1969)).

3.2 Evaluating density forecasts

We evaluate the (combined) density forecasts by computing the average log score over the evaluation sample, and by testing forecast accuracy relative to the "true," but unobserved, density using the probability integral transforms (pits). As described above, the (average) log score is an intuitive measure of density fit, while the pits summarize the properties of the densities and may help us judge whether the densities are biased in a particular direction and whether the width of the densities have been roughly correct on average. More precisely, the pits represent the ex-ante inverse predictive cumulative distributions, evaluated at the ex-post actual observations.

We gauge calibration by examining whether the pits are uniform and identically and (for one-step ahead forecasts) independently distributed over the interval $[0, 1]$. Several candidate tests exist, but few offer a composite test of both uniformity and independence, as would be appropriate for one-step ahead forecasts.

Thus, we conduct several different tests. We use a test of uniformity of the pits proposed by Berkowitz (2001). The Berkowitz test works with the inverse normal cumulative density function transformation of the pits, which permits testing for normality instead of for uniformity. For one-step ahead forecasts, the null hypothesis is that the transformed pits are iid N(0,1). The test statistic is χ^2, with three degrees of freedom. For longer horizons, we do not test for independence, and thus the null hypothesis is that the transformed pits are identically standard normally distributed. The test statistics are then χ^2, with two degrees of freedom. Other tests of uniformity employed are the Anderson-Darling (AD) test (see Noceti et al. (2003)) and a Pearson chi-squared test, as suggested by Wallis (2003). Note that the latter two tests are more suitable for small samples. Independence of the pits is tested using a Ljung-Box test, based on autocorrelation coefficients of up to four for one-step ahead forecasts. For forecast horizon $h > 1$, we test for autocorrelation with lags equal to or greater than h using a modified Ljung-Box test. See Corradi and Swanson (2006) and Hall and Mitchell (2007) for more elaborate descriptions of the different tests.

Finally, note that passing the various pits tests is necessary, but not sufficient, for a forecast density to be considered the true density, conditional on the information set at the time the forecast is made.

4 Empirical exercise and ordering of data blocks

We perform a real-time out-of-sample nowcasting exercise for quarterly U.S. GDP growth for the period 1990Q2-2010Q3. The exercise is constructed as follows: For each vintage of GDP values, we estimate all models and compute density nowcasts (for all individual models, model classes and combinations) for every new data release within the quarter until publication of the first GDP estimate. This occurs approximately three weeks after the end of the quarter. By then, the nowcast will have become a backcast for that quarter.

Our dataset consists of 120 monthly variables. Series that have similar release dates and are similar in content are grouped together in blocks. The structure of the unbalancedness changes when a new block is released. In total, we have defined 15 different monthly blocks, where the number of variables in each block varies from 30, in "Labor Market," to only one, in "Initial Claims." On some dates, more than one block is released. However, our results are robust to alternative orderings of the blocks.

In Table 2, we illustrate the data release calendar and show how the 15 different blocks are released throughout each month and quarter until the first release of the GDP estimate. The table shows, for each model class, the number of individual models that update their nowcast after each new data release. It also indicates whether the GDP nowcast is a two-step ahead or a one-step ahead forecast. Nowcasts for all three model classes are updated with every new data release. However, while the nowcast of models in the FM class changes

Table 2. *Structure of data releases and models updated from the start of the quarter until the first estimate of GDP is released.*

					Number of models updated			
	Release	**Block**	**Time**	**Horizon**	**Bridge**	**FM**	**MF-VAR**	**Combination**
Nowcast	1	Interest rate	January	2	3	4	3	10
	2	Financials		2	12	4	12	28
	3	Surveys 2		2	6	4	6	16
	4	Labor market		2	30	4	30	64
	5	Money & Credit		2	2	4	2	8
	6	Mixed 1		2	5	4	5	14
	7	Ind. Production		2	16	4	16	36
	8	Mixed 2		2	11	4	11	26
	9	PPI		2	7	4	7	18
	10	CPI		2	13	4	13	30
	11	GDP		1	120	4	120	244
	12	GDP & Income		1	7	4	7	18
	13	Housing		1	3	4	3	10
	14	Survey 1		1	4	4	4	16
	15	Initial Claims		1	1	4	1	6
	16	Interest rate	February	1	3	4	3	10
	17	Financials		1	12	4	12	28
	18	Surveys 2		1	6	4	6	16
	19	Labor market		1	30	4	30	64
	20	Money & Credit		1	2	4	2	8
	21	Mixed 1		1	5	4	5	14
	22	Ind. Production		1	16	4	16	36
	23	Mixed 2		1	11	4	11	26
	24	PPI		1	7	4	7	18
	25	CPI		1	13	4	13	30
	26	GDP		1				
	27	GDP & Income		1	7	4	7	18
	28	Housing		1	3	4	3	10
	29	Survey 1		1	4	4	4	16
	30	Initial Claims		1	1	4	1	6
	31	Interest rate	March	1	3	4	3	10
	32	Financials		1	12	4	12	28
	33	Surveys 2		1	6	4	6	16
	34	Labor market		1	30	4	30	64
	35	Money & Credit		1	2	4	2	8
	36	Mixed 1		1	5	4	5	14
	37	Ind. Production		1	16	4	16	36
	38	Mixed 2		1	11	4	11	26
	39	PPI		1	7	4	7	18
	40	CPI		1	13	4	13	30
	41	GDP		1				
	42	GDP & Income		1	7	4	7	18
	43	Housing		1	3	4	3	10
	44	Survey 1		1	4	4	4	16
	45	Initial Claims		1	1	4	1	6
Backcast	46	Interest rate	April	1	3	4	3	10
	47	Financials		1	12	4	12	28
	48	Surveys 2		1	6	4	6	16
	49	Labor market		1	30	4	30	64
	50	Money & Credit		1	2	4	2	8
	51	Mixed 1		1	5	4	5	14
	52	Ind. Production		1	16	4	16	36
	53	Mixed 2		1	11	4	11	36
	54	PPI		1	7	4	7	18
	55	CPI		1	13	4	13	30

Note: The table illustrates a generic quarter of real-time out-of-sample forecasting experiments. Our forecast evaluation period runs from 1990Q2 to 2010Q3, which gives us more than 80 observations to evaluate, for each data release. All models that are updated are re-estimated at each point in time throughout the quarter. In total, we re-estimate and simulate (bootstrap) the individual models 2,000 times for every block in a given quarter.

with every data release (because the factors are affected), the nowcasts of models of the Bridge and MF-VAR classes only change if the newly released data contains information that has historically improved the log score, that is, if models that revise their nowcasts have non-zero weight.

5 Results

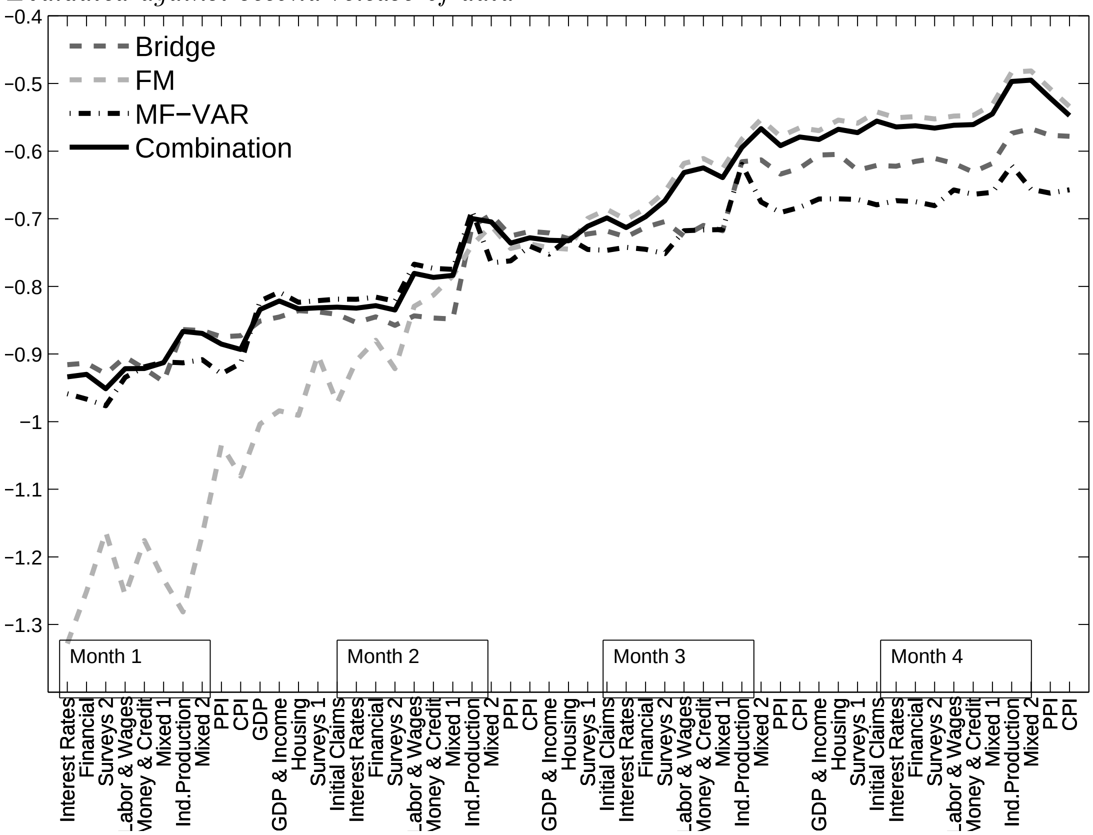
In this section, we analyze the performance of our density combination approach. The main goal of the exercise is to examine how the predictive densities improve as more data are available during the quarter. In doing so, we wish to evaluate both the accuracy of the density nowcasts (section 5.1) and whether they are well-calibrated (section 5.2).

