Taylor & Francis
Taylor & Francis Group

Journal of Business & Economic Statistics

ISSN: 0735-0015 (Print) 1537-2707 (Online) Journal homepage: https://amstat.tandfonline.com/loi/ubes20

Combined Density Nowcasting in an Uncertain Economic Environment

Knut Are Aastveit, Francesco Ravazzolo & Herman K. van Dijk

To cite this article: Knut Are Aastveit, Francesco Ravazzolo & Herman K. van Dijk (2018) Combined Density Nowcasting in an Uncertain Economic Environment, Journal of Business & Economic Statistics, 36:1, 131-145, DOI: 10.1080/07350015.2015.1137760

To link to this article: https://doi.org/10.1080/07350015.2015.1137760

Accepted author version posted online: 18 Feb 2016.
Published online: 27 Apr 2017.

Supplementary materials for this article are available online. Please go to http://tandfonline.com/r/JBES

Combined Density Nowcasting in an Uncertain Economic Environment

Knut Are Aastveit
Department, Norges Bank, Oslo, Norway (*Knut-Are.Aastveit@norges-bank.no*)

Francesco Ravazzolo
Faculty of Economics and Management, Free University of Bozen/Bolzano, Bozen/Bolzano, Italy (*Francesco.Ravazzolo@norges-bank.no*)

Herman K. van Dijk
Econometric Institute, Erasmus University Rotterdam, Econometrics Department, VU University Amsterdam, and Tinbergen Institute, Rotterdam, The Netherlands (*hkvandijk@ese.eur.nl*)

We introduce a combined density nowcasting (CDN) approach to dynamic factor models (DFM) that in a coherent way accounts for time-varying uncertainty of several model and data features to provide more accurate and complete density nowcasts. The combination weights are latent random variables that depend on past nowcasting performance and other learning mechanisms. The combined density scheme is incorporated in a Bayesian sequential Monte Carlo method which rebalances the set of nowcasted densities in each period using updated information on the time-varying weights. Experiments with simulated data show that CDN works particularly well in a situation of early data releases with relatively large data uncertainty and model incompleteness. Empirical results, based on U.S. real-time data of 120 monthly variables, indicate that CDN gives more accurate density nowcasts of U.S. GDP growth than a model selection strategy and other combination strategies throughout the quarter with relatively large gains for the two first months of the quarter. CDN also provides informative signals on model incompleteness during recent recessions. Focusing on the tails, CDN delivers probabilities of negative growth, that provide good signals for calling recessions and ending economic slumps in real time.

KEY WORDS: Density forecast combination; Survey forecast; Bayesian filtering; Sequential Monte Carlo Nowcasting; Real-time data.

1. INTRODUCTION

Economic forecast and decision making in real time have, in recent years, been made under a high degree of uncertainty. One prominent feature of this uncertainty is that many key statistics are released with a long delay, are subsequently revised and are available at different frequencies. Therefore, professional economists in business and government, whose job is to track swings in the economy and to make forecasts that inform decision-makers in real time, prefer to examine a large number of potential relevant time series.

In this context, factor models provide a convenient and efficient tool to exploit information in a large panel of time series in a systematic way by allowing for information reduction in a parsimonious manner while retaining forecasting power, see, for example, Stock and Watson (2002a,b, 2006), Forni et al. (2005), and Boivin and Ng (2005). A recent study by Giannone, Reichlin, and Small (2008) shows that these models are particularly suitable for *nowcasting*. The basic principle of nowcasting is the exploitation of information that is published early and possibly at higher frequencies than the target variable of interest to obtain an "early estimate" before the official number becomes available, see Evans (2005) and Banbura, Giannone, and Reichlin (2011). A key challenge is dealing with the differences in data release dates that cause the available information set to differ over points in time within the quarter. This is what Wallis (1986) coined the "ragged edge" of data. Giannone, Reichlin, and Small (2008) evaluated point nowcasts from a dynamic factor model and highlight the importance of using nonsynchronous data release, showing that the root mean square forecasting error decreases monotonically with each release.

Recent academic literature on factor models and nowcasting focused on developing single models that increase forecast accuracy in terms of point nowcasts, see, among others, Banbura and Modugno (2014) and Banbura and Rünstler (2011). As there is considerable uncertainty regarding several features of the model specification, for example, choice of variables to include in the large dataset, choice of number of factors, choice of lag length, and so on, Koop and Potter (2004) and Clark and McCracken (2009, 2010) suggested to follow the idea of Bates and Granger (1969) and combining forecasts from a wide range

January 2018, Vol. 36, No. 1
DOI: 10.1080/07350015.2015.1137760
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/jbes.

of models with different features to reduce these problems.[1] Surprisingly however, few studies in the nowcasting literature focus on combining nowcasts from different models, Kuzin, Marcellino, and Schumacher (2013) and Aastveit et al. (2014) being notable exceptions. Furthermore, building on earlier work in statistics by, for example, West and Crosse (1992) and West (1992), the research interest in forecast combination has more recently focused on the construction of combinations of predictive densities, see, for example, Hall and Mitchell (2007) and Jore, Mitchell, and Vahey (2010). An extension to density forecasting is to allow for time-varying model weights with learning and model set incompleteness, see Billio et al. (2013). Using a combination scheme that allows for model set incompleteness seems particularly suitable for nowcasting, as economic decision makers produce their nowcasts based on both incomplete data information (ragged edge problem) and uncertainty about the true data-generating process.

In this article, we introduce a combined density nowcasting (CDN) approach to dynamic factor models (DFMs) that accounts for time-varying uncertainty of several model and data features to provide more accurate and complete density nowcasts. The latent weights of the combination scheme depend on past nowcasting performance and other learning mechanisms. Our weights can therefore be interpreted as the density equivalent to the dynamic model averaging proposed by Koop and Korobilis (2012). The combined density scheme is incorporated in a Bayesian sequential Monte Carlo method, which rebalances the set of nowcasted densities in each period using updated information on the time-varying weights.[2] In this way, we are able to weight data uncertainty, parameter uncertainty, model uncertainty, including model incompleteness, and uncertainty in the combination of weights in a coherent way. We address the aforementioned sources of uncertainty using a large unbalanced real-time macroeconomic dataset for the United States that consists of 120 monthly indicators and combine predictive density nowcasts from four different DFMs that vary in terms of the number of factors included.

In statistical terms, CDN results in a convolution of several probability density functions consisting of the density of the nowcasts of individual models, the density of the latent weights of the combination scheme, and the density of the combination scheme. The integral of this product of densities with respect to the nowcasts of the individual models and the latent weights is what we are interested in. It does not have a closed form solution and, therefore, has to be evaluated numerically. The algorithm that we use is an extension of the nonlinear filtering methods of Billio et al. (2013) for the case of dynamic factor models with model incompleteness and data uncertainty. The application of the proposed sequential Monte Carlo filtering method leads to a good approximation, but the procedure is computationally intensive when the number of models to combine increases substantially. By making use of recent advances in computing power and parallel programming technique, it is feasible to apply nonlinear time-varying weights to four factor models at different points in time during the quarter. In doing so, we apply the MATLAB package DeCo (density combination), developed by Casarin et al. (2015), which provides an efficient implementation of the algorithm in Billio et al. (2013) based on CPU and GPU parallel computing.

We first implement simulation experiments to understand the role of incompleteness for nowcasting. We distinguish between data incompleteness (ragged edge problem) and model set incompleteness (the true model is not a part of the forecasters' model space) and compare point and density nowcasting performance from CDN with the performance of a Bayesian model averaging (BMA) approach, the optimal combination of density forecasts approach (OptComb) suggested by Hall and Mitchell (2007) and Geweke and Amisano (2011), equal weights and the ex post best individual model. The results illustrate that all approaches provide accurate point and density nowcasts when there is no incompleteness. However, when data incompleteness and/or model set incompleteness is present, the point and density nowcasting performance from CDN is superior to BMA, OptComb, equal weights and the ex post best individual model, providing considerably more accurate nowcasts, in particular at early data releases with relatively large data uncertainty and model incompleteness.

Next, we show the usefulness of CDN when it is applied to four different DFMs for nowcasting GDP growth using U.S. real-time data that consist of 120 monthly indicators. We divide data into different blocks, according to their release date within the quarter, and update the density nowcasts at three different points in time during each month of the quarter for the evaluation period 1990Q2–2010Q3. Our experiment refers to a professional economist who is interested in dealing with both data and model uncertainty. We find that CDN outperforms BMA, OptComb, equal weights, a selection strategy, and even the ex-post best individual model in terms of density nowcasting performance for all blocks. Also, empirically, we find relatively large gains in terms of improved density nowcasts for the first blocks of the quarter compared to the final blocks of the quarter.

By studying the standard deviation of the combination residuals, we show that this is higher for the earlier blocks in the quarter than for the later blocks in the quarter, indicating that incompleteness plays a larger role in the early part of the quarter. Thus, there are clear gains in terms of improved nowcasting performance from using CDN when incompleteness is present. We emphasize that the standard deviations of the combination residuals fluctuate over time and seem to increase during economic downturns, providing informative signals on model incompleteness.

Finally, we document that CDN also performs well with respect to focusing on the tails and delivers probabilities of negative growth that provide timely warning signals for calling a recession and ending economic slumps. These are in line with forecasts from the Survey of Professional Forecasters.

The structure of the article is as follows. Section 2 introduces CDN. Section 3 describes the data. Section 4 contains results using simulated data and Section 5 provides results of the application of the proposed method to nowcasting U.S. growth. Section 6 concludes.

[1]The idea of combining forecasts from different models has been widely used for economic forecasting. Timmermann (2006) provided an extensive survey of different combination methods.

[2]Note the analogy with dynamic portfolio management of a set of assets, where a periodic rebalancing of the assets occurs depending on the dynamic pattern of the weights that incorporate past performance of the assets.

2. THE COMBINED DENSITY NOWCASTING FOR DYNAMIC FACTOR MODELS

There is considerable empirical evidence that dynamic factor models (DFMs) provide accurate short-term forecasts, see, for example, Giannone, Reichlin, and Small (2008), Banbura and Modugno (2014) and Foroni and Marcellino (2014) for point forecasts and Aastveit et al. (2014) and Marcellino, Porqueddu, and Venditti (2015) for density forecasts. These models are particularly useful in a data-rich environment, where common latent factors and shocks are assumed to drive the co-movements between aggregate and disaggregate variables and the real-time data flow is inherently high dimensional with data released at different frequencies. Still, there is considerable uncertainty regarding several features of the model specification, for example, choice of variables to include in the large dataset, choice of number of factors and choice of lag length. Selection criteria and various testing procedure have been proposed to address such problems, see, for example, Bai and Ng (2006). However, as pointed out by Koop and Potter (2004), a potential problem with model selection based on sequential testing procedures of information criteria is that statistical evidence from other plausible models will be ignored. Koop and Potter (2004) therefore provided, in the context of factor models, a theoretical justifications for averaging across models, as opposed to selecting a single model. We build on this literature and propose a general model structure which can deal with both uncertainty related to data due to different sample frequencies and data releases, and uncertainty regarding model specification, such as selecting the number of factors k with $k = 1, \ldots, K$ and other components of the information set I_K.

