working paper

12 27

Real-Time Nowcasting with a Bayesian Mixed Frequency Model with Stochastic Volatility

Andrea Carriero, Todd E. Clark, and
Massimiliano Marcellino

FEDERAL RESERVE BANK OF CLEVELAND

Working Paper 12-27 November 2012

Real-Time Nowcasting with a
Bayesian Mixed Frequency Model with Stochastic Volatility

Andrea Carriero, Todd E. Clark, and Massimiliano Marcellino

This paper develops a method for producing current-quarter forecasts of GDP growth with a (possibly large) range of available within-the-quarter monthly observations of economic indicators, such as employment and industrial production, and financial indicators, such as stock prices and interest rates. In light of existing evidence of time variation in the variances of shocks to GDP, we consider versions of the model with both constant variances and stochastic volatility. We also evaluate models with either constant or time-varying regression coefficients. We use Bayesian methods to estimate the model, in order to facilitate providing shrinkage on the (possibly large) set of model parameters and conveniently generate predictive densities. We provide results on the accuracy of nowcasts of real-time GDP growth in the U.S. from 1985 through 2011. In terms of point forecasts, our proposal is comparable to alternative econometric methods and survey forecasts. In addition, it provides reliable density forecasts, for which the stochastic volatility specification is quite useful, while parameter time-variation does not seem to matter.

Keywords: Prediction, forecasting, Bayesian methods, mixed frequency models.

JEL Codes: E37, C53, C22.

Andrea Carriero is at Queen Mary, University of London, and can be reached at a.carriero@qmul.ac.uk. Todd E. Clark is at the Federal Reserve Bank of Cleveland and can be reached at todd.clark@clev.frb.org. Massimiliano Marcellino is at the European University Institute, Bocconi University, and CEPR and can be reached at massimiliano.marcellino@eui.eu. The authors gratefully acknowledge helpful conversations with Brent Meyer, helpful comments from Knut Are Aastveit, Marta Banbura, Domenico Giannone, Christian Schumacher, and seminar participants at the Cleveland Fed, and research assistance from Claudia Foroni and John Lindner.

1 Introduction

Nowcasting has come to be commonly viewed as an important and unique forecasting problem; see, e.g., Banbura, Giannone, and Reichlin (2011) and Banbura, Giannone, Modugno, and Reichlin (2012). It is important because current-quarter forecasts of GDP growth and inflation provide useful summaries of recent news on the economy and because these forecasts are commonly used as inputs to forecasting models, such as some of the DSGE models in use at central banks, that are effective in medium-term forecasting but not necessarily short-term forecasting. As studies such as Faust and Wright (2009, 2012) have emphasized, initial-quarter forecasts often play a key role in the accuracy of forecasts at subsequent horizons. Nowcasting is unique in that, to some degree, it involves "simply" adding up information in data releases for the current quarter. 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 — what Wallis (1986) refers to as the "ragged edge" of data.

The nowcasting approach we will propose in this paper is motivated by not only some of the previous nowcasting work, reviewed in section 2, but also by three other key findings in the broader forecasting literature. First, prior work, particularly De Mol, Giannone, and Reichlin (2008), Banbura, Giannone, and Reichlin (2010) and Carriero, Kapetanios, and Marcellino (2011), has shown that, with large data sets, estimation with Bayesian shrinkage is a viable alternative to factor model methods. Second, Clark (2011), Carriero, Clark, and Marcellino (2012), and D'Agostino, Gambetti, and Giannone (2012) find it useful for forecasting purposes to incorporate stochastic volatility into VAR models, for both point and density forecasts. Third, some other prior work has shown that direct multi-step methods of forecasting can be at least as accurate as iterated methods (e.g., Marcellino, Stock and Watson (2006)) for multi-step forecasting. At a forecast horizon of $h > 1$, the direct approach rests on estimates of a model relating y_{t+h} to information in period t. The iterated approach involves a model relating y_{t+1} to information in period t and iterating forward to obtain 2-step forecasts from 1-step, etc. The direct approach can be more accurate than the iterated approach in the presence of model misspecification and does not require modeling the behavior of the explanatory variables, thus making univariate modelling sufficient. To be clear, the model we develop isn't literally a direct multi-step model, but it is clearly in the spirit of such a forecasting specification.