5.1 Nowcast accuracy

We study the impact of different data releases on density nowcasting accuracy, measured by average log score. Figure 1 depicts the end of sample average log scores for the combined density nowcasts and for the three model classes after each data block release, over the period between the beginning of a quarter and the first release of GDP. The first 10 observations of the quarter are actually two-step ahead forecasts, while the 11 final observations are backcasts (see also table 2).

The figure reveals two interesting results. First, forecasting performance improves when new information becomes available. The log scores of the final combined predictive densities and of the three model classes increase as new information arrives during the quarter. Second, the ranking of the model classes changes during the quarter, as new data are released, while the combined density nowcast always performs well compared to the model classes. For example, the Bridge and MF-VAR classes outperform the FM class in the early stages of the quarter. This is not surprising, as the factor estimates are highly

Figure 1. *End of sample average log scores for forecasts after different block releases. Evaluated against second release of data*

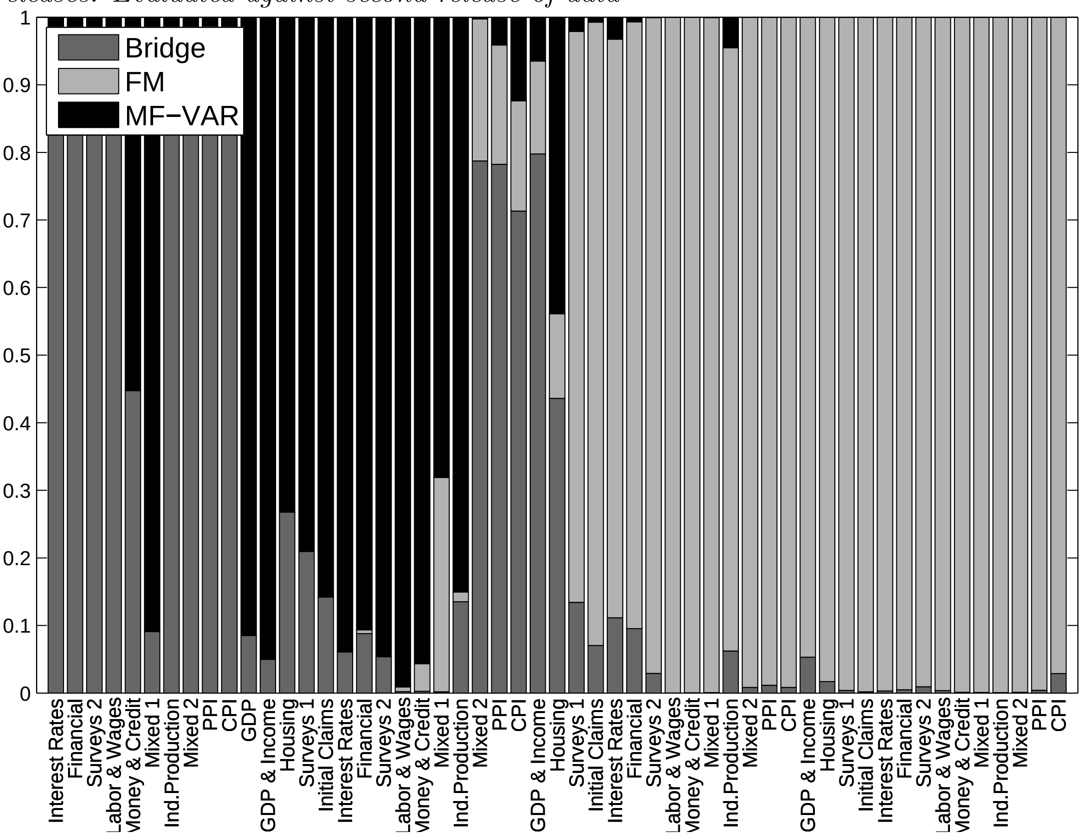


Note: The nowcasts from the individual models and model classes have been combined using the linear opinion pool and log score weights. The evaluation period runs from 1990Q2 to 2010Q3.

uncertain in the early stages of the quarter. However, as new information continues to arrive, factor uncertainty decreases and the relative performance of the FM class improves. Towards the end of the quarter, the FM class is the best performing model class.

Figure 2 shows how the weights attached to each model class in the combined density nowcast change after each data block release. The figure illustrates the weights at the end of the evaluation period. As the weights are based on past log score performance, the same pattern as that observed in the average log score comparison arises. That is, the Bridge and MF-VAR class have high weight in the early periods of the quarter, while the FM class winds up having nearly all the weight towards the end of the quarter. The reader, however, should not interpret this as attaching all weight to one unique model, as the FM class is in fact a combination of four factor models. Finally, note that the average log score of the combined density nowcast is almost identical to that of the best performing model class throughout the quarter. This illustrates the main advantage of using forecast combinations.

Figure 2. *End of sample weights attached to the different model classes after different block releases. Evaluated against second release of data*



Note: The nowcasts from the individual models and model classes have been combined using the linear opinion pool and log score weights. The evaluation period runs from 1990Q2 to 2010Q3.

5.2 Calibration

We evaluate the predictive densities relative to the "true," but unobserved, density, using the pits (see figure 3). Table 3 shows p-values for the four different tests described in section 3.2, applied to the combined forecast at five different points in time. The latter correspond to the start of the first month (Block 1), the end of the first month (Block 15), the end of the second month (Block 30), the end of the third month (Block 45) and the middle of the fourth month (Block 55). P-values equal to or higher than 0.05 mean that we cannot reject, at the 5% significance level, the hypothesis that the combined predictive density is correctly calibrated.

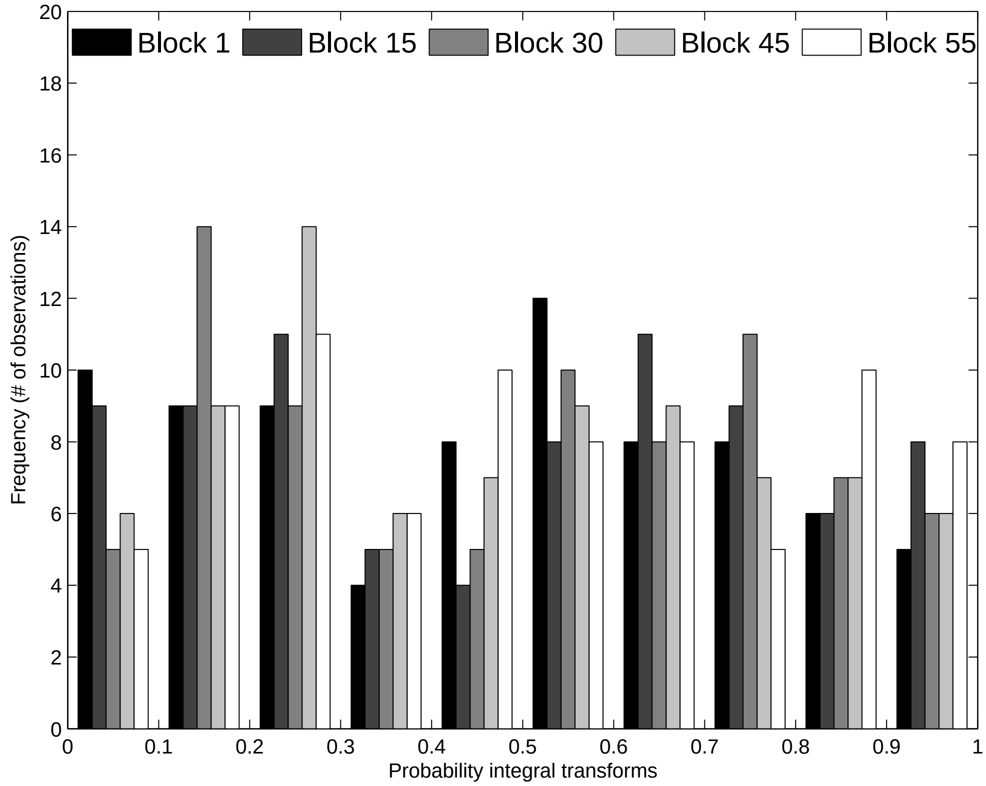
Table 3. Pits tests for evaluating density forecasts for GDP growth