We start by specifying the convolution of the three probability density functions that involve a novel combination scheme that deals with model uncertainty including model incompleteness. Next, we describe the individual factor models and how they cope with data uncertainty. We end with a brief description of the algorithms used to evaluate the convolution of densities and a description of alternative combination schemes used for comparison.

2.1 A Convolution of Combination, Weights, and Individual Model Predictive Densities for Multi-Period Ahead Nowcasting

In this section, we present our novel combination approach. Koop and Potter (2004) and Strachan and Dijk (2013) suggested to rely on Bayesian combination of several model features. We suggest to follow this line and extend their approach of using fixed model weights to the situation where we combine a set of predictive densities of model and data features using time-varying latent weights while allowing for model incompleteness, meaning that the true model is not necessarily included in the model set. To the extent that we use time-varying latent weights, our method can be viewed as the density equivalent of the dynamic model averaging approach in Koop and Korobilis (2012). Note, however, that the combination scheme in Koop and Korobilis (2012) does not allow for model incompleteness.

The combined density is a convolution of the density of the combination scheme, the density of the latent weights, and the predictive densities of the different individual models. Since there are K specifications of different models, we propose to compute the combined nowcast density of GDP growth $p(y_{t+h}|I_K)$ as

$$p(y_{t+h}|I_K) = \int_{\widetilde{Y}_{t+h}} \int_{W_{t+h}} p(y_{t+h}|\widetilde{y}_{t+h}, w_{t+h}, I_K) \times p(w_{t+h}|w_t) p(\widetilde{y}_{t+h}|I_K) dw_{t+h} d\widetilde{y}_{t+h}, \quad (1)$$

where $\widetilde{y}_{t+h}$ is an element of $\widetilde{Y}_{t+h} \in \mathcal{Y} \subset \mathbb{R}^K$, w_{t+h} is an element of W_{t+h}, the K-dimensional simplex. The density $p(y_{t+h}|\widetilde{y}_{t+h}, w_{t+h}, I_K)$ specifies the combination scheme and $p(w_{t+h}|w_t)$ is the density of the $(K \times 1)$ latent weights w_{t+h}. The density $p(\widetilde{y}_{t+h}|I_K)$ is the joint predictive density for the variable $\widetilde{y}_{t+h}$ with K different initial conditions. In Section 2.2, we will describe how to estimate the set of predictive densities $p(\widetilde{y}_{t+h}|\widetilde{F}_{t+h}, k)$ and $p(\widetilde{F}_{t+h}|k)$ to obtain $p(\widetilde{y}_{t+h}|k) = p(\widetilde{y}_{t+h}|\widetilde{F}_{t+h}, k) \times p(\widetilde{F}_{t+h}|k)$ for each individual model k, with $k = 1, \ldots, K$, that lead to $p(\widetilde{y}_{t+h}|I_K)$. We note that the combined density $p(y_{t+h}|I_K)$ is computed in a recursive way depending on past data. The combination weights w_{t+h} and the combination scheme are computed using a *direct* approach, see Marcellino, Stock, and Watson (2006). Most combination methods rely on the direct approach, see, for example, BMA, and although an iterated updating approach to evaluate the weights is computationally feasible and theoretically attractive under correct model specification, we aim to compare our strategy to standard combination schemes.

Given that we make use of a direct approach to nowcasting the weights, $p(w_{t+h}|w_t)$ is not h-order Markovian, but it can be interpreted as a degenerate h-order Markov process. Take the case of two periods nowcasting, that we use in practice, and define the transition function $p(w_{t+2}, w_{t+1}|w_{t+1}, w_t)$ as equal to $p(w_{t+2}|w_t)\delta_{w_{t+1}}(w_{t+1})$. That is, we have a "partially degenerate" random variable, and the Dirac delta, $\delta_{w_{t+1}}(w_{t+1})$, takes account of the fact that w_{t+1} is given in this step. For convenience, we write explicitly the joint $h = 1$, 2-step ahead nowcast density:

$$\begin{aligned} p(y_{t+2}, y_{t+1}|I_K) = \int_{(\widetilde{Y}_{t+2}, \widetilde{Y}_{t+1})} \int_{(W_{t+2}, W_{t+1})} & p(y_{t+2}|\widetilde{y}_{t+2}, w_{t+2}, I_K) \\ & \times p(w_{t+2}|w_t) p(\widetilde{y}_{t+2}|I_K) \\ & \times p(y_{t+1}|\widetilde{y}_{t+1}, w_{t+1}, I_K) p(w_{t+1}|w_t) \\ & \times p(\widetilde{y}_{t+1}|I_K) dw_{t+2} d\widetilde{y}_{t+2} dw_{t+1} d\widetilde{y}_{t+1}, \end{aligned} \quad (2)$$

where $p(y_{t+2}|\widetilde{y}_{t+2}, w_{t+2}, I_K)$ and $p(w_{t+2}|w_t)$ are computed using direct forecasting and $p(\widetilde{y}_{t+2}|I_K)$ is computed using iterative forecasting.

We make use of a Gaussian density for the combination scheme, which allows for model incompleteness via the following specification:

$$p(y_{t+h}|\widetilde{y}_{t+h}, w_{t+h}, I_K) \propto \exp\left\{ -\frac{1}{2\sigma^2} \left(y_{t+h} - \widetilde{y}_{t+h}^{'} w_{t+h} \right)^2 \right\}, \quad (3)$$

where we repeat that w_{t+h} is a vector containing the K values for the combination weights and $\widetilde{y}_{t+h}$ contains the K predicted values from a distribution with density $p(\widetilde{y}_{t+h}|I_K)$. For the use of alternative density functions for the combination scheme, see appendix B.1 in Billio et al. (2013).

In our modeling strategy, combination disturbances are estimated and their distribution follows a Gaussian process with mean zero and standard deviation σ, providing a probabilistic measure of the incompleteness of the model set. In other words, the model that is specified in Equation (3) can be written as

$$y_{t+h} = \widetilde{y}'_{t+h} w_{t+h} + \zeta_{t+h} \tag{4}$$

with $\zeta_{t+h} \sim \mathcal{N}(0, \sigma^2)$.

Second, the combination weights w_{t+h} have a probabilistic distribution in the standard simplex. We model them as logistic transforms, given as

$$w_{k,t+h} = \frac{\exp\{z_{k,t+h}\}}{\sum_{k=1}^{K} \exp\{z_{k,t+h}\}}, \quad k = 1, \ldots, K, \tag{5}$$

where the $(K \times 1)$ vector of latent weights $z_{t+h} = (z_{1,t+h}, \ldots, z_{K,t+h})'$ has a distribution with density given as

$$p(z_{t+h}|z_{t+h-1}, \widetilde{y}_{t-\tau:t}) \propto \exp\left\{ -\frac{1}{2} (\Delta z_{t+h} - \Delta e_{t+h})' \times \Lambda^{-1} (\Delta z_{t+h} - \Delta e_{t+h}) \right\} \tag{6}$$

with $\Delta z_{t+h} = z_{t+h} - z_{t+h-1}$ and $\Delta e_{t+h} = e_{t+h} - e_{t+h-1}$. The vector $e_{t+h} = (e_{1,t+h}, \ldots, e_{K,t+h})'$ is specified as a learning function based on past predictive performance given as

$$e_{k,t+h} = (1 - \lambda) \sum_{i=\tau}^{t} \lambda^{i-1} e_{k,i}, \quad k = 1, \ldots, K \tag{7}$$

with $\lambda \in (0, 1)$ as discount factor and $(t - \tau + 1)$ as the length of the interval for the learning parameter. In the simulation exercises and in the empirical application, we set $\lambda = 0.95$ and $\tau = 1$. Results in Stock and Watson (2004) suggest that low discounting, setting the discount factor to 0.95 or 0.99, provides better point forecasts than higher discounting, setting the discount factor to 0.9 or lower. Similar results were also obtained for density forecasts by Billio et al. (2013) in a macroeconomic application, and Pettenuzzo and Ravazzolo (2015) in a financial application. We note that in principle the parameter λ could be estimated from the data, and one possibility would be to rely on a grid search to estimate it (see Billio et al. (2013) for a discussion of this option). Setting $\tau = 1$ implies that the learning function is based on an expanding window. Thus, z_{t+h} is a latent process evolving over time with dynamics following an h-order Markov specification depending on past performances which describes the contribution of each model in the combination. The logistic transformation restricts weights to be in the unit interval.

Following the discussion in Gneiting (2011), we note that different scoring rules may be applied depending on the user preference. That is, a user interested in point forecasting may focus on mean square prediction errors; a user with a more general loss function may focus on scores that are based on density forecasting, such as the log score, see Section 2.5. A user just interested to standard Bayesian updating and no learning based on past performance scores can set $\lambda = 1$ and weights will be driven by a process equal to the previous values plus a news component normally distributed with zero mean and Λ covariance matrix. We define this as the combined density nowcasting approach without learning (CDNNL).

2.2 Individual Factor Models

We employ dynamic factor model specifications as suggested by Giannone, Reichlin, and Small (2008). Assume we have a monthly (m) unbalanced dataset, where the unbalancedness is due to data being released at different points in time (ragged edge). Let $X_{t_m^x} = (x_{1,t_m^x}, \ldots, x_{N,t_m^x})'$ be a vector of observable and stationary monthly variables which have been standardized to have a 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 $t_m = 3, 6, 9, \ldots, T_m$) in the same way as in Giannone, Reichlin, and Small (2008) and Aastveit et al. (2014). Quarterly differences are therefore calculated as $x_t = x_{t_m}^{(3)} = (1 - L_m^3)(1 + L_m + L_m^2)Z_{t_m^x}$, where L_m is the monthly lag operator and Z_{t_m} is the raw data. Likewise quarterly growth rates are calculated as $x_t = x_{t_m}^{(3)} = (1 - L_m^3)(1 + L_m + L_m^2)\log Z_{t_m^x}$. For alternative approaches of dealing with the ragged edge in the monthly dataset and the bridge between monthly and quarterly variables, see Marcellino and Schumacher (2010). A dynamic factor model is then given by the following observation equation:

$$X_{t_m^x} = \chi_{t_m^x} + \xi_{t_m^x} = \Lambda F_{t_m^x} + \xi_{t_m^x}, \qquad \xi_{t_m^x} \sim N(0, \Sigma_\xi), \tag{8}$$

where Λ is a $(n \times k)$ matrix of factor loadings, $F_{t_m^x} = (f_{1t_m^x}, \ldots, f_{kt_m^x})'$ is the static common factors and $\epsilon_{t_m^x} = (\epsilon_{1t_m^x}, \ldots, \epsilon_{nt_m^x})'$is an idiosyncratic component with zero expectation and $\Psi_{t_m^x} = E[\epsilon_{t_m^x}\epsilon'_{t_m^x}]$ as covariance matrix.