Building on this past work, we develop a new Bayesian Mixed Frequency (BMF) model with Stochastic Volatility (SV) for point and density nowcasting. Our formulation also readily allows the regression coefficients to be time-varying. In particular, we produce current-quarter forecasts of GDP growth with a (possibly large) range of available within-the-quarter monthly observations of economic indicators, such as employment and industrial production, and financial indicators, such as stock prices and interest rates. Each time series of monthly indicators is transformed into three quarterly time series, each containing observations for, respectively, the first, second or third month of the quarter. Hence, there can be missing observations at the end of some of these three time series, depending on the specific month of the quarter we are in. We then include in the model only the constructed quarterly series without missing observations at the moment in time the forecast is formed. This approach, which is in the spirit of direct multi-step forecasting, addresses the ragged edge of the data.

We use Bayesian methods to estimate the resulting model, which expands in size as more monthly data on the quarter become available. Bayesian estimation facilitates providing shrinkage on estimates of a model that can be quite large, conveniently generates predictive densities, and readily allows for stochastic volatility and time-varying parameters.

We provide results on the accuracy of the resulting nowcasts of real-time GDP growth in the U.S. from 1985 through 2011. While most prior nowcasting research has focused on the accuracy of point forecasts of GDP growth (with the notable exception of Aastveit, et al. (2011) and Marcellino, et al. (2012)), we consider both point and density forecasts. It turns out that in terms of point forecasts our proposal is comparable to alternative mixed frequency econometric methods and survey forecasts. In addition, it provides reliable density forecasts, for which the stochastic volatility specification is quite useful, while parameter time-variation does not seem to matter.

As we will detail below, in the schema of Banbura, Giannone, Modugno, and Reichlin (2012), our proposed model falls in the 'partial models' set. Our approach is most closely related to the U-MIDAS specification of Foroni, et al. (2011), which they estimate with classical methods and the same number of regressors in each time period, with the actual regressor set determined by the BIC criterion. Compared to other partial model approaches or studies, the key innovations of our analysis include the use of Bayesian shrinkage to be able to consider a possibly large set of indicators; the allowance for time variation in coefficients; the inclusion of stochastic volatility; and the analysis of density nowcasts.

Our approach of including different variables corresponding to different months of the quarter also has some similarities to vector autoregressive (full system) approaches used in recent studies by Ghysels (2012), Foroni and Marcellino (2012) and McCracken and Sekhposyan (2012). For example, in Ghysel's VAR(p) with mixed frequency data, an equation for GDP growth in quarter t includes p lags of quarterly GDP and employment in each of months 1-3 of quarter $t-1$, quarter $t-2$, etc., up to quarter $t-p$. Ghysels suggests using MIDAS-type restrictions to limit the number of parameters. In our case, in which the model of interest is univariate, we propose using Bayesian shrinkage to limit the potential effects of parameter estimation error on forecast accuracy. And, again, our approach has the advantage of making stochastic volatility and time-varying parameters tractable.

The paper is structured as follows. After a brief review of the related literature in Section 2, we proceed by detailing our model and estimation method in section 3, introducing competing nowcasts in section 4, describing the data in section 5, presenting results in section 6, and providing some concluding remarks in section 7.

2 A Brief Review of Related Nowcasting Approaches

A number of model-based approaches are commonly used for nowcasting, primarily with a focus on point forecasts. Banbura, Giannone, Modugno, and Reichlin (2012) distinguish 'partial model' methods from approaches thought to be more comprehensive, which might simply be labeled 'full system' methods. In this section, we briefly review some of the key methods falling in these two categories. In section 3, after detailing our models, we will describe how our proposed approach relates to these key methods.

Among partial model methods, bridging and MIDAS models are most commonly used. Bridging models, considered in such studies as Baffigi, Golinelli and Parigi (2004), Diron (2008) and Bencivelli, Marcellino and Moretti (2012), relate the period t value of the quarterly variable of interest, such as GDP growth, to the period t quarterly average of key monthly indicators.[1] The period t average of each monthly indicator is obtained with data

[1] One common approach is to use bridging equations to forecast components of GDP growth and then add up forecasts of components to obtain a forecast of GDP growth.

available within the quarter and forecasts for other months of the quarter (obtained typically from an autoregressive model for the monthly indicator).

MIDAS-based models, developed in Ghysels, Santa-Clara and Valkanov (2004) for financial applications and applied in, e.g., Clements and Galvao (2008) and Guerin and Marcellino (2012) for macroeconomic applications, relate the period t value of the quarterly variable of interest to a constrained distributed lag of monthly or weekly or even daily data on the predictors of interest. The resulting model is then estimated by non-linear least squares and used to forecast the variable of interest from constrained distributed lags of the available data. Foroni, Marcellino and Schumacher (2012) propose the use of unconstrained distributed lags of the high frequency indicators. They label the resulting model unrestricted MIDAS, or U-MIDAS, and show that it generally outperforms MIDAS when the frequency mismatch between the low frequency target variable and the high frequency indicators is limited. Rodriguez and Puggioni (2010) discuss Bayesian estimation of unrestricted MIDAS equations. They also allow for time-varying parameters but not for stochastic volatility, and consider point but not density forecasts.