Block	h	Berkowitz	χ^2	LB1	LB2	LB3	Anderson-Darling
Block 1	2	**0.13**	**0.72**	**0.94**	**0.73**	**0.80**	**1.03**
Block 15	1	**0.29**	**0.13**	**0.72**	**0.60**	**0.58**	**0.56**
Block 30	1	**0.70**	**0.59**	**0.16**	**0.83**	**0.20**	**0.55**
Block 45	1	**0.65**	**0.49**	**0.07**	**0.60**	**0.15**	**0.61**
Block 55	1	**0.87**	**0.85**	0.01	**0.59**	0.03	**0.50**

Note: For Block 15, Block 30 and Block 45, the nowcast is a one-step ahead forecast, while it is a two-step ahead forecast for Block 1. Block 55 is a one-step ahead backcast. All numbers are p-values, except for the Anderson - Darling test. The null hypothesis in the Berkowitz test is that the inverse normal cumulative distribution function transformed pits are $idN(0,1)$*, and for* $h = 1$ *are independent.* χ^2 *is the Pearson chi-squared test suggested by Wallis (2003) of uniformity of the pits histogram in eight equiprobable classes. LB1, LB2 and LB3 are Ljung-Box tests of independence of the pits in the first, second and third power, respectively, at lags greater than or equal to the horizon. Assuming independence of the pits, the Anderson-Darling test statistic for uniformity of the pits has a 5% small-sample (simulated) critical value of 2.5.*

The combined density nowcast, where the nowcast corresponds to a two-step ahead forecast, passes all the tests for Block 1. Turning to the one-step ahead forecast (Block 15 - Block 55), the combined density nowcast also seems to be well-calibrated. Based on the Berkowitz test, the Anderson-Darling test and the Pearson chi-squared test, we cannot reject, at a 5% significance level, the null hypothesis that the combined density is well-calibrated. One exception is that the null hypothesis in the Ljung-Box tests (LB1 and LB3) are rejected for Block 55.

Figure 3. *Pits of the combined density forecasts at five points during the quarter. The pits are the ex ante inverse predictive cumulative distributions, evaluated at the ex post actual observations.*

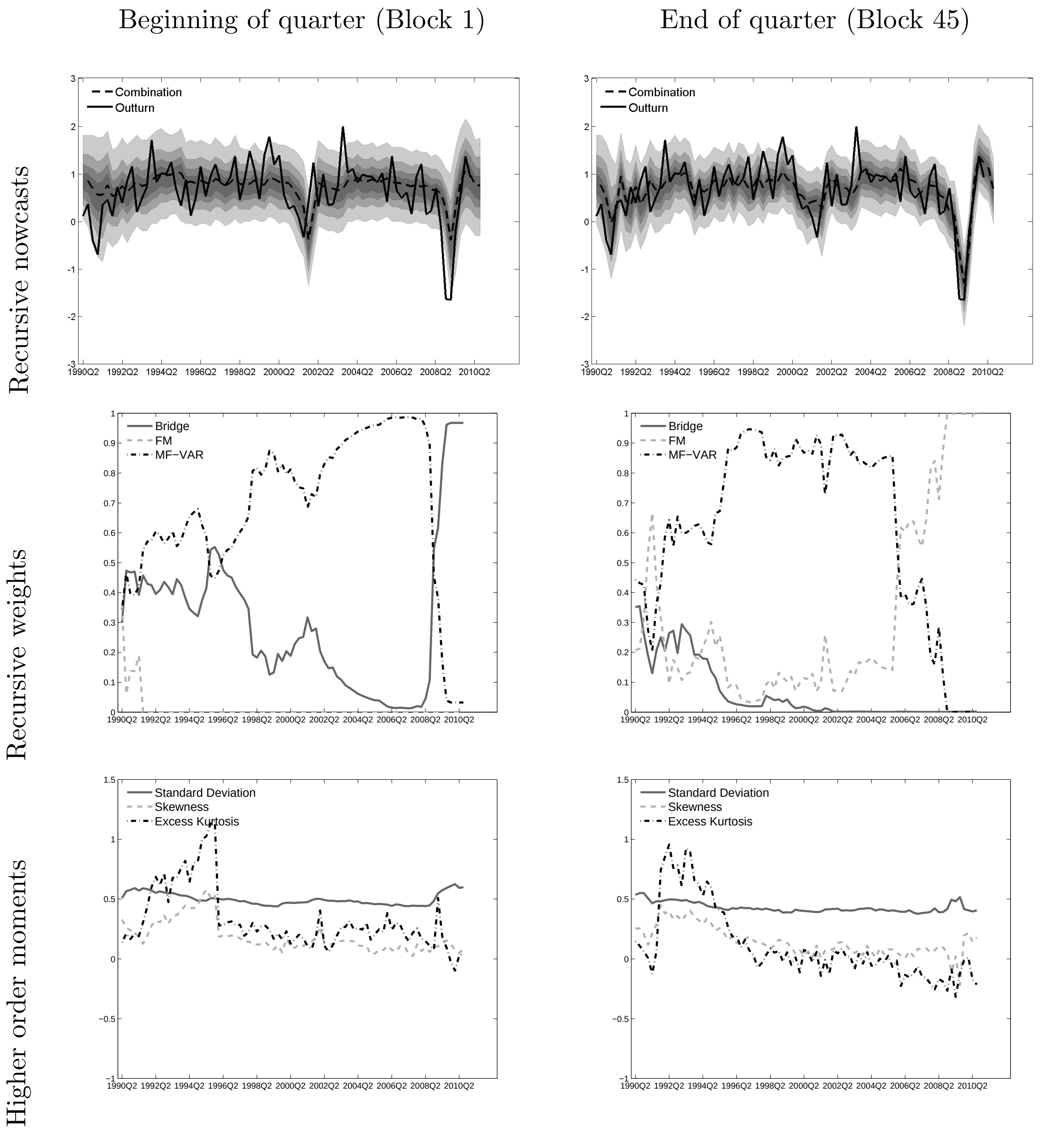


Note: The pits of predictive densities should have a standard uniform distribution, if the model is correctly specified.

5.3 Properties of the density nowcasts

Some properties of the density nowcasts are illustrated in figure 4. In the first row, the figure shows recursive real-time out-of-sample density nowcasts for U.S. GDP growth for the period 1990Q2-2010Q3. Recursive nowcasts made on the first day of the quarter (Block 1) are shown in the left panel, while recursive nowcasts made on the last day of the quarter (Block 45) are shown in the right panel. The two panels illustrate how the precision of the predictive densities improves as more information becomes available.

Figure 4. Recursive real-time out-of-sample density nowcasts for quarterly U.S. GDP



Note: Results from Block 1 are in the left column, and results from Block 45 are in the right column.

The second row in the figure shows how the recursive weights change over time. There are large movements in the weights related to the start of the Great Recession, for nowcasts made at Block 1 and Block 45. For nowcasts made at Block 45, there is also a shift in the weights during the expansion of 2006-2007. This illustrates the flexibility of our density combination framework.

Finally, as noted in section 3.1, using a linear opinion pool to combine density nowcasts may yield a predictive density that deviates from normality. The lower row in the figure shows how the behavior of higher-order moments of the combined predictive density evolve over time. The standard deviation is rather stable over time, but increases, in particular for Block 1 nowcasts, during the Great Recession. There are larger movements in skewness and excess kurtosis over time. For Block 1 and Block 45 nowcasts, there is evidence of positive skewness and positive excess kurtosis in the early parts of the sample. Also, at the end of the sample, the density nowcasts appear to deviate from normality. The movements in the higher-order moments correspond with changes in the weights attached to the different model classes.