The dynamics of the common factors follows a VAR process:

$$F_{t_m^x} = AF_{t_m^x-1} + Bu_{t_m^x} \tag{9}$$

where $u_{t_m^x} \sim WN\,(0, I_s)$, B is a $(k \times s)$ matrix of full rank s, A is a $(k \times k)$ matrix where all roots of $\det(I_k - Az)$ lie outside the unit circle. The idiosyncratic and VAR residuals are assumed to be independent:

$$\begin{bmatrix} \xi_{t_m^x} \\ u_{t_m^x} \end{bmatrix} \sim \text{iid}N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} R & 0 \\ 0 & Q \end{bmatrix}\right) \tag{10}$$

with R set to be diagonal. Note that the estimates are robust to violations of this assumption, see, for example, Banbura et al. (2012).

The factor model, Equations (8) and (9), is estimated in a two-step procedure using principal components and the Kalman filter. The unbalanced part of the dataset 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, Reichlin, and Small (2008).

Finally, predictions of quarterly GDP growth, y_t, are obtained by using a bridge equation where nowcasts of quarterly GDP growth (y_t) are expressed as a linear function of the expected common factors: to obtain quarterly aggregates of the monthly factors, $(F^{(3)}_{t_m^y+h_m|t_m^x})$, the monthly factors are forecasted over the remainder of the quarter using Equation (9). Then, the quarterly growth rate of GDP, y_t is regressed on the resulting factor values using a bridge equation like:

$$y_{t+h} = y_{t_m+h_m} = \alpha + \beta' F^{(3)}_{t_m+h_m|t_m+\omega} + e_{t_m+h_m},$$
$$e_{t_m+h_m} \sim N(0, \Sigma_e), \tag{11}$$

where β is an $k \times 1$ vector of parameters. Accordingly, forecasts of GDP growth $(y_{t_m+h_m|t_m^x})$ are constructed from Equation (11), conditional on the estimated parameters and the factor forecasts. The conditional information, t_m^x, varies over the months in each quarter depending on which data are available and therefore the joint density forecast from all K models, $p(\widetilde{y}_{t_m+h_m}|I_K)$, also changes. For the sake of notation in the rest of the section and in Equation (1) we generally refer to it as $p(\widetilde{y}_{t+h}|I_k)$. For the use of an alternative bridge equation that relies on a more general estimated weighting scheme; see Foroni, Marcellino, and Schumacher (2015) and Carriero, Clark, and Marcellino (2015). Note, that Equation (11) implies that h-step ahead nowcasts are computed as iterative forecasts. An alternative approach, not considered in this article, is the direct multi-step ahead forecasting suggested by Marcellino, Stock, and Watson (2006).

To estimate Equations (8), (9), and (11), one can make use of Bayesian approaches based on Monte Carlo or frequentist estimation principles. In our case, we take a pragmatic approach and make use of standard frequentist approaches based on bootstrapping to estimate Equations (8), (9), and (11), and then compute $p(\widetilde{F}_{t+h}|k)$ and $p(\widetilde{y}_{t+h}|\widetilde{F}_{t+h}, k)$ for a given model $k = 1, \dots, K$ and generate predicted values $\widetilde{y}_{t+h}$, conditional upon generated predicted values $\widetilde{F}_{t+h}$. Here, we apply the residual-based bootstrapping approach developed by Aastveit et al. (2014). The bootstrapping algorithm is explained in the online Appendix. Thus, motivated by Fernandez, Ley, and Steel (2001) and Sala-I-Martin, Doppelhofer, and Miller (2004), our approach is one of Bayesian averaging of frequentist estimates, extending their Bayesian averaging approach to account for time-varying weights and model set incompleteness.

2.3 Algorithm, Estimation, and Parallelization

If the three densities in Equation (1) all belonged to the normal family with no dynamics, the integral in Equation (1) could be solved analytically or by simple numerical methods like direct Monte Carlo simulation. In the case of a dynamic model structure with a normal distribution and also normal dynamics for the weights, one can make use of standard normal filtering methods. In our case, however, there exists a perfect analogy between the setup of the equations in our CDN approach and the model specification in the *nonlinear* state-space literature. We interpret CDN in terms of a nonlinear state-space formulation and apply a sequential Monte Carlo filtering method. That is, Equation (3) is analogous to the measurement or observable equation; Equations (5) and (6) are nonlinear transition equations and Equations (8), (9), and (11) can be interpreted as being equivalent to the parameter equations in the nonlinear state space. These latter equations can, alternatively, be specified as being part of a more general state-space model, where the nonlinear filtering methods are also used to approximate the densities. Thus, Equation (1) accounts for several sources of uncertainty, including different sample frequencies, different data releases, different information sets, and model specifications.

The convolution has such useful properties like commutative, associative, and distributive laws that enable us to be flexible in the order of integration and other properties under usual regularity conditions. As mentioned, we use sequential Monte Carlo integration to solve part of the integral in (1) by using the regularized version of the Liu and West (2001) filtering procedure for the weights and combination scheme and we make, further, use of draws from the K individual predictive densities.

Our methodology is very general and allows the evaluation of predictive densities provided by various methods (parametric Bayesian and frequentist models as well as nonparametric methods), given the condition that all three densities are proper. We repeat that in the empirical applications in Section 5, we construct predictive densities using frequentist bootstrapping methods and combine these predictive densities using Bayesian inference. The algorithm is explained in detail in the online Appendix.

2.4 Alternative Combination Approaches

In the simulation exercise and the empirical exercise, we compare our CDN approach with three other alternative combination approaches.

2.4.1 Bayesian Model Averaging. For Bayesian model averaging (BMA, henceforth), the individual predictive densities are combined into a composite-weighted predictive distribution $p(y_{t+h}|I_K)$, given by

$$p(y_{t+h}|I_K) = \sum_{k=1}^{K} P(M_k)\, p(\widetilde{y}_{t+h}|k), \tag{12}$$

where $P(M_k)$ is the posterior probability of model k, based on the predictive likelihood for model k. Mitchell and Hall (2005) discussed the analogy of the log score in a frequentistic framework to the log predictive likelihood in a Bayesian framework, and how it relates to the Kullback–Leibler divergence. See also Hall and Mitchell (2007), Jore, Mitchell, and Vahey (2010), and Geweke and Amisano (2010) for a discussion on the use of the log score as a ranking device for the forecast ability of different models and Hoeting et al. (1999) for a review on BMA.

We note that BMA assumes that the true model is included in the model set. It can be shown that the BMA combination weights converge (in the limit) to select the true model. However, as noted by Diebold (1991), all models could be false, and as a result the model set could be misspecified.

2.4.2 Optimal Combination. As an alternative to BMA, Hall and Mitchell (2007) and Geweke and Amisano (2011) proposed to use a linear prediction pool:

$$p(y_{t+h}|I_K) = \sum_{k=1}^{K} w_k\, p(\widetilde{y}_{t+h}|k), \tag{13}$$

where the individual model weights w_k are computed by maximizing the log predictive likelihood, or log score (LS), of the linear prediction pool:

$$\sum_{\tau=1}^{t-1} \log\left[\sum_{k=1}^{K} w_k \times \exp\left(\mathrm{LS}_{k,\tau+1}\right)\right] \tag{14}$$

with $\mathrm{LS}_{k,\tau+1}$ denoting the recursively computed log score for model k at time $\tau+1$, see Equation (17). Hall and Mitchell (2007) and Geweke and Amisano (2011) labeled the approach optimal combination (OptComb) and optimal prediction pools, respectively. Geweke and Amisano (2011) showed that the model weights, computed in this way, no longer converge to a unique solution, except in the case where there is a dominant model in terms of Kullback–Leibler divergence. Note that common to both BMA and the approach by Hall and Mitchell (2007) and Geweke and Amisano (2011), is the assumption that the model combination weights are constant over time.

2.4.3 Equal Weights. Several studies combine forecasts using simple averages based on equal weights (EW), see, for example, Timmermann (2006), Stock and Watson (2006), Clark and McCracken (2010), and Angelini et al. (2011). In fact, such a simple combination of forecasts is often found to outperform more sophisticated adaptive forecast combination methods. This result is often referred to as the forecast combination puzzle. We compute equally weighted forecasts using the linear prediction pool, Equation (13), where the weights attached to each model are set to $w_k = 1/K$.

2.5 Nowcast Evaluation

The aim of this article is to provide an efficient methodology which deals with various sources of uncertainty to improve nowcast accuracy. As most other papers focusing on nowcasting do, we first provide some results on point nowcasts. However, as these nowcasts are only optimal for a small and restricted group of loss functions, our main focus is on density nowcasting.

To shed light on the predictive ability of our methodology, we consider several evaluation statistics for point and density nowcasts previously proposed in the literature. For a generic density nowcast of GDP growth, $p(y_{t+h}|I_K)$, we compare point forecasts in terms of mean square prediction errors (MSPE)

$$\text{MSPE} = \frac{1}{t^*}\sum_{t=\underline{t}}^{\bar{t}} e_{t+h}^2,$$

where $t^* = \bar{t} - \underline{t} + h$, $\bar{t}$ and $\underline{t}$ denote the beginning and end of the evaluation period, and e_{t+h} is the h-step ahead square prediction error associated to the density $p(y_{t+h}|I_K)$. In the simulation exercise and the empirical applications, we use the mean of the density as point forecast.