Full system methods for nowcasting include factor models and mixed frequency VARs.[2] In particular, the factor model of Giannone, Reichlin, and Small (2008) provides a sophisticated method for accommodating the ragged edge of data and using a large information set to forecast the variable of interest. Their model relates the variable of interest, such as quarterly GDP growth, to current and possibly lagged values of a set of factors that summarize the information in a large data set of monthly or even weekly indicators. The factors are modeled as following a VAR process. By casting the factor model in state space form with a measurement equation relating monthly data to the factors and a state equation based on the VAR model for the factors, Giannone, Reichlin, and Small (2008) are able to use the Kalman Filter to obtain current quarter factor estimates or forecasts that reflect all

[2]Some prior studies have proposed using factors based on large datasets in underlying modeling approaches that fall in the partial model category, or combining factor model approaches with others. In particular, Marcellino and Schumacher (2010) develop a factor-MIDAS approach to incorporate large datasets of high frequency information into a nowcast model. In their application to nowcasting German GDP, the factor-MIDAS approach compares well with the Kalman filter-based factor model approach. Kuzin, Marcellino and Schumacher (2012) propose pooling the nowcasts resulting from a large set of mixed frequency single indicator models. Finally, Aastveit, et al. (2011) develop a nowcasting system that combines forecasts from VAR models, bridging equations, and factor models. In contrast to most of the nowcasting literature, Aastveit, et al. focus on density forecasts.

available information, taking account of the ragged edge of data.

While much recent full system work has focused on large factor models, some studies have developed smaller factor models. Mariano and Murasawa (2003) extended the factor model of Stock and Watson (1989) to handle the ragged edge and mixed frequency data. Their approach is used to model in a Kalman filter context the demand and supply components of euro area GDP by Frale, Marcellino, Mazzi and Proietti (2011), with all the resulting index estimates later combined into euromind, a monthly indicator of euro area economic conditions. A related approach is proposed by Camacho and Perez-Quiros (2010). The method is further refined in Marcellino, Porqueddu and Venditti (2012) to allow for stochastic volatility in the common and idiosyncratic components, and provide density forecasts.

Still other studies have developed mixed frequency VAR models, introduced by Zadrozny (1988), for nowcasting. Foroni, Ghysels, and Marcellino (2012) provide an overview of this method. Most recently, Schorfheide and Song (2012) develop a mixed frequency Bayesian VAR for forecasting that they find to work well in real-time forecasting.[3] For nowcasting Euro area GDP, Kuzin, Marcellino and Schumacher (2011) find mixed frequency VARs to be comparable to MIDAS-based approaches.

3 The Bayesian Mixed Frequency Model with Stochastic Volatility (BMF-SV)

This section details our proposed nowcasting models, starting with the case of constant error variances and then taking up the cases of stochastic volatility and time-varying regression coefficients. The section then discusses how our approach relates to others in the literature. The section concludes by detailing the indicators used in the models and the priors and algorithms used in estimation.

[3] Chui, et al. (2011) also develop a mixed frequency BVAR. In addition, Ghysels (2012) and Foroni and Marcellino (2012) analyze structural mixed frequency VAR models, showing that higher frequency information can be helpful also in this context.

3.1 (BMF) Model with constant volatility

Starting with our specifications that treat the error variance of the model as constant over time, we consider nowcasting the quarterly growth rate of GDP in month m of the current quarter based on the regression:

$$y_t = X'_{m,t}\beta_m + v_{m,t}, \quad v_{m,t} \sim i.i.d.N(0, \sigma_m^2), \tag{1}$$

where the vector $X_{m,t}$ contains the available predictors at the time the forecast is formed, t is measured in quarters, and m indicates a month (as detailed below).

The specification of the regressor vector $X_{m,t}$ is partly a function of the way we sample the monthly variables. For each monthly variable, we first transform it at the monthly frequency as necessary to achieve stationarity. At the quarterly frequency, for each monthly variable, we then define three different variables, by sampling the monthly series separately for each month of the quarter.

Exactly what variables are included in $X_{m,t}$ also depends on when in the quarter the forecast is formed. We consider four different timings for forecasting period t GDP growth: forecasting at the end of the first week of month 1 of quarter t ($m = 1$), at the end of the first week of month 2 of quarter t ($m = 2$), at the end of the first week of month 3 ($m = 3$), and at the end of the first week of month 1 of quarter $t + 1$ ($m = 4$). These points in time are chosen to correspond to the usual timing of the publication of employment data: employment data for month s are normally published at the end of the first week of month $s+1$. At these points in time, the availability of other indicators varies across indicators. As a consequence, the model specification periodically changes in each month of the quarter, and this also takes care of the ragged edge of the data.