5.4 Robustness

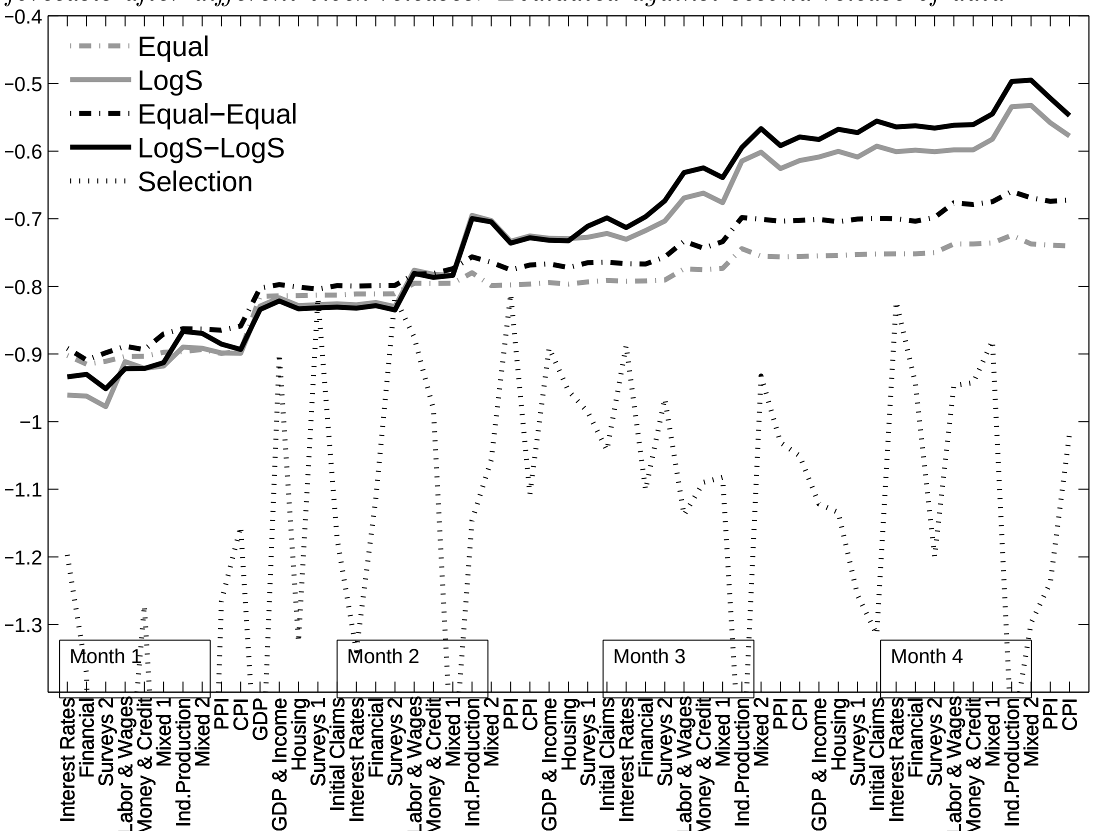
In this section, we perform three robustness checks: first, with respect to alternative weighting schemes; second, with respect to point forecasting; and finally, with respect to the choice of benchmark vintage for GDP.

5.4.1 Alternative weighting schemes for the combination

Several papers have found that simple combination forecasts, with forecasts equally weighted, outperform more sophisticated adaptive forecast combination methods. This is often referred to as the forecast combination puzzle. While Jore et al. (2010) seem to find some evidence of gains in adaptive log score weights for density combination forecasts, this remains a debated issue. We investigate robustness with respect to the following weighting schemes: 1) combination of all individual nowcasts in one step, applying equal weights (Equal); 2) combination of all individual nowcasts in one step, applying log score weights (LogS); 3) combination of nowcasts in two steps, applying equal weights in both steps (Equal-Equal); and 4) a selection strategy, where we try to pick the nowcast of the "best" model. We have constructed the latter by recursively choosing the best model among all 244 models at each point in time throughout the evaluation period, and used this model to forecast the next period. The preferred combination of nowcasts in two steps, applying log score weights in each, is denoted as LogS-LogS.

Figure 5 compares the average log scores for the different weighting schemes. We highlight five results. First, overall, all combined density nowcasts yield a steady increase in average log scores, as more information becomes available. This is not the case for the selection strategy, which produces large and volatile changes in the average log score after every data block release. Second, the selection strategy typically produces the poorest density nowcast in terms of average log score. Third, the difference between "Equal" and "Equal-Equal" can be seen as the gain from using a two-step approach, where models are first grouped into model classes and then combined. It is evident from the figure that "Equal-Equal" always performs better than "Equal." Fourth, there is less difference between "LogS" and "LogS-LogS," as the log score weights discriminate rather sharply between nowcasts of the different models. Finally, no weighting scheme is superior throughout the quarter, but our preferred two-step combination approach ("LogS-LogS") is the best performing strategy for most of the quarter.

Figure 5. *Comparing different weighting schemes. End of sample average log scores for forecasts after different block releases. Evaluated against second release of data*

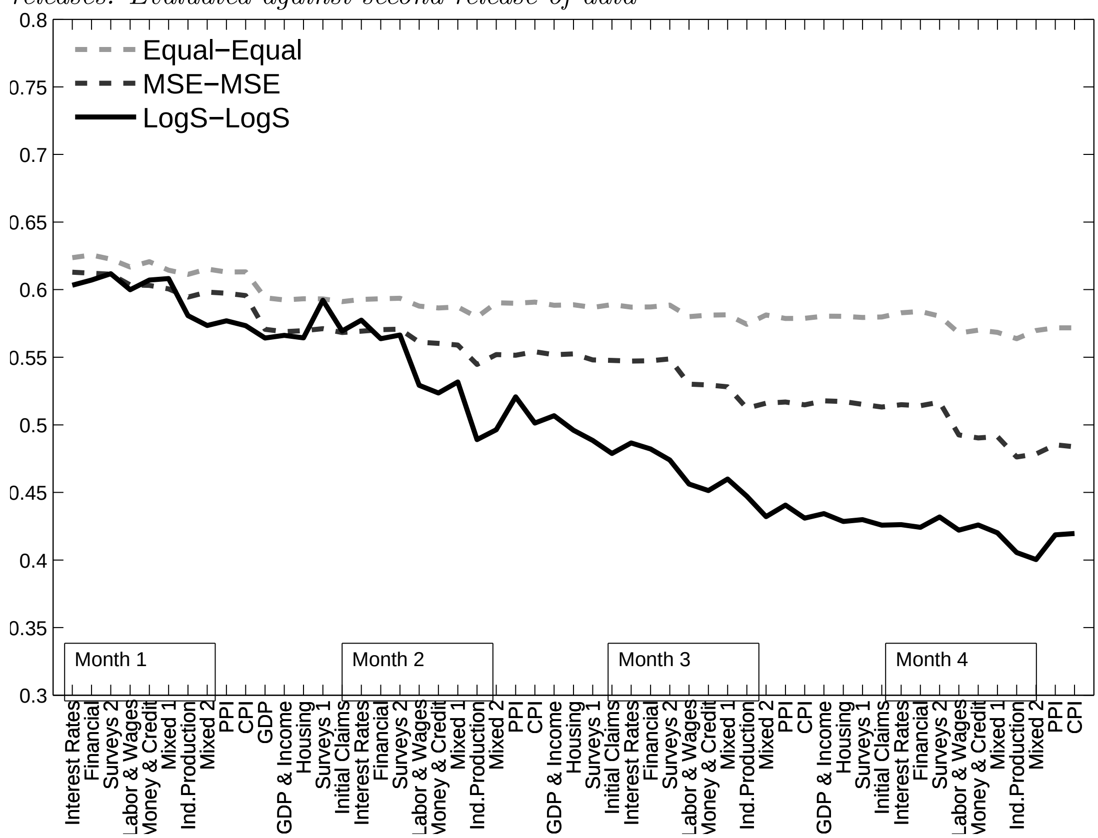


Note: Equal and LogS indicate that nowcasts from all individual models are combined in one step using the linear opinion pool and applying equal weights or log score weights, respectively. Equal-Equal and LogS-LogS indicate combination of nowcasts in two steps using the linear opinion pool and applying equal weights or log score weights, respectively. Selection refers to a strategy of "picking" the best model among all 244 models at each point in time throughout the evaluation period. The evaluation period runs from 1990Q2 to 2010Q3.

5.4.2 Point forecasting

We investigate robustness of our results by evaluating point nowcasting performance. We do this by comparing three different combination strategies. First, we use the "LogS-LogS" approach, calculating point nowcasts as the mean of the combined density nowcast. Second, we combine nowcasts in two steps, applying inverse MSE weights in both (MSE-MSE), and calculate point nowcasts. Finally, we calculate the point nowcast, using the "Equal-Equal" approach. We evaluate the point nowcasts of the three different combination approaches, using the root mean squared prediction error (RMSE). The remainder of the experiment is similar to what we have described above.

Figure 6. *Comparing different weighting schemes. RMSE for forecasts after different block releases. Evaluated against second release of data*



Note: Equal-Equal, MSE-MSE and LogS-LogS indicate that the individual models within each model class and the model classes have been combined using the linear opinion pool and equal weights, MSE weights and log score weights, respectively. The evaluation period runs from 1990Q2 to 2010Q3.

Figure 6 depicts the RMSE for the combined nowcasts, using the three strategies, after each data block release. The figure displays two key results. First, for all strategies, nowcasting errors steadily decline as more information becomes available throughout the quarter until the first GDP estimate is released. This result accords with the findings of earlier nowcasting experiments that used mean squared prediction error evaluation (see, e.g., Giannone et al. (2008)).