The complete predictive densities are evaluated using the Kullback–Leibler information criterion (KLIC) based measure, using the expected difference in the logarithmic scores of the candidate nowcast densities; see, for example, Mitchell and Hall (2005), Hall and Mitchell (2007), Amisano and Giacomini (2007), and Kascha and Ravazzolo (2010). The KLIC chooses the model that on average gives the higher probability to events that actually occurred. Specifically, the KLIC distance between the true density $f(y_{t+h}|I_K)$ of a random variable y_{t+h} and the candidate density $p(y_{t+h}|I_K)$ is defined as

$$\begin{aligned}\text{KLIC}_{t+h} &= \int f(y_{t+h}|I_K)\ln\frac{f(y_{t+h}|I_K)}{p(y_{t+h}|I_K)}dy_{t+h},\\ &= \mathbb{E}_t[\ln f(y_{t+h}|I_K) - \ln p(y_{t+h}|I_K))],\end{aligned} \tag{15}$$

where $\mathbb{E}_t(\cdot) = \mathbb{E}(\cdot|I_K)$ is the conditional expectation given information set I_K at time t. An estimate can be obtained from the average of the sample information, $y_{\underline{t}+1}, \ldots, y_{\bar{t}+1}$, that is part of the information set I_K, on $f(y_{t+h}|I_K)$ and $p(y_{t+h}|I_K)$:

$$\overline{\text{KLIC}} = \frac{1}{t^*}\sum_{t=\underline{t}}^{\bar{t}}[\ln f(y_{t+h}|I_K) - \ln p(y_{+h}|I_K)]. \tag{16}$$

Although we do not pursue the approach of finding the true density, we can still rank different densities. For the comparison of two competing models, it is sufficient to consider the logarithmic score (LS), which corresponds to the latter term in the above sum,

$$\text{LS} = -\frac{1}{t^*}\sum_{t=\underline{t}}^{\bar{t}}\ln p(y_{t+h}|I_K), \tag{17}$$

for different densities and to choose the model for which it is minimal, or, as we report in our tables, its opposite is maximal.

Finally, following Diebold, Gunther, and Tay (1998), we also evaluate density forecasting accuracy by testing goodness of fit relative to the "true," but unobserved density using the probability integral transforms (pits). 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. The pits at time t are:

$$\text{pits}_{t+h} = \int_{-\infty}^{y_{t+h}} p(z_{t+h}|I_K)dz_{t+h} \tag{18}$$

and should be uniformly, independently and identically distributed if the forecast densities $f(z_{t+h}|I_K)$ conditional on some information set, I_K, are correctly calibrated.

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]. We test for correct calibration by applying the raw-moments test, recently developed by Knüppel (2015). The raw-moments test is based on the standardized pits instead of the inverse normal transforms and accounts for possible serial correlation in the pits. The test is therefore also suitable for multi-step ahead forecasts and performs better in terms of size and power than other commonly used tests, such as the test provided by Berkowitz (2001).

3. DATA

We consider in total 120 monthly variables to nowcast quarterly GDP growth in the United States. Our real-time dataset is similar to the one used in Aastveit et al. (2014).[3] As in that article, we use the last available data vintage as real-time observations for consumer prices and survey data if the real-time data vintage is not available. For other series, such as disaggregated measures of industrial production, real-time vintage data exist only for parts of the evaluation period. For these variables, we use the first available real-time vintage and truncate these series backward recursively. Finally, for financial data, we construct monthly averages of daily observations.

Following Banbura and Rünstler (2011), we divide the data into "soft data" and "hard data." The first set includes 38 surveys and financial indicators and reflects market expectations,

[3]The main source is the ALFRED (Archival Federal Reserve Economic Data) database maintained by the Federal Reserve Bank of St. Louis. In addition some series are also collected from the Federal Reserve Bank of Philadelphia's Real-Time Data Set for Macroeconomists, see Croushore and Stark (2001) and the online appendix.

Table 1. Block information

Block	Time	Horizon
	Nowcasting	
1	Start of first month of quarter	2 steps ahead
2	10th of first month of quarter (after inflation release)	2 steps ahead
3	Around 20–25th of first month of quarter (after GDP release)	1 step ahead
4	Start of second month of quarter	1 step ahead
5	10th of second month of quarter (after inflation release)	1 step ahead
6	Around 20–25th of second month of quarter	1 step ahead
7	Start of third month of quarter	1 step ahead
8	10th of third month of quarter (after inflation release)	1 step ahead
9	Around 20–25th of third month of quarter	1 step ahead
	Backcasting	
10	Start of fourth month of quarter	1 step ahead
11	10th of fourth month of quarter (after inflation release)	1 step ahead

NOTE: The table shows time in the quarter and nowcast horizon for the 11 blocks.

as opposed to the latter set that includes 82 measures of GDP components (e.g., industrial production), the labor market and prices. Although soft data are often more timely (i.e., released early in the quarter), while real activity data are published with a significant delay, the latter category is considered to contain a more precise signal for GDP forecasting.

The full nowcast evaluation period runs from 1990Q2 to 2010Q3. We use monthly real-time data with quarterly vintages from 1990Q3 to 2010Q4, that is, we do not take account of data revisions in the monthly variables within a quarter. The quarterly vintages reflect information available just before the first release of the GDP estimate. The starting point of the estimation period is 1982M1. We study nowcasts at nine different points in time during a quarter. They correspond to the beginning, middle and end of each month in the quarter. Since GDP measures are released approximately 20–25 days after the end of the quarter, our exercise also includes two backcasts, calculated at the beginning and the middle of the first month after the quarter of interest. See Table 1 for information on the final 11 blocks. When nowcasting GDP growth, the choice of a benchmark for the "actual" measure of GDP is not obvious (see Stark and Croushore (2002) for a discussion of alternative benchmarks). We follow Romer and Romer (2000) in using the second available estimate of GDP as the actual measure.

4. SIMULATION EXPERIMENTS WITH DATA AND MODEL INCOMPLETENESS

In this section, we implement several simulation exercises to understand the roles of data incompleteness and model incompleteness in nowcasting. In practice, economic decision makers produce their nowcasts based on incomplete data information (ragged edge problem) and uncertainty about the true data-generating process (DGP). In the simulation exercises below, we therefore distinguish between different degrees of incompleteness. *Weak incompleteness* is the case where the nowcaster produces nowcasts based on missing observations of data (i.e., the ragged edge problem). The DGP is in this case assumed to be a part of the nowcasters' model space. *Strong incompleteness* refers to the case where the DGP is not a part of the nowcasters' model space.

We run five simulation exercises, where in each exercise we produce recursive density nowcasts for 60 quarters. For the first three simulation exercises, we simulate nowcasted values assuming that the DGP (DGP1) follows a dynamic factor model, described in Section 2.2, with two factors extracted at the end of the sample (corresponding to the information set at Block 11). In the two final simulation exercises, we assume that the DGP follows a VAR(4) in GDP growth, the unemployment rate, core PCE inflation, and the federal funds rate. We distinguish between two cases, DGP2 where the DGP follows a VAR(4) with a constant variance and DGP3 where the DGP follows a VAR(4) with stochastic volatility. Note that DGP2 and DGP3 are estimated from a balanced panel at the end of the sample. In each simulation exercise, we compare the performance of our CDN approach, both in terms of point nowcasts (MSPE), density nowcasts (LS) and calibration (PITS), with the BMA, OptComb and EW approaches (discussed in Section 2.4), as well as the best ex-post individual model (Best Model). We report results for both the CDN approach with learning (CDN) and the CDN approach without learning (CDNNL). In addition, to get a sense of the absolute performance of the different approaches, we also report results from the true model (true). For the true model, the only source of forecast errors are the future shocks, that is, there is no estimation and model uncertainty.

In the first simulation exercise, (Sim1), we estimate (and combine) four individual DFMs with 1–4 factors extracted from a panel corresponding to the information at Block 11. Thus, in this exercise the DGP is a part of the model space and there is therefore no model set incompleteness and no data incompleteness. We introduce weak incompleteness in the second simulation exercise (Sim2). We estimate (and combine) the same individual DFMs with 1–4 factors. The only difference from Sim1 is that the models are now estimated with incomplete data information. More precisely, the models are estimated using data that corresponds to the information that is available when nowcasting at the middle of the quarter (i.e., Block 5). Hence, there is data incompleteness, but no model incompleteness.

The last three simulation exercises focus on cases of strong incompleteness (cases where both data incompleteness and model incompleteness is present). In the third simulation exercise, (Sim3), we estimate (and combine) 4 individual DFMs. However, we assume that for some reason, the factors are only estimated based on the "hard data" variables in our dataset (i.e., we assume that no survey data are available to the forecaster). Thus, there is model incompleteness, since the "true" model (which is a DFM with two factors extracted from the full dataset) is within the model space, but all the models are misspecified in terms of using the wrong dataset (i.e., using just a subset of all the "true" data series to extract the factors). In addition, we also assume that there is data incompleteness as in Sim2.

In the final two simulation exercises (Sim4 and Sim5), we assume a different DGP. In these cases, we assume that the DGP follows a VAR(4) in GDP growth, the unemployment

Table 2. Simulation results

	BMA	EW	OptComb	Best model	CDNNL	CDN	True
				Sim1: No incompleteness			
LS	−0.251	0.030	−0.048	**0.224**	0.061	0.074	0.297
PITS	0.360	0.192	0.260	0.159	0.200	0.264	
MSPE	0.028	0.025	0.027	0.025	0.027	**0.024**	0.013
				Sim2: Weak incompleteness			
LS	−3.882	−3.825	−3.048	−3.105	−0.531**	**−0.459****	0.297
PITS	0.103	**0.034**	**0.088**	0.126	0.464	0.442	
MSPE	0.198	0.169*	0.180*	0.198	0.148*	**0.147****	0.013
				Sim3: Strong incompleteness			
LS	−2.959	−3.056	−3.082	−2.928	−0.499**	**−0.457****	0.297
PITS	0.117	0.119	0.118	0.109	0.684	0.721	
MSPE	0.241	0.251	0.226	0.230	0.172*	**0.169****	0.013
				Sim4: Strong incompleteness			
LS	−0.567	−0.501	−0.493	−0.555	−0.353**	**−0.325****	−0.271
PITS	0.524	0.404	0.318	0.100	0.162	0.214	
MSPE	0.205	0.186*	0.179	0.186	**0.111****	0.112**	0.091
				Sim5: Strong incompleteness			
LS	−0.993	−1.049	−0.947	−1.187	−0.732**	**−0.623****	−0.622
PITS	0.490	0.183	**0.044**	**0.079**	0.622	0.539	
MSPE	0.347	0.343	0.335	0.364	0.213**	**0.209****	0.208

NOTES: The table reports results from the five simulation exercises, showing the average log score (LS), p-values for the pits test by Knüppel (2015) (PITS) and mean square prediction error (MSPE) for six different prediction methods applied to dynamic factor models: Bayesian model averaging based on the predictive likelihood (BMA), forecast combination with equal weights (EW), optimal combination (OptComb), the ex-post best performing model (Best Model), our combined density nowcasting approach without learning (CDNNL) and our combined density nowcasting approach with learning (CDN). In addition, we report results from the true model (True) with just future shocks (i.e., no estimation or model uncertainty) as a comparison. Bold numbers in the rows for LS and RMSE indicate the most accurate model for the different statistics. Statistically significant differences between the BMA approach and the alternative combination approaches are denoted by one and two asterisks corresponding to significance levels of 10% and 5%, respectively. Bold numbers in the rows for PITS indicate a rejection of the hypothesis of correctly calibrated densities at a 10% significance level.

rate, inflation, and the interest rate, while we again estimate and combine individual DFMs with 1–4 factors extracted from all the available data series (i.e., our estimated models are similar to the ones in the Sim2 exercise). However, we distinguish between a DGP following a constant variance VAR(4) (DGP2) and a VAR(4) with stochastic volatility (DGP3). The latter case is included since several studies have recently highlighted that it is important to account for stochastic volatility to obtain well-calibrated densities, see, for example, Clark (2011), Carriero, Clark, and Marcellino (2015), and Marcellino, Porqueddu, and Venditti (2015). It is therefore interesting to assess whether allowing for time-varying weights when combining forecasts from models with constant parameters, can capture some of the changes in the variance.