Second, the density combination approach ("LogS-LogS") performs better, in terms of RMSE, than the strategy of applying inverse MSE weights. As far as we are aware, this is a new finding in the nowcasting literature. We believe this result is linked to the properties of the log score weights, in particular, that they distinguish more sharply between forecasts from different models than inverse MSE weights. Comparing figure 2 and figure 7 in the appendix, we see that as new information arrives during the quarter, the log score weights adapt faster than the inverse MSE weights. This does not a priori need to improve the nowcasts, but we may argue that in our case log score weighting attaches higher weight to models with new and relevant information than alternative weighting approaches.

5.4.3 Alternative benchmark vintages

The choice of benchmark vintage is a key issue in any application using real-time vintage data (see Croushore (2006) for a survey of forecasting with real-time macroeconomic data). In our application, we use the second release of the GDP estimate as a benchmark. Figure 8 in the appendix shows results with, respectively, the fifth release of GDP and the last available vintage of GDP as benchmarks. Clearly the figures show that the nowcasting performance of the different model classes varies with the choice of benchmark vintage. Hence, the weights attached to the different model classes also vary. However, the result that the density combination nowcast always performs well is robust.

6 Conclusion

In this paper we have used a density combination framework to produce combined density nowcasts for U.S. quarterly GDP growth. We use a system of three different model classes widely used for short-term forecasting: bridge equation models, factor models and mixed-frequency VARs. The density nowcasts are combined in a two-step procedure. In the first step, nowcasts from all individual models within a model class are combined, using the log score to compute the weights. This yields a combined predictive density nowcast for each of the three different model classes. In the second step, these three predictive densities are combined into a single density nowcast, again using log score weights. The density nowcasts are updated for every new data release during a quarter until the first estimate of GDP is available. Our recursive nowcasting exercise is applied to U.S. real-time data and evaluated for the period 1990Q2-2010Q3.

We show that log scores for the predictive densities increase almost monotonically, as new information arrives during the quarter. The densities also seem well-calibrated. In addition, while the ranking of the model classes changes during the quarter as new data are released, the combined density nowcast always performs well compared to the three model classes. Finally, the density combination approach is superior to a simple model selection strategy, and the density combination framework actually performs better, in terms of point forecast evaluation, than standard point forecast combination methods.

The results are robust to the choice of benchmark (real-time) vintage. While the nowcasting performance of different model classes may vary according to benchmark vintage, the density combination nowcast always performs well.

References

Aiolfi, M. and A. Timmermann (2006). Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics 135*(1-2), 31–53.

Amisano, G. and R. Giacomini (2007). Comparing density foreasts via weighted likelihood ratio tests. *Journal of Business and Economic Statistics 25*(2), 177–190.

Angelini, E., G. Camba-Mendez, D. Giannone, L. Reichlin, and G. Rünstler (2011). Short-term forecasts of euro area GDP growth. *Econometrics Journal 14*(1), C25–C44.

Bache, I. W., A. S. Jore, J. Mitchell, and S. P. Vahey (2011). Combining VAR and DSGE forecast densities. *Journal of Economic Dynamics and Control 35*(10), 1659–1670.

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

Banbura, M., D. Giannone, and L. Reichlin (2011). Nowcasting. In M. P. Clements and D. F. Hendry (Eds.), *The Oxford Handbook of Economic Forecasting*. Oxford Handbooks in Economics.

Banbura, 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.

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

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business and Economic Statistics 19*(4), 465–474.

Bjørnland, H. C., K. Gerdrup, A. S. Jore, C. Smith, and L. A. Thorsrud (2011). Weights and pools for a Norwegian density combination. *The North American Journal of Economics and Finance 22*(1), 61–76.

Clark, T. E. and M. W. McCracken (2009). Improving forecast accuracy by combining recursive and rolling forecasts. *International Economic Review 50*(2), 363–395.

Clark, T. E. and M. W. McCracken (2010). Averaging forecasts from VARs with uncertain instabilities. *Journal of Applied Econometrics 25*(1), 5–29.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting 5*, 559–581.

Clements, M. P. and A. B. Galvão (2008). Macroeconomic forecasting with mixed-frequency data. *Journal of Business & Economic Statistics 26*, 546–554.

Clements, M. P. and A. B. Galvão (2009). Forecasting US output growth using leading indicators: an appraisal using MIDAS models. *Journal of Applied Econometrics 24*(7), 1187–1206.

Corradi, V. and N. R. Swanson (2006). *Predictive Density Evaluation*, Volume 1 of *Handbook of Economic Forecasting*, Chapter 5, pp. 197–284. Elsevier.

Croushore, D. (2006). *Forecasting with Real-Time Macroeconomic Data*, Volume 1 of *Handbook of Economic Forecasting*, Chapter 17, pp. 961–982. Elsevier.

Croushore, D. and T. Stark (2001). A real-time data set for macroeconomists. *Journal of Econometrics 105*(1), 111–130.

Evans, M. D. (2005). Where are we now? Real-time estimates of the macro economy. *International Journal of Central Banking 1*(2), 127–175.

Faust, J. and J. H. Wright (2009). Comparing Greenbook and reduced form forecasts using a large realtime dataset. *Journal of Business & Economic Statistics 27*(4), 468–479.

Garratt, A., J. Mitchell, and S. P. Vahey (2009). Measuring output gap uncertainty. Reserve Bank of New Zealand Discussion Paper Series DP2009/15, Reserve Bank of New Zealand.

Geweke, J. and G. Amisano (2011). Optimal prediction pools. *Journal of Econometrics 164*(1), 130–141.

Ghysels, E. and J. H. Wright (2009). Forecasting professional forecasters. *Journal of Business & Economic Statistics 27*(4), 504–516.

Giannone, D., L. Reichlin, and S. Simonelli (2009). Nowcasting euro area economic activity in real time: The role of confidence indicators. *National Institute Economic Review 210*(1), 90–97.

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.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association 106*(494), 746–762.

Hall, S. G. and J. Mitchell (2007). Combining density forecasts. *International Journal of Forecasting 23*(1), 1–13.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999). Bayesian model averaging: A tutorial. *Statistical Science 14*(4), 382–417.

Jore, A. S., J. Mitchell, and S. P. Vahey (2010). Combining forecast densities from VARs with uncertain instabilities. *Journal of Applied Econometrics 25*(4), 621–634.

Kascha, C. and F. Ravazzolo (2010). Combining inflation density forecasts. *Journal of Forecasting 29*(1-2), 231–250.

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 with many predictors: An application to German GDP. *Journal of Applied Econometrics* (forthcoming).

Mariano, R. S. and Y. Murasawa (2010). A coincident index, common factors, and monthly real GDP. *Oxford Bulletin of Economics and Statistics 72*(1), 27–46.

Mitchell, J. and S. G. Hall (2005). Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESR 'fan' charts of inflation. *Oxford Bulletin of Economics and Statistics 67*(s1), 995–1033.

Mitchell, J., G. L. Mazzi, and G. Montana (2013). Density nowcasts and model combination: nowcasting Euro-area GDP growth over the 2008-9 recession. *Oxford Bulletin of Economics and Statistics* (forthcoming).

Mitchell, J. and K. F. Wallis (2011). Evaluating density forecasts: forecast combinations, model mixtures, calibration and sharpness. *Journal of Applied Econometrics 26*(6), 1023–1040.

Noceti, P., J. Smith, and S. Hodges (2003). An evaluation of tests of distributional forecasts. *Journal of Forecasting 22*(6-7), 447–455.

Rünstler, G., K. Barhoumi, S. Benk, R. Cristadoro, A. D. 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.

Romer, C. D. and D. H. Romer (2000). Federal reserve information and the behavior of interest rates. *American Economic Review 90*(3), 429–457.

Stark, T. and D. Croushore (2002). Forecasting with a real-time data set for macroeconomists. *Journal of Macroeconomics 24*(4), 507–531.

Stock, J. H. and M. W. Watson (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics 20*(2), 147–62.

Stock, J. H. and M. W. Watson (2004). Combining forecasts of output growth in seven-country data set. *Journal of Forecasting 23*, 405–430.

Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1, pp. 136–96. Amsterdam: Elsevier.

Wallis, K. F. (2003). Chi-squared tests of interval and density forecasts, and the Bank of England's fan charts. *International Journal of Forecasting 19*(3), 165–175.

Wallis, K. F. (2005). Combining density and interval forecasts: A modest proposal. *Oxford Bulletin of Economics and Statistics 67*(s1), 983–994.

A Models and model classes

Table 1 in the main text lists the three model classes employed in the nowcasting experiment. A brief description of the estimation and simulation procedure for each model class is given below.

A.1 Bridge equations (Bridge)

A bridge equation is estimated from quarterly aggregates of monthly data. Monthly indicators are typically available earlier than GDP growth (as described in table 2 and appendix C). When nowcasting, we want to exploit this information. Following the notation in Kuzin et al. (2013), quarterly GDP growth is denoted y_{t_q}, where t_q is the quarterly time index $t_q = 1, 2, 3, \ldots, T_q$ and T_q is the sample length of quarterly GDP growth. GDP growth can also be expressed at a monthly frequency by setting $y_{t_m} = y_{t_q}$, which holds for the monthly time index $t_m = 3, 6, 9, \ldots, T_m$ with $T_m = 3T_q$. Let x_{i,t_m} denote a generic stationary monthly indicator, transformed to correspond to a quarterly quantity when observed at the end of the quarter, that is when $t_m = 3, 6, 9, \ldots, T_m$. Predictions of quarterly GDP growth, y_{t_q} are then obtained in two steps. First, monthly indicators x_{i,t_m} are forecasted over the remainder of the quarter using a simple univariate autoregression (AR(1)). In a second estimation step, the quarterly growth rate of GDP, y_{t_q}, is regressed on the resulting values using the bridge equation:

$$y_{t_q} = y_{t_m} = \alpha + \beta' x_{i,t_m} + e_{t_m}, \qquad e_{t_m} \sim N(0, \Sigma_e) \tag{3}$$

which holds for $t_m = 3, 6, 9, \ldots, T_m$

Forecasts are constructed, conditional on the estimated parameters and the transformed monthly indicator forecasts. See Baffigi et al. (2004) and Angelini et al. (2011) for a more detailed discussion of alternative bridge equations.

A.2 Mixed-frequency vector autoregressive models (MF-VAR)

In contrast to the bridge equation methodology, the MF-VAR methodology takes into account the possible joint dynamics between the particular indicator used and GDP growth. MF-VARs has recently attracted increased research interest (see, e.g., Giannone et al. (2009), Mariano and Murasawa (2010) and Kuzin et al. (2011)).

The intuitive appeal of the MF-VAR approach is that it operates at the highest sampling frequency of the time series included in the model, while the lower frequency variables are interpolated according to their stock-flow nature. We specify one bivariate MF-VAR model for each of the leading indicators, together with unobserved monthly GDP.

The estimation framework follows the procedure outlined in Kuzin et al. (2011), and we refer to their work for details. In brief, we work with the following time-aggregation restriction, to relate unobserved month-on-month GDP growth, $y^*_{t_m}$, to observed quarterly growth of GDP, y_{t_q}

$$y_{t_q} = y_{t_m} = \frac{1}{3}y^*_{t_m} + \frac{2}{3}y^*_{t_m-1} + y^*_{t_m-2} + \frac{2}{3}y^*_{t_m-3} + \frac{1}{3}y^*_{t_m-4} \tag{4}$$

which holds for $t_m = 3, 6, 9, \ldots, T_m$.

The joint dynamics between the latent month-on-month growth of GDP, $y^*_{t_m}$, and the corresponding monthly indicator, x_{i,t_m}, are modeled as simple bivariate VARs. Each VAR is specified with one lag only, and the forecasts are generated by iterating the VAR process forward.

As described in Kuzin et al. (2011), the high-frequency VAR, together with the time-aggregation restriction, can be cast in state-space form and estimated by maximum likelihood. In this framework, the Kalman filter can handle missing values at the end of the sample and address the mixed-frequency nature of the data.

A.3 Factor models (FM)

Factor models summarize the information contained in large datasets by reducing the parameter space (see, e.g., Stock and Watson (2002)). The factor model specification we employ is similar to Giannone et al. (2008) (see also Banbura and Rünstler (2011) for an extension. Assume we have a vector of n observable and stationary monthly variables $X_{t_m} = (x_{1,t_m}, \dots, x_{n,t_m})'$ which have been standardized to have mean equal to zero and variance equal to one. The monthly variables are transformed so as to correspond to a quarterly quantity when observed at the end of the quarter (i.e., when $t_m = 3, 6, 9, \dots, T_m$). The factor model is then given by the following two equations:

$$X_{t_m} = \chi_{t_m} + \xi_{t_m} = \Lambda F_{t_m} + \xi_{t_m}, \qquad \xi_{t_m} \sim N(0, \Sigma_\xi) \tag{5}$$

$$F_{t_m} = \sum_{i=1}^{p} A_i F_{t_m - i} + B u_{t_m}, \qquad u_{t_m} \sim N(0, I_u) \tag{6}$$

Equation 5 relates the monthly time series X_{t_m} to a common component χ_{t_m} plus an idiosyncratic component $\xi_{t_m} = (\xi_{1,t_m}, \dots, \xi_{n,t_m})'$. The former is given by an $r \times 1$ vector of latent factors $F_{t_m} = (f_{1,t_m}, \dots, f_{r,t_m})'$ times an $n \times r$ matrix of factor loadings Λ, while the latter is assumed to be multivariate white noise. Equation 6 describes the law of motion for the latent factors with lags $1, \dots, p$. The factors are driven by q-dimensional standardized white noise u_{t_m}, where B is an $r \times q$ matrix, and where $q \leq r$. Finally, $A_1, \dots, A_p$ are $r \times r$ matrices of parameters.

The factor model, equations 5 and 6, is estimated in a two-step procedure using principal components and the Kalman filter. The unbalanced part of the data set can be incorporated through the use of the Kalman filter, where missing monthly observations are interpreted as having an infinitely large noise to signal ratio. For more details about this estimation procedure, see Giannone et al. (2008).

Finally, predictions of quarterly GDP growth, y_{t_q}, are then obtained in same way as for the bridge equations. Now the monthly factors F_{t_m} are forecasted over the remainder of the quarter using equation 6. Then the quarterly growth rate of GDP, y_{t_q} is regressed on the resulting factor values, that is, by replacing x_{i,t_m} in equation 3 (bridge equation) with the factors, F_{t_m}:

$$y_{t_q} = y_{t_m} = \alpha + \beta' F_{t_m} + e_{t_m}, \qquad e_{t_q} \sim N(0, \Sigma_e) \tag{7}$$

which holds for $t_m = 3, 6, 9, \ldots, T_m$ and where β is an $r \times 1$ vector of parameters.

Forecasts of GDP growth are constructed, conditional on the estimated parameters and the forecasted factors in equation 7.

A.4 Constructing densities

To take account of both parameter and forecast uncertainty we apply a bootstrapping procedure to all models. All models can be represented in a state space form by suitable reformulations and restrictions on the different parameter vectors.

For simplicity, we describe our bootstrapping procedure in terms of the state space representation given in equations 5 and 6 and the bridge equation 7. For each model within each model class, we first follow the estimation steps described in section A.3. This yields the parameters: $A^0 = [A_1, \ldots, A_p]$, u^0, ξ^0, and Λ^0. In addition, for the Bridge equation and the Factor model class, we have α^0, β^0, and e^0. Let y_{h_q} denote the forecast vector for quarterly GDP growth h_q quarters ahead. Then, for $i = 1, ..., 2000$:

1. Simulate $\tilde{F}_{t_m} = A^0 \tilde{F}_{t_m - 1} + B u^*_{t_m}$, where $u^*_{t_m}$ is re-sampled from u^0
2. Simulate $\tilde{X}_{t_m} = \Lambda^0 \tilde{F}_{t_m} + \xi^*_{t_m}$, where $\xi^*_{t_m}$ is re-sampled from ξ^0
3. Based on $\tilde{X}_{t_m}$, re-estimate the model to get: Λ^i, F^i and A^i

4. Forecast F^i h_m-months ahead, where $h_m = 3h_q, 3h_q - 1$ or $3h_q - 2$ (depending on the number of available monthly observations), using Λ^i and A^i and take into account forecast uncertainty by re-sampling from u^0. Let $F^i_{h_m}$ denote the simulated forecasts

   - Bridge equation models and Factor models: $F^i_{h_m}$ denote, respectively, forecast of the indicator variable and of the monthly factors.
     (a) Estimate α^i, β^i based on F^i
     (b) Forecast $y^i_{h_q}$ conditional on α^i, β^i, and $F^i_{h_m}$, where forecast uncertainty is taken into account by re-sampling from e^0
   - MF-VAR class: $F^i_{h_m}$ denote forecasts of month-to-month GDP growth. Forecasts of quarterly GDP growth can be obtained by feeding $F^i_{h_m}$ through the observation equation to get $y^i_{h_q}$
   - Save $y^i_{h_q}$