We study absolute accuracy by testing if the density forecasts are correctly calibrated, using the test in Knüppel (2015). The relative forecasting performance of the different models is assessed in terms of LS for density nowcasts and MSPE for point nowcasts. To provide a rough gauge of whether the differences in forecast accuracy between the BMA approach and the alternative combination approaches are significant, we follow Clark and Ravazzolo (2015) and apply a Diebold and Mariano (1995) t-tests for equality of the average loss (with loss defined as LS and MSPE). Statistically significant differences are denoted by asterisks corresponding to significance levels of 10% and 5%, respectively.

Table 2 reports results from the simulation exercises. When there is no model incompleteness, the best individual model, CDN and the other combination approaches perform very similarly in terms of point nowcasts. As expected, the best individual model outperforms CDN and the other combination approaches in terms of density nowcasting. Still, the results indicate that the CDN approach works well in the case where there is no data and model incompleteness. Note that although not statistically significant, the CDN approach provides better point and density nowcasts than the other combination approaches.

When introducing data and model incompleteness, there are clear gains from using our CDN approach relative to the other strategies. Starting with the case of weak incompleteness (i.e., Sim 2 where only data incompleteness is present), our CDN approach significantly improves upon the BMA approach, both in terms of point and density nowcast performance. Interestingly, the CDN approach also outperforms all other combination strategies and the ex-post best individual model. The latter result is rather striking, as the only source of incompleteness is missing data observations (ragged edge problem). Thus, this indicates that using a combination scheme that allows for model incompleteness is important in the case where data observations are missing. Compared with the results from the true model, the forecast errors measured in either LS or MSPE are large for some of the combination approaches. This highlights that missing data observations are an important source of forecasts errors.

The relative improvements, compared to the other strategies, are even more evident in the cases of strong incompleteness (Sim3–Sim5). Comparing the nowcasting performance from our CDN approach with the other strategies, indicates that there is scope for substantial improvements in performance by using a combination scheme that allows for model incompleteness when both data and model set incompleteness are present. In all cases, our CDN approach significantly improves upon the BMA approach and is superior to the other alternative combination approaches both in terms of density and point forecast performance.[4] Interestingly, when comparing the LS and MSPE from the CDN with the true model in Sim5, the values are very similar. In contrast, when comparing the LS and MSPE from the CDN without learning with the true model, differences occur. This suggests that applying time-varying weights are helpful for capturing changes in the variance, even in the case where forecasts are combined from models with constant parameters.

On a final note, in most of the simulation exercises, the density nowcasts seem to be well calibrated for all combination approaches. The exceptions are EW and OptComb in Sim2 and OptComb and Best Model in Sim5, where the null hypothesis of correctly calibrated densities are rejected at a 10% significance level.

5. EMPIRICAL APPLICATION

In this section, we analyze the performance of our CDN approach for nowcasting U.S. real GDP growth. The main goal of the exercise is to examine the nowcasting performance of our CDN approach and to study the role of model incompleteness for nowcasting.

5.1 Point and Density Nowcasts of GDP Growth

We produce density nowcasts/backcasts for GDP growth at 11 different points in time, described in Section 3, using four different DFMs. The models differ in terms of the numbers of factors included.[5] Our exercise refers to a researcher who constructs nowcasts in real time accounting for various forms of uncertainty, including uncertainty related to model specification. We consider six different model specification strategies: (1) Bayesian model averaging based on predictive likelihood (BMA); (2) a selection strategy where we recursively pick the model with the highest realized cumulative log score at each point in time throughout the evaluation period (SEL); (3) forecast combination with equal weights attached to the models (EW); (4) the optimal combination approach suggested by Hall and Mitchell (2007) and Geweke and Amisano (2011); (5) combined density nowcasting without learning (CDNNL); and (6) combined density nowcasting with learning (CDN).

Table 3 reports results for the six different model specification strategies at the 11 different points in time (blocks) during the quarter. In addition, we also report results for the best performing ex-post individual model (labeled Ex-Post). Note that this model is the DFM with two factors, similar to the specification in Giannone, Reichlin, and Small (2008). The first column shows the LS and MSPE for BMA, while all other columns report measures relative to the BMA performance. The table reveals four interesting results.

First, with the exception of the results for Block 1 and Block 2, the point nowcasting accuracy from the different models is very similar.

Second, CDN provides more accurate density nowcasts than BMA, EW, OptComb, and SEL for all of the blocks. For Block 1 to Block 4, the differences in terms of LS performance between the BMA approach and the CDN approach are statistically significant at a 10% significance level. The CDN approach also provides more accurate density nowcasts for all blocks than the ex-post best individual model, with the only exceptions being the results for Block 8 and Block 10, where Ex Post performs slightly better than the CDN. Overall, this indicates that there are gains in terms of improved nowcasting performance from CDN when we take into account the whole density shape of the nowcasts. Note that for all blocks the CDN approach which includes learning is superior to the CDN approach without learning (CDNNL). This shows that there are gains from applying time-varying weights.

Third, the relative gains in terms of improved density nowcasts are larger for the first blocks of the quarter than for the last blocks of the quarter. This supports the findings from the simulation exercises in Section 4, which showed that the gains from CDN are larger when uncertainty is high, and thus the incompleteness is strong. The data incompleteness (denoted as weak incompleteness) is larger in the early part of the quarter than in the latter part of the quarter. In addition, when data uncertainty is high, it is also more likely that it becomes harder to detect the "true" DGP than when the data uncertainty is low. That is, it is also more likely that model incompleteness is present when data uncertainty is high.

Fourth, the nowcasts from the CDN approach seems to be well-calibrated for all blocks, as the null hypothesis of correctly calibrated densities, based on the test in Knüppel (2015), cannot be rejected at the 10% significance levels.

5.2 Signals of Model Incompleteness

To illustrate the role of more substantial incompleteness, Figure 1 shows the standard deviations of the combination residuals for the incomplete model sets, see equation (4), over time for Block 1, Block 5, and Block 11. The figure reveals two interesting observations.

First, for most of the time observations, the standard deviation of the combination residuals is higher for Block 1 than Block 5 and Block 11, and higher for Block 5 than Block 11. This observation therefore confirms that incompleteness is higher in the early part of the quarter than in the later part of the quarter.

Second, the standard deviations of the combination residuals fluctuate over time. Interestingly, the standard deviation of the combination residual is high in 2001 and in the latter part of 2008 and the early part of 2009. This coincides with the U.S. economy being in a recession. The high standard deviation is

[4] Note that since DGP1 is rather different from DGP2 and DGP3, it may be misleading to compare the absolute performance for each model from the different simulation exercises (Sim3, Sim4, and Sim5).

[5] We obtained very similar results when using 12 different DFMs: four models extracting factors from the hard data; four models using the soft data; and four models using all the data. For each group, we then considered one to four factors, resulting in four different DFM specifications for each data group. In general, the models using factors extracted from all the data series were superior to the models extracting factors from either hard or soft data. For brevity, and to save computational time, we therefore only report results when combining four different DFMs.

Table 3. Point and density nowcasting

	BMA	SEL	EW	OptComb	Ex Post	CDNNL	CDN
				Block 1			
LS	−1.441	1.124	0.749	0.814	0.926	0.641**	**0.590****
PITS	0.119	0.198	**0.059**	0.167	0.147	0.389	0.708
MSPE	0.583	0.988	0.809	0.686	**0.524**	0.567	0.542
				Block 2			
LS	−1.101	1.117	0.920	0.892	0.954	0.782**	**0.715****
PITS	**0.094**	0.236	0.111	0.223	**0.068**	0.941	0.705
MSPE	0.317	1.032	1.052	0.988	0.959	1.014	**0.924**
				Block 3			
LS	−0.980	0.987	1.005	0.944	0.977	0.862	**0.814****
PITS	**0.049**	0.197	**0.040**	0.149	**0.041**	0.313	0.450
MSPE	0.289	0.989	1.141	1.036	**0.983**	1.086	1.025
				Block 4			
LS	−0.892	0.997	1.027	0.971	0.978	0.922	**0.862***
PITS	0.102	0.155	**0.087**	0.155	**0.075**	0.326	0.465
MSPE	0.275	0.991	1.162	1.046	**0.977**	1.108	1.007
				Block 5			
LS	−0.768	0.991	0.962	0.985	0.961	0.953	**0.897**
PITS	0.254	0.292	0.272	0.340	0.309	0.456	0.594
MSPE	0.241	0.990	1.269	1.098	**0.969**	1.181	1.002
				Block 6			
LS	−0.788	0.993	0.992	0.949	0.964	0.987	**0.882**
PITS	0.246	0.354	0.238	0.387	0.256	0.435	0.610
MSPE	0.247	0.989	1.234	1.073	**0.969**	1.157	0.984
				Block 7			
LS	−0.743	0.990	1.022	0.983	0.953	1.039	**0.911**
PITS	0.216	0.307	0.243	0.296	0.246	0.405	0.725
MSPE	0.242	0.991	1.261	1.089	**0.958**	1.158	0.969
				Block 8			
LS	−0.619	1.000	1.219	1.146	0.968	1.040	**0.995**
PITS	0.610	0.547	0.638	0.609	0.718	0.602	0.625
MSPE	0.203	0.995	1.362	1.125	**0.972**	1.243	1.024
				Block 9			
LS	−0.655	0.998	1.132	0.987	0.965	1.079	**0.949**
PITS	0.649	0.518	0.635	0.545	0.741	0.612	0.693
MSPE	0.218	1.002	1.294	1.091	0.979	1.164	**0.973**
				Block 10			
LS	−0.594	1.023	1.190	1.006	**0.951**	1.123	0.998
PITS	0.565	0.825	0.501	0.854	0.532	0.722	0.512
MSPE	0.189	1.011	1.460	1.123	**0.980**	1.278	1.031
				Block 11			
LS	−0.610	0.995	1.284	0.992	0.952	1.048	**0.931**
PITS	0.813	0.716	0.813	0.858	0.883	0.778	0.449
MSPE	0.187	0.991	1.364	1.118	**0.974**	1.224	0.989

NOTES: The table shows the average log score (LS), p-values for the pits test by Knüppel (2015) (PITS) and mean square prediction error (MSPE) for seven different prediction methods applied to dynamic factor models: Bayesian model averaging based on predictive likelihood (BMA), selecting the model with highest recursive score at each point in time (SEL), forecast combination with equal weights (EW), optimal combination (OptComb), the ex-post best performing model our combined density nowcasting approach without learning (CDNNL) and our combined density nowcasting approach with learning (CDN). The results in Columns 2–7 show LS and MSPE relative to the BMA measure. Bold numbers in the rows for LS and MSPE indicate the most accurate model for different statistics. Statistically significant differences between the BMA approach and the alternative combination approaches are denoted by one and two asterisks corresponding to significance levels of 10% and 5%, respectively. Bold numbers in the rows for PITS indicate a rejection of the hypothesis of correctly calibrated densities at a 10% significance level. See Table 1 for information on different blocks.

evident for Block 1 and Block 5 for the 2001 recession, and even more pronounced for the Great Recession, increasing the standard deviation for the combination residual for all blocks. In Section 5.3, we will study the performance of CDN during economic downturns in more detail.