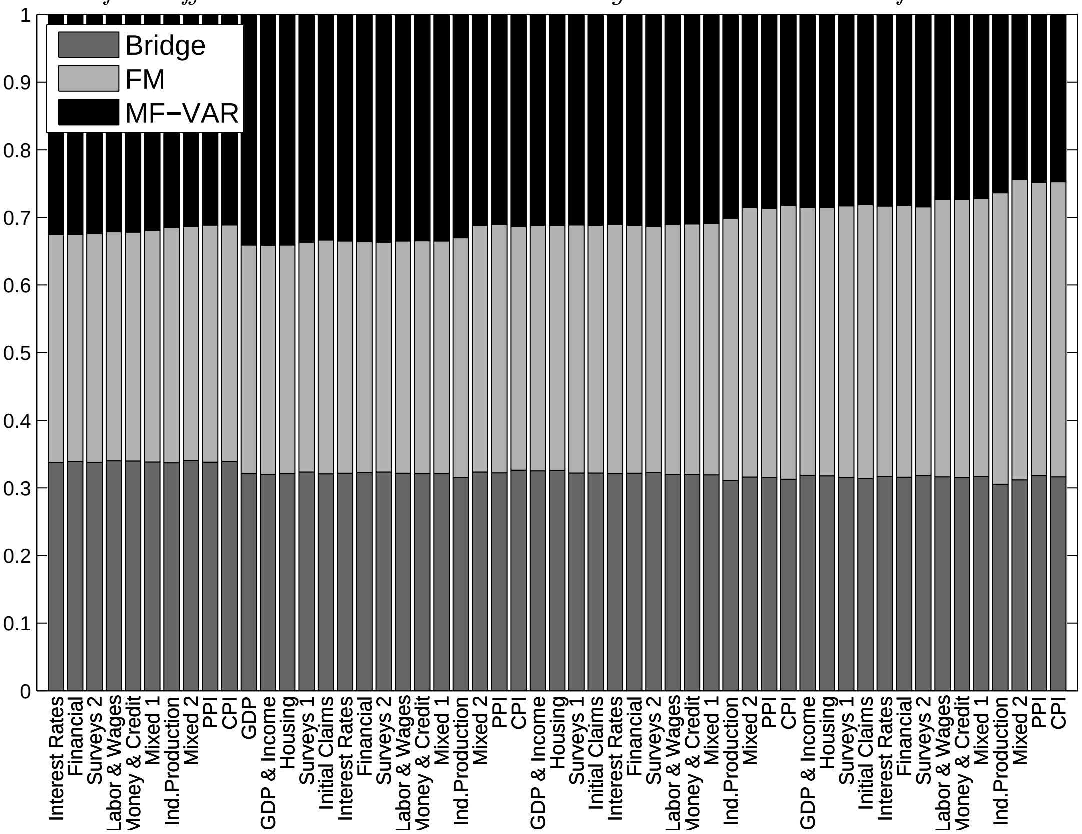
5. Return to 1 (2000 times)

6. Construct the combined predictive density by applying a kernel smoothing function on the vector $y^i_{h_q}$, evaluated on 401 equally spaced points.

The bootstrapping procedure takes into account that we use generated regressors in the models. For the MF-VARs the procedure is particularly time consuming, as the likelihood must be maximized at each simulation step (step 3 above). We therefore simplify the simulation steps 1 to 3 for this model class. Hence, F^i is re-sampled based on the time series of the covariance matrix of the state variables (assuming a multivariate normal distribution), computed with the Kalman Filter when estimating F^0. A^i is then re-sampled by applying OLS using F^i. Still, as we use 244 models in total, the whole forecasting experiment, which involves re-estimation and simulation of every model for every new block and vintage of data (55 blocks, 82 vintages), involves considerable computation time.

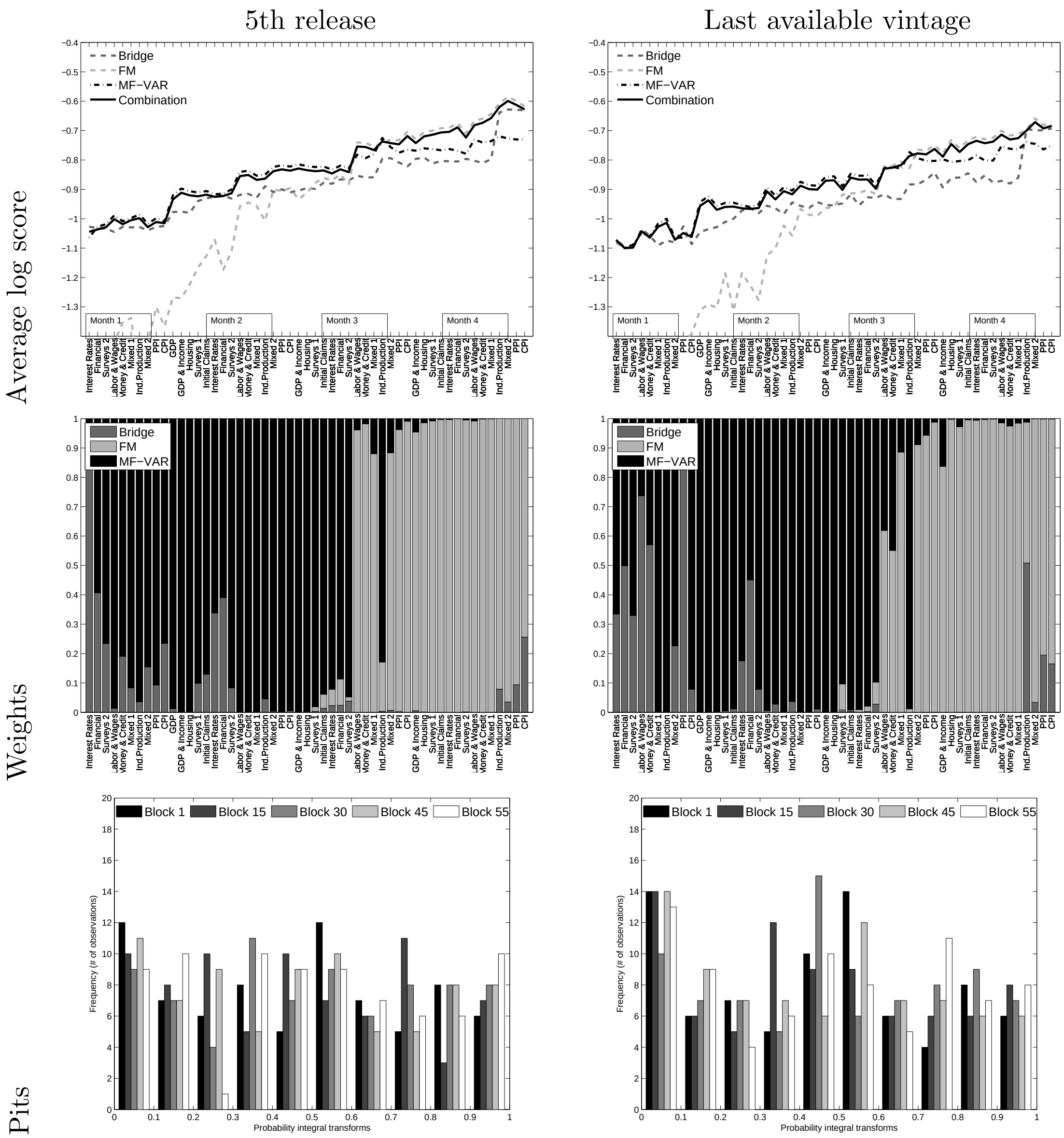
B Robustness

Figure 7. *Inverse MSE weights. End of sample weights attached to the different model classes after different block releases. Evaluated against second release of data*



Note: The nowcasts from the individual models and model classes have been combined using the linear opinion pool and inverse MSE weights. The evaluation period runs from 1990Q2 to 2010Q3.

Figure 8. Results evaluated against fifth release and last available vintage of GDP.



Note: The nowcasts from the individual models and model classes have been combined using the linear opinion pool and log score weights. The evaluation period runs from 1990Q2 to 2010Q3.

C Data description

Block	Block Name	Description	Publication Lag	Start Vintage
1	Interest Rates	Federal funds rate	One month	Last vintage
1	Interest Rates	3 month Treasury Bills	One month	Last vintage
1	Interest Rates	6 month Treasury Bills	One month	Last vintage
2	Financials	Spot USD/EUR	One month	Last vintage
2	Financials	Spot USD/JPY	One month	Last vintage
2	Financials	Spot USD/GBP	One month	Last vintage
2	Financials	Spot USD/CAD	One month	Last vintage
2	Financials	Price of gold on the London market	One month	Last vintage
2	Financials	NYSE composite index	One month	Last vintage
2	Financials	Standard & Poors 500 composite index	One month	Last vintage
2	Financials	Standard & Poors dividend yield	One month	Last vintage
2	Financials	Standard & Poors P/E Ratio	One month	Last vintage
2	Financials	Moodys AAA corporate bond yield	One month	Last vintage
2	Financials	Moodys BBB corporate bond yield	One month	Last vintage
2	Financials	WTI Crude oil spot price	One month	Last vintage
3	Surveys 2	Purchasing Managers Index (PMI)	One month	03.03.1997
3	Surveys 2	ISM mfg index, Production	One month	02.11.2009
3	Surveys 2	ISM mfg index, Employment	One month	02.11.2009
3	Surveys 2	ISM mfg index, New orders	One month	02.11.2009
3	Surveys 2	ISM mfg index, Inventories	One month	02.11.2009
3	Surveys 2	ISM mfg index, Supplier deliveries	One month	02.11.2009
4	Labor Market	Civilian Unemployment Rate	One month	05.01.1990
4	Labor Market	Civilian Participation Rate	One month	07.02.1997
4	Labor Market	Average (Mean) Duration of Unemployment	One month	05.01.1990
4	Labor Market	Civilians Unemployed - Less Than 5 Weeks	One month	05.01.1990
4	Labor Market	Civilians Unemployed for 5-14 Weeks	One month	05.01.1990
4	Labor Market	Civilians Unemployed for 15-26 Weeks	One month	05.01.1990
4	Labor Market	Civilians Unemployed for 27 Weeks and Over	One month	05.01.1990
4	Labor Market	Employment on nonag payrolls: Total nonfarm	One month	05.01.1990
4	Labor Market	Employment on nonag payrolls: Total Private Industries	One month	05.01.1990
4	Labor Market	Employment on nonag payrolls: Goods-Producing Industries	One month	05.01.1990
4	Labor Market	Employment on nonag payrolls: Construction	One month	05.01.1990
4	Labor Market	Employment on nonag payrolls: Durable goods	One month	05.01.1990
4	Labor Market	Employment on nonag payrolls: Nondurable goods	One month	05.01.1990
4	Labor Market	Employment on nonag payrolls: Manufacturing	One month	05.01.1990
4	Labor Market	Employment on nonag payrolls: Mining and logging	One month	05.01.1990
4	Labor Market	Employment on nonag payrolls: Service-Providing Industries	One month	05.01.1990
4	Labor Market	Employment on nonag payrolls: Financial Activities	One month	05.01.1990
4	Labor Market	Employment on nonag payrolls: Education & Health Services	One month	06.06.2003
4	Labor Market	Employment on nonag payrolls: Retail Trade	One month	05.01.1990
4	Labor Market	Employment on nonag payrolls: Wholesale Trade	One month	05.01.1990
4	Labor Market	Employment on nonag payrolls: Government	One month	05.01.1990
4	Labor Market	Employment on nonag payrolls: Trade, Transportation & Utilities	One month	05.01.1990
4	Labor Market	Employment on nonag payrolls: Leisure & Hospitality	One month	06.06.2003
4	Labor Market	Employment on nonag payrolls: Other Services	One month	05.01.1990
4	Labor Market	Employment on nonag payrolls: Professional & Business Services	One month	06.06.2003
4	Labor Market	Average weekly hours of PNW: Total private	One month	Last vintage
4	Labor Market	Average weekly overtime hours of PNW: Mfg	One month	Last vintage
4	Labor Market	Average weekly hours of PNW: Mfg	One month	Last vintage
4	Labor Market	Average hourley earnings:Construction	One month	Last vintage
4	Labor Market	Average hourly earnings: Mfg	One month	Last vintage
5	Money & Credit	M1 Money Stock	One month	30.01.1990
5	Money & Credit	M2 Money Stock	One month	30.01.1990
6	Mixed 1	Consumer credit: New car loans at auto finance companies, loan-to-value	Two months	Last vintage
6	Mixed 1	Consumer credit: New car loans at auto finance companies, amount financed	Two months	Last vintage
6	Mixed 1	Federal government total surplus or deficit	One month	Last vintage
6	Mixed 1	Exports of goods, total census basis	Two months	Last vintage
6	Mixed 1	Imports of goods, total census basis	Two months	Last vintage
7	Ind. Production	Industrial Production Index	One month	17.01.1990
7	Ind. Production	Industrial Production: Final Products (Market Group)	One month	14.12.2007
7	Ind. Production	Industrial Production: Consumer Goods	One month	14.12.2007
7	Ind. Production	Industrial Production: Durable Consumer Goods	One month	14.12.2007
7	Ind. Production	Industrial Production: Nondurable Consumer Goods	One month	14.12.2007
7	Ind. Production	Industrial Production: Business Equipment	One month	14.12.2007

Block	Block Name	Description	Publication Lag	Start Vintage
7	Ind. Production	Industrial Production: Materials	One month	14.12.2007
7	Ind. Production	Industrial Production: Durable Materials	One month	14.12.2007
7	Ind. Production	Industrial Production: nondurable Materials	One month	14.12.2007
7	Ind. Production	Industrial Production: Manufacturing (NAICS)	One month	14.12.2007
7	Ind. Production	Industrial Production: Durable Manufacturing (NAICS)	One month	14.12.2007
7	Ind. Production	Industrial Production: Nondurable Manufacturing (NAICS)	One month	14.12.2007
7	Ind. Production	Industrial Production: Mining	One month	14.12.2007
7	Ind. Production	Industrial Production: Electric and Gas Utilities	One month	14.12.2007
7	Ind. Production	Capacity Utilization: Manufacturing (NAICS)	One month	05.12.2002
7	Ind. Production	Capacity Utilization: Total Industry	One month	15.11.1996
8	Mixed 2	Housing starts: Total new privately owned housing units started	One month	18.01.1990
8	Mixed 2	New private housing units authorized by building permits	One month	17.08.1999
8	Mixed 2	Phily Fed Buisness outlook survey, New orders	Current month	Last vintage
8	Mixed 2	Phily Fed Buisness outlook survey, General business activity	Current month	Last vintage
8	Mixed 2	Phily Fed Buisness outlook survey, Shipments	Current month	Last vintage
8	Mixed 2	Phily Fed Buisness outlook survey, Inventories	Current month	Last vintage
8	Mixed 2	Phily Fed Buisness outlook survey, Unfilled orders	Current month	Last vintage
8	Mixed 2	Phily Fed Buisness outlook survey, Prices paid	Current month	Last vintage
8	Mixed 2	Phily Fed Buisness outlook survey, Prices received	Current month	Last vintage
8	Mixed 2	Phily Fed Buisness outlook survey, Number of employees	Current month	Last vintage
8	Mixed 2	Phily Fed Buisness outlook survey, Average workweek	Current month	Last vintage
9	PPI	Producer Price Index: Finished Goods	One month	12.01.1990
9	PPI	Producer Price Index: Finished Goods Less Food & Energy	One month	11.12.1996
9	PPI	Producer Price Index: Finished Consumer Goods	One month	11.12.1996
9	PPI	Producer Price Index: Intermediate Materials: Supplies & Components	One month	12.01.1990
9	PPI	Producer Price Index: Crude Materials for Further Processing	One month	12.01.1990
9	PPI	Producer Price Index: Finished Goods Excluding Foods	One month	11.12.1996
9	PPI	Producer Price Index: Finished Goods Less Energy	One month	11.12.1996
10	CPI	Consumer Prices Index: All Items (urban)	One month	18.01.1990
10	CPI	Consumer Prices Index: Food	One month	12.12.1996
10	CPI	Consumer Prices Index: Housing	One month	Last vintage
10	CPI	Consumer Prices Index: Apparel	One month	Last vintage
10	CPI	Consumer Prices Index: Transportation	One month	Last vintage
10	CPI	Consumer Prices Index: Medical care	One month	Last vintage
10	CPI	Consumer Prices Index: Commodities	One month	Last vintage
10	CPI	Consumer Prices Index: Durables	One month	Last vintage
10	CPI	Consumer Prices Index: Services	One month	Last vintage
10	CPI	Consumer Prices Index: All Items Less Food	One month	12.12.1996
10	CPI	Consumer Prices Index: All Items Less Food & Energy	One month	12.12.1996
10	CPI	Consumer Prices Index: All items less shelter	One month	Last vintage
10	CPI	Consumer Prices Index: All items less medical care	One month	Last vintage
11	GDP	Real Gross Domestic Product	One quarter	28.01.1990
12	GDP & Income	Real Disposable Personal Income	One month	29.01.1990
12	GDP & Income	Real Personal Consumption Expenditures	One month	29.01.1990
12	GDP & Income	Real Personal Consumption Expenditures: Durable Goods	One month	29.01.1990
12	GDP & Income	Real Personal Consumption Expenditures: Nondurable Goods	One month	29.01.1990
12	GDP & Income	Real Personal Consumption Expenditures: Services	One month	29.01.1990
12	GDP & Income	Personal Consumption Expenditures: Chain-type Price Index	One month	01.08.2000
12	GDP & Income	Personal Consumption Expenditures: Chain-Type Price Index Less Food & Energy	One month	01.08.2000
13	Housing	New one family houses sold	One month	30.07.1999
13	Housing	New home sales: Ratio of houses for sale to houses sold	One month	Last vintage
13	Housing	Existing home sales: Single-family and condos	One month	Last vintage
14	Surveys 1	Chicago Fed MMI Survey	One month	Last vintage
14	Surveys 1	Composite index of 10 leading indicators	One month	Last vintage
14	Surveys 1	Consumer confidence surveys: Index of consumer confidence	Current month	Last vintage
14	Surveys 1	Michigan Survey: Index of consumer sentiment	Current month	31.07.1998
15	Initial Claims	Average weekly initial claims	Current month	Last vintage