Figure 2 shows the weights associated with the four dynamic factor models for Block 1, Block 5, and Block 11. We notice the large uncertainty on the weights, with substantial variation over time. There is a clear indication that DFMs with either one or two factors obtain higher weights than DFMs with three and four factors. Moreover, the weights also change between the blocks. Finally, the red dotted line in each subfigure shows the corresponding weights obtained by the BMA approach. Comparing the CDN weights with the BMA weights, we see two interesting differences. First, the medians of the CDN weights and BMA weights differ substantially, with much larger

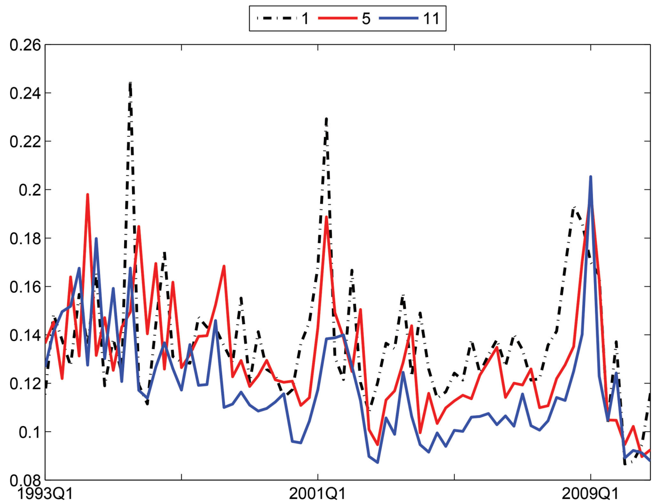

Figure 1. Standard deviation of the combination residuals. Standard deviation of the combination residuals for incomplete model sets from Equation (4), for Block 1, Block 5, and Block 11.

movements over time from the BMA weights. Second, BMA selects much more extreme weights, attaching almost all the weights to one single model, consistent with findings in Amisano and Geweke (2013). The main difference between CDN and BMA is that our weighting scheme allows for model incompleteness (the BMA weights based on predictive likelihood will also take into account past predictive performance scores).

Finally, Figure 3 shows a full set of recursive real-time out-of-sample density nowcasts for U.S. GDP growth for the period 1990Q2–2010Q3 at three different blocks (Blocks 1, 5, and 11). The three panels illustrate how the precision of the predictive densities improves, that is, being more narrow and centered around the actual GDP values as more information becomes available.

5.3 CDN Nowcasting of Negative Growth in the Business Cycle

In a previous subsection, we have shown that CDN provides accurate nowcasts when focusing on the entire distribution of GDP growth. The distribution of CDN can also be used to compute probabilities to be in specific phases of the business cycle. There is a large literature on estimation and timely detection of turning points and economic downturns, see, for example, Harding and Pagan (2002), Chauvet and Piger (2008), Hamilton (2011), Guèrin and Marcellino (2013), Stock and Watson (2014), and Foroni, Guérin, and Marcellino (2015). The individual economists in the Survey of Professional Forecasters (SPF)

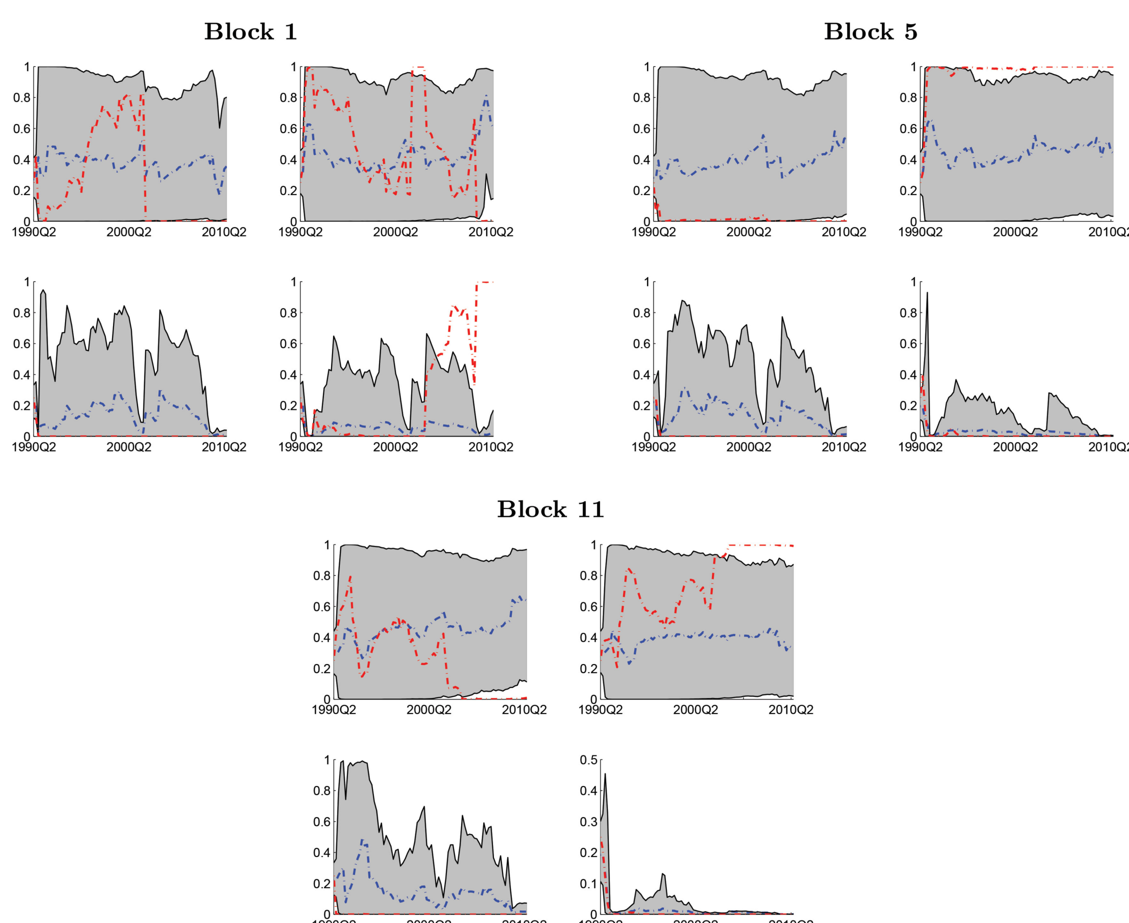

Figure 2. Time-varying weights. The figures plot the 90% credibility intervals of the model posterior weights and their medians (blue dotted lines) for Blocks 1, 5, and 11. The first row of each subfigure shows weights for DFM models with one and two factors. The second row of each subfigure shows weights for DFM models with three and four factors. The red dotted line shows the weights attached to each model using BMA.

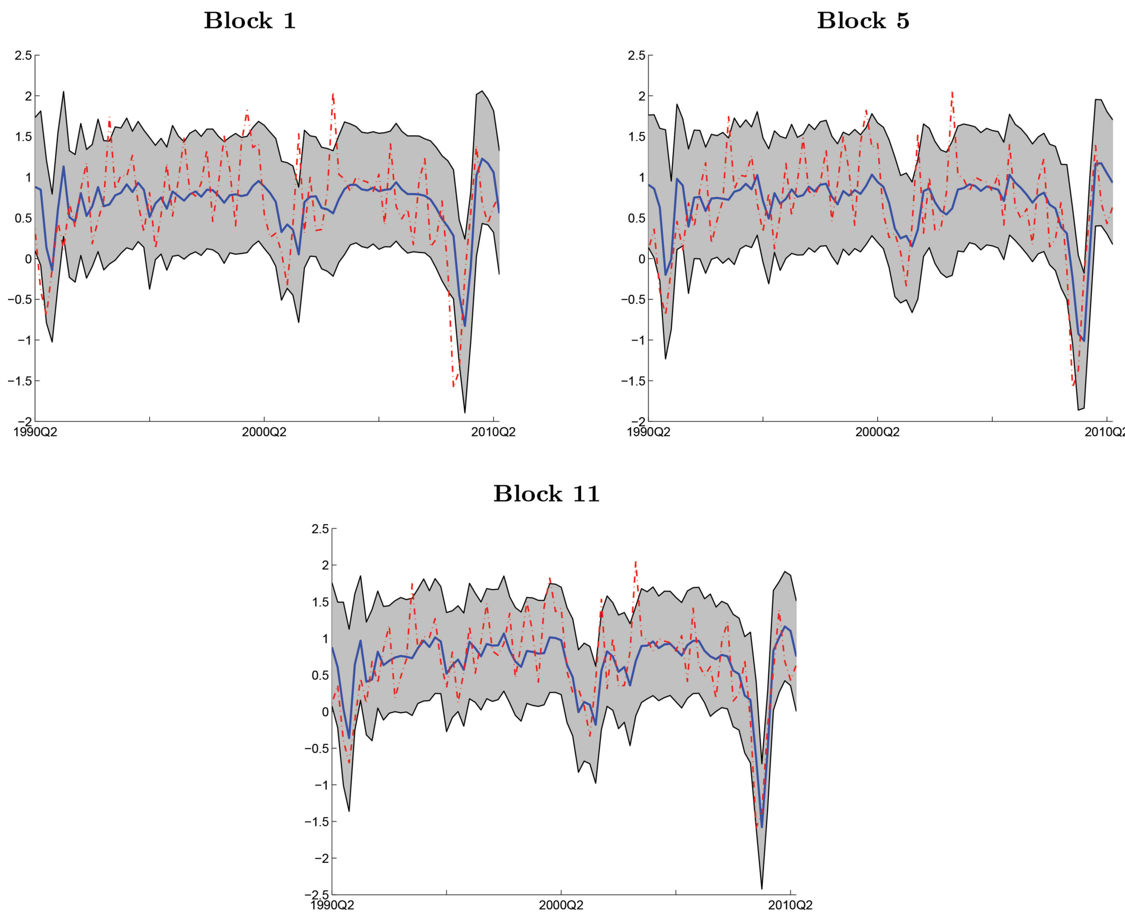

Figure 3. Recursive nowcasts. The figures plot recursive nowcasts for Block 1, 5, and 11. The shaded areas show the 90% credibility intervals of the predictive densities and their medians (blue dotted lines). The red dotted line shows actual GDP, measured as the second release.

also report forecasts of the probability of a decline in the level of real GDP in the current quarter and the following four quarters. Motivated by this, we use CDN to study the probability of negative growth in the current quarter (i.e., GDP growth nowcasts below 0).

Figure 4 compares the recursive probabilities of negative growth in the current quarter from CDN with the mean responses for the probability of negative growth in the current quarter provided by the SPF. To ensure that the information set used to construct the CDN nowcasts are as similar as possible to the information available when the SPF forecasts were made, we report CDN nowcasts for Block 5. Block 5 corresponds to the information set a few days prior to the release of the SPF forecasts. By comparing CDN and SPF forecasts with actual GDP growth (shown by the bars), we find that both CDN and SPF forecasts deliver timely and accurate forecasts of negative growth.

To provide insights about which method is more accurate, we compute concordance statistics (CS), which count the proportion of time during which the predicted and the actual GDP series are in the same state. For convenience, we assume here two states, a state of negative growth and a state of positive growth. We say that a model predicts negative growth for the current quarter if the probability of negative growth is 50% or larger. Comparing the CS for CDN with SPF, we find that both perform equally well with CS $= 0.963$.

Finally, Figure 5 shows the recursive probabilities of negative growth in the current quarter during the period 2007Q1–2009Q4 from the CDN approach for Block 1, Block 5, and Block 11. The figure reveals three interesting observations.

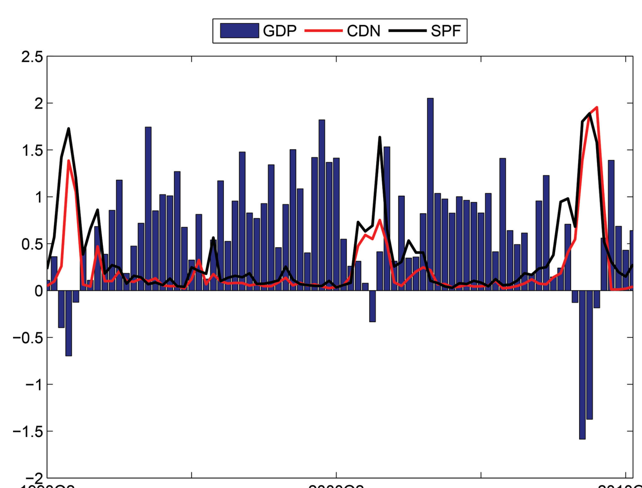

Figure 4. Probabilities of negative growth. Probabilities over time of negative quarterly growth given by the CDN approach and SPF. The red and black lines plot the probabilities scaled by two (therefore covering the interval [0,2]); the bars plot the realization.

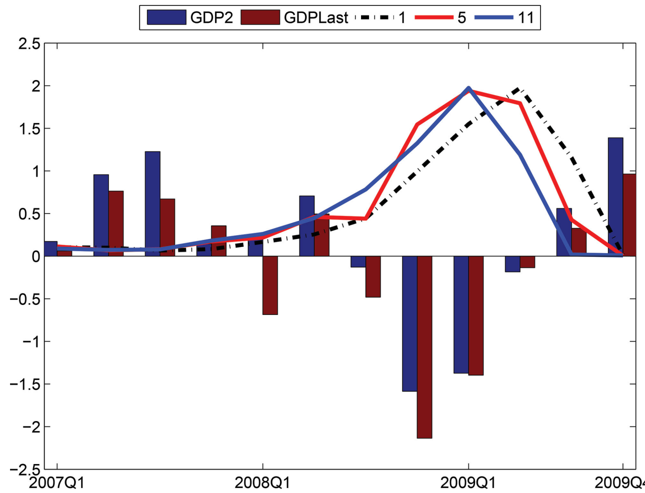

Figure 5. Probabilities of negative growth during the Great Recession period. Probabilities of negative quarterly growth during the Great Recession period provided by the CDN approach at different blocks during the quarter. The black dotted line, and the red and black solid lines plot the probabilities scaled by two (therefore covering the interval [0,2]) from the CDN approach at Block 1, Block 5, and Block 11, respectively. The blue and red bars plot the realizations measured as the second available estimate of GDP and the last available estimate of GDP (November 2014 vintage), respectively, as the actual measure.

First, the probability of obtaining negative growth in the current quarter is very low for all of the blocks during the first quarters of 2007, but starts to increase from 2007Q4. The probability of negative growth for the current quarter continues to increase for each of the quarters throughout 2008. Interestingly, within each quarter the probability of negative growth increases as more information becomes available (i.e., the probability of negative growth is higher for Block 11 than Block 5 and Block 1, and higher for Block 5 than Block 1).

Second, the probability of negative growth in the current quarter starts to fall from May 2009 (Block 5 in 2009Q2). In mid-August 2009 (Block 5 in 2009Q3), the probability of negative growth in the current quarter is for the first time below 0.5 and this probability continues to fall when more information is available throughout the quarter (see Block 11 for 2009Q3). This is consistent with 2009Q3 being the first quarter, where the actual measure of GDP growth is positive. This shows that the CDN not only delivers timely and accurate forecasts for economic downturns, but also provides timely and accurate forecasts of when the economic slump ended.

Third, by comparing the blue and the red bars, which show the realizations of GDP growth measured as the second available estimate of GDP and the last available estimate of GDP (November 2014 vintage), respectively, the figure illustrates that the GDP growth numbers have been revised downwards for all of the quarters in 2008 and 2009, with 2009Q2 as a notable exception. For several of these quarters the downward revisions have been large, exceeding changes of 0.5 percentage point in the quarterly growth rate. This reminds us of how difficult it is to call recessions (or negative growth rates) in real time.

6. CONCLUSION

In this article, we introduced a combined density nowcasting (CDN) approach to dynamic factor models that accounts for the time-varying uncertainty of several model and data features to provide more accurate and complete density nowcasts. The combination weights depend on past nowcasting performance and other learning mechanisms that are incorporated in a Bayesian sequential Monte Carlo method which rebalances the set of nowcasted densities in every period using the updated information on the time varying weights. In this way, we are able to weight data uncertainty, parameter uncertainty, model uncertainty, including model incompleteness, and uncertainty in the combination of weights in a coherent way.

We first implemented simulation experiments to understand the role of incompleteness for nowcasting, distinguishing between data incompleteness (ragged edge problem) and model set incompleteness (the true model is not a part of the forecasters' model space). By comparing point and density nowcasting performance from CDN with the performance of a Bayesian model averaging (BMA) approach, the optimal combination of density forecasts approach (OptComb), equal weights and the ex post best individual model, we find that CDN provides superior nowcasts, particularly at early data releases where there is relatively large data uncertainty and model incompleteness.

We then show the usefulness of CDN when it is applied to four different DFMs for nowcasting GDP growth using U.S. real-time data. The experiment refers to a professional economist who is interested in dealing with various forms of uncertainty in real time. We therefore divide data into different blocks according to their release date within the quarter, and update the density nowcasts at three different points in time during each month of the quarter for the evaluation period 1990Q2–2010Q3.

We find that the CDN outperforms BMA, OptComb, equal weights, a selection strategy, and even the ex-post best individual model in terms of density nowcasting performance for all blocks. The relative gains in terms of improved density nowcasts are also in the empirical analysis larger for the first blocks than for the last blocks of a quarter.

By studying the standard deviation of the combination residuals, we show that this is higher for the earlier blocks in the quarter than for the later blocks in the quarter, confirming that incompleteness plays a larger role in the early part of the quarter. Thus, there are clear gains in terms of improved nowcasting performance from using CDN when incompleteness is present.

Finally, the standard deviations of the combination residuals fluctuate over time and increase during economic downturns. We document that CDN also performs well with respect to focusing on the tails and delivers probabilities of stagnation, measured as the probability of negative growth, that are timely and in line with forecasts from the Survey of Professional Forecasters.

SUPPLEMENTARY MATERIALS

The supplementary materials contain an online appendix where we document the estimation algorithm for our combined density nowcasting (CDN) approach, as well as providing additional details on the data set that we use.

ACKNOWLEDGMENTS

The authors gratefully acknowledge comments and suggestions from the editor, associate editor, two anonymous referees, Roberto Casarin, Dimitris Korobilis, Per Mykland, Richard Paap, Barbara Rossi, Shaun Vahey, Jonathan Wright and Mark Watson as well as seminar and conference participants at the Bank of Portugal, Columbia University, Norges Bank, the ASSA meeting in Philadelphia, the ESEM meeting in Toulouse, the IAEE meeting in London, the ISF meeting in Rotterdam, the Real-Time Data Analysis, Methods and Applications Conference at the Federal Reserve Bank of Philadelphia, the Workshop on Short-term Forecasting at the National Bank of Poland, and the 2015 NBER-NSF Time Series Conference at Vienna University of Economics and Business. The views expressed are those of the authors and do not necessarily reflect those of Norges Bank.

[Received December 2014. Revised December 2015.]

REFERENCES

Aastveit, K., Gerdrup, K., Jore, A., and Thorsrud, L. (2014), "Nowcasting GDP in Real Time: A Density Combination Approach," *Journal of Business & Economic Statistics*, 32, 48–68. [132,133,134,135,136]

Amisano, G., and Geweke, J. (2013), "Prediction Using Several Macroeconomic Models," Working Paper Series, European Central Bank 1537, European Central Bank. [141]

Amisano, G., and Giacomini, R. (2007), "Comparing Density Foreasts via Weighted Likelihood Ratio Tests," *Journal of Business & Economic Statistics*, 25, 177–190. [136]

Angelini, E., Camba-Mendez, G., Giannone, D., Reichlin, L., and Rünstler, G. (2011), "Short-Term Forecasts of Euro Area gdp Growth," *Econometrics Journal*, 14, C25–C44. [136]

Bai, J., and Ng, S. (2006), "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-augmented Regressions," *Econometrica*, 74, 1133–1150. [133]

Banbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2012), "Nowcasting and the Real-time Data Flow," in *Handbook of Economic Forecasting*(Vol. 2), eds. G. Elliott and A. Timmermann, North Holland: Elsevier. [134]

Banbura, M., Giannone, D., and Reichlin, L. (2011), "Nowcasting," in *The Oxford Handbook of Economic Forecasting*, eds. M. P. Clements and D.F. Hendry, Oxford: Oxford Handbooks in Economics. [131]

Banbura, M., and Modugno, M. (2014), "Maximum Likelihood Estimation of Factor Models on Data Sets With Arbitrary Pattern of Missing Data," *Journal of Applied Econometrics*, 29, 133–160. [131,133]

Banbura, M., and Rünstler, G. (2011), "A Look Into the Factor Model Black Box: Publication Lags and the Role of Hard and Soft Data in Forecasting Gdp," *International Journal of Forecasting*, 27, 333–346. [131,136]

Bates, J., and Granger, C. (1969), "The Combination of Forecasts," *Operations Research Quarterly*, 20, 451–468. [131]

Berkowitz, J. (2001), "Testing Density Forecasts, With Applications to Risk Management," *Journal of Business and Economic Statistics*, 19, 465–474. [136]

Billio, M., Casarin, R., Ravazzolo, F., and van Dijk, H. K. (2013), "Time-Varying Combinations of Predictive Densities Using Nonlinear Filtering," *Journal of Econometrics*, 177, 213–232. [132,133,134]

Boivin, J., and Ng, S. (2005), "Understanding and Comparing Factor-based Forecasts," *International Journal of Central Banking*, 1, 117–151. [131]

Carriero, A., Clark, T. E., and Marcellino, M. (2015), "Realtime Nowcasting With a Bayesian Mixed Frequency Model With Stochastic Volatility," *Journal of the Royal Statistical Society*, Series A, 178, 837–862. [135,138]

Casarin, R., Grassi, S., Ravazzolo, F., and van Dijk, H. K. (2015), "Parallel Sequential Monte Carlo for Efficient Density Combination: The DeCo MATLAB Toolbox," *Journal of Statistical Software*, 68. [132]

Chauvet, M., and Piger, J. (2008), "Comparison of the Real-time Performance of Business Cycle Dating Methods," *Journal of Business and Economic Statistics*, 26, 42–49. [141]

Clark, T. E. (2011), "Real-Time Density Forecasts From Bayesian Vector Autoregressions With Stochastic Volatility," *Journal of Business & Economic Statistics*, 29, 327–341. [138]

Clark, T. E., and McCracken, M. W. (2009), "Improving Forecast Accuracy by Combining Recursive and Rolling Forecasts," *International Economic Review*, 50, 363–395. [131]

——— (2010), "Averaging Forecasts From Vars With Uncertain Instabilities," *Journal of Applied Econometrics*, 25, 5–29. [131,136]

Clark, T. E., and Ravazzolo, F. (2015), "The Macroeconomic Forecasting Performance of Autoregressive Models With Alternative Specifications of Time-varying Volatility," *Journal of Applied Econometrics*, 30, 551–575. [138]

Croushore, D., and Stark, T. (2001), "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics*, 105, 111–130. [136]

Diebold, F. X. (1991), "A Note on Bayesian Forecast Combination Procedures," in *Economic Structural Change: Analysis and Forecasting*, eds. A. Westlund and P. Hackl, New York: Springer-Verlag, pp. 225–232. [135]

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998, November), "Evaluating Density Forecasts With Applications to Financial Risk Management," *International Economic Review*, 39, 863–83. [136]

Diebold, F. X., and Mariano, R. S. (1995), "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13, 253–63. [138]

Evans, M. D. (2005), "Where Are We Now? Real-time Estimates of the Macro Economy," *International Journal of Central Banking*, 1, 127–175. [131]

Fernandez, C., Ley, E., and Steel, M. F. J. (2001), "Model Uncertainty in Cross-country Growth Regressions," *Journal of Applied Econometrics*, 16, 563–576. [135]

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005), "The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting," *Journal of the American Statistical Association*, 100, 830–840. [131]

Foroni, C., Guérin, P., and Marcellino, M. (2015), "Markov-Switching Mixed-Frequency VAR Models," *International Journal of Forecasting*, 31, 692–711. [141]

Foroni, C., and Marcellino, M. (2014), "A Comparison of Mixed Frequency Approaches for Nowcasting Euro Area Macroeconomic Aggregates," *International Journal of Forecasting*, 30, 554–568. [133]

Foroni, C., Marcellino, M., and Schumacher, C. (2015), "Unrestricted Mixed Data Sampling (MIDAS): MIDAS Regressions With Unrestricted Lag Polynomials," *Journal of the Royal Statistical Society*, Series A, 178, 57–82. [135]

Geweke, J., and Amisano, G. (2010), "Comparing and Evaluating Bayesian Predictive Distributions of Asset Returns," *International Journal of Forecasting*, 26, 216–230. [135]

——— (2011), "Optimal Prediction Pools," *Journal of Econometrics*, 164, 130–141. [132,135,139]

Giannone, D., Reichlin, L., and Small, D. (2008), "Nowcasting: The Real-Time Informational Content of Macroeconomic Data," *Journal of Monetary Economics*, 55, 665–676. [131,133,134,139]

Gneiting, T. (2011), "Making and Evaluating Point Forecasts," *Journal of the American Statistical Association*, 106, 746–762. [134]

Guèrin, P., and Marcellino, M. (2013), "Markov-Switching MIDAS Models," *Journal of Business & Economic Statistics*, 31, 45–56. [141]

Hall, S. G., and Mitchell, J. (2007), "Combining Density Forecasts," *International Journal of Forecasting*, 23, 1–13. [132,135,136,139]

Hamilton, J. D. (2011), "Calling Recessions in Real Time," *International Journal of Forecasting*, 27, 1006–1026. [141]

Harding, D., and Pagan, A. (2002), "Dissecting the Cycle: A Methodological Investigation," *Journal of Monetary Economics*, 49, 365–381. [141]

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999), "Bayesian Model Averaging: A Tutorial," *Statistical Science*, 14, 382–417. [135]

Jore, A. S., Mitchell, J., and Vahey, S. P. (2010), "Combining Forecast Densities From Vars With Uncertain Instabilities," *Journal of Applied Econometrics*, 25, 621–634. [132,135]

Kascha, C., and Ravazzolo, F. (2010), "Combining Inflation Density Forecasts," *Journal of Forecasting*, 29, 231–250. [136]

Knüel, M. (2015), "Evaluating the Calibration of Multi-Step-Ahead Density Forecasts Using Raw Moments," *Journal of Business & Economic Statistics*, 33, 270–281. [136,138,140]

Koop, G., and Korobilis, D. (2012), "Forecasting Inflation Using Dynamic Model Averaging," *International Economic Review*, 53, 867–886. [132,133]

Koop, G., and Potter, S. (2004), "Forecasting in Dynamic Factor Models Using Bayesian Model Averaging," *Econometrics Journal*, 7, 550–565. [131,133]

Kuzin, V., Marcellino, M., and Schumacher, C. (2013), "Pooling Versus Model Selection for Nowcasting GDP With Many Predictors: Empirical Evidence for Six Industrialized Countries," *Journal of Applied Econometrics*, 28, 392–411. [132]

Liu, J., and West, M. (2001), "Combined Parameter and State Estimation in Simulation-Based Filtering," in *Sequential Monte Carlo Methods in Practice*, eds. A. Doucet, J.D. Freitas, and, N. Gordon, New York: Springer-Verlag, pp. 197–217. [135]

Marcellino, M., Porqueddu, M., and Venditti, F. (2015), "Short-Term GDP Forecasting With a Mixed Frequency Dynamic Factor Model With Stochastic Volatility," *Journal of Business & Economic Statistics,* 34, 118–127. [133,138]

Marcellino, M., and Schumacher, C. (2010), "Factor MIDAS for Nowcasting and Forecasting With Ragged-Edge Data: A Model Comparison for

German GDP," *Oxford Bulletin of Economics and Statistics*, 72, 518–550. [134]

Marcellino, M., Stock, J., and Watson, M. (2006), "A Comparison of Direct and Iterated Multistep ar Methods for Forecasting Macroeconomic Time Series," *Journal of Econometrics*, 135, 499–526. [133,135]

Mitchell, J., and Hall, S. G. (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, 995–1033. [135,136]

Pettenuzzo, D., and Ravazzolo, F. (2015), "Optimal Portfolio Choice Under Decision-based Model Combinations," *Journal of Applied Econometrics*, doi:10.1002/jae.2502. [134]

Romer, C. D., and Romer, D. H. (2000), "Federal Reserve Information and the Behavior of Interest Rates," *American Economic Review*, 90, 429–457. [137]

Sala-I-Martin, X., Doppelhofer, G., and Miller, R. I. (2004), "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach," *American Economic Review*, 94, 813–835. [135]

Stark, T., and Croushore, D. (2002), "Forecasting With a Real-Time Data Set for Macroeconomists," *Journal of Macroeconomics*, 24, 507–531. [137]

Stock, J., and Watson, M. (2002a), "Forecasting Using Principal Components From a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179. [131]

——— (2002b), "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business & Economic Statistics*, 20, 147–62. [131]

——— (2004), "Combining Forecasts of Output Growth in Seven-country Data Set," *Journal of Forecasting*, 23, 405–430. [134]

——— (2006), *Forecasting With Many Predictors of Handbook of Economic Forecasting* (Vol. 1), Chapter 10, New York: Elsevier, pp. 515–554. [131,136]

——— (2014), "Estimating Turning Points Using Large Data Sets," *Journal of Econometrics*, 178, 368–381. [141]

Strachan, R. W., and Dijk, H. K. V. (2013), "Evidence on Features of a Dsge Business Cycle Model From Bayesian Model Averaging," *International Economic Review*, 54, 385–402. [133]

Timmermann, A. (2006), "Forecast Combinations," in *Handbook of Economic Forecasting*(Vol. 1), eds. G. Elliott, C. W. J. Granger, and A. Timmermann, Amsterdam: Elsevier, pp. 136–96. [132,136]

Wallis, K. F. (1986), "Forecasting With an Econometric Model: The 'Ragged Edge' Problem," *Journal of Forecasting*, 5, 1–13. [131]

West, M. (1992), "Modelling Agent Forecast Distributions," *Journal of the Royal Statistical Society*, Series B, 54, 553–567. [132]

West, M., and Crosse, J. (1992), "Modelling of Probabilistic Agent Opinion," *Journal of the Royal Statistical Society*, Series B, 54, 285–299. [132]