At each of the four forecast origins we consider for each quarter t, the regressor set $X_{m,t}$ is specified to include the subset of variables available for t (details provided below). It is in this sense that our proposed approach is very much in the spirit of direct multi-step forecasting. Under a direct approach, for forecasting some variable in the future as of period t, one puts on the right hand side of the regression model the variables available in period t. Our model is similar in that we define the set of explanatory variables at each moment in time to consist only of the variables for which period t observations are available. We should stress that this approach does not involve bridging methods. Bridging methods require forecasting monthly observations of monthly variables for any months of quarter t for which data are not yet available. We do not use such forecasts. Rather, we only put on the right hand side of the regression model the actual monthly observations that are available for the quarter, in the form of different quarterly variables associated with the different months of the quarter. In this sense, what we do is similar to the blocking approach in the engineering literature, see, e.g., Chen, et al. (2012).

Our nowcasting models of the form of the above equation (1) also include a constant and one lag of GDP growth. In most cases, this means the models include GDP growth in period $t-1$. However, in the case of models for forecasting at the end of the first week of month 1 of quarter t, the value of GDP growth in period $t-1$ is not actually available in real time. In this case, the model includes GDP growth in period $t-2$. This is consistent with our general direct multi-step specification of the forecasting models. Finally, while the preceding discussion has focused on current quarter values of the monthly variables, for most of the models we also consider versions in which the period $t-1$ (previous quarter) values of every variable is included as a predictor.

The largest model we consider includes over 50 explanatory variables in $X_{m,t}$ (reflecting 9 different monthly indicators in this largest model, 2-3 quarterly series (corresponding to different months of the quarter) for each of them, and both current and one period lagged values of the quarterly variables). Accordingly, with simple OLS estimation, parameter estimation error would have large adverse effects on forecast accuracy. Our Bayesian approach to estimation incorporates shrinkage to help limit the effects of parameter estimation error on forecast accuracy. We ran some checks with some of our basic models to verify the importance of this shrinkage to nowcast accuracy. These checks showed that models without shrinkage yielded RMSEs 14 to 26 percent higher and average log scores 9 to 21 percent lower than the same models estimated with shrinkage (specifically, with the prior settings described below).[4]

[4]More specifically, in these checks we compared the results for our specifications with just coincident indicators against results for the same models estimated with extremely loose (essentially OLS-replicating) priors on the regression coefficients, using $\lambda_1 = \lambda_2 = 1000$.

3.2 (BMF-SV) Model with stochastic volatility

In the stochastic volatility (SV) case, our proposed forecasting model takes the form

$$\begin{aligned} y_t &= X'_{m,t}\beta_m + v_{m,t} \\ v_{m,t} &= \lambda^{0.5}_{m,t}\epsilon_{m,t},\ \epsilon_{m,t} \sim i.i.d.N(0,1) \\ \log(\lambda_{m,t}) &= \log(\lambda_{m,t-1}) + \nu_{m,t},\ \nu_{m,t} \sim\ i.i.d.N(0,\phi_m), \end{aligned} \tag{2}$$

with $m = 1,2,3,4$. Following the approach pioneered in Cogley and Sargent (2005) and Primiceri (2005), the log of the conditional variance of the error term in equation (2) follows a random walk process. In a vector autoregressive context, studies such as Clark (2011), Carriero, Clark, and Marcellino (2012), and D'Agostino, Gambetti, and Giannone (2012) have found that this type of stochastic volatility formulation improves the accuracy of both point and density forecasts. Apart from the modification of the volatility process, the model takes the same form given in the preceding subsection, in terms of timing and variables included in $X_{m,t}$.[5]

3.3 (BMF-TVP and BMF-TVP-SV) Models with time-varying parameters

Some previous forecasting analyses (e.g., Rodriguez and Puggioni (2010) and D'Agostino, Gambetti, and Giannone (2012)) have found time-varying parameters (TVP) to improve forecast accuracy. One of the advantages of our BMF and BMF-SV models is that they can be readily extended to allow the regression coefficient vector β_m to be time varying. In these cases, the coefficient vector becomes $\beta_{m,t}$, which follows a random walk process:

$$\beta_{m,t} = \beta_{m,t-1} + n_{m,t}, \quad n_{m,t} \sim\ i.i.d.N(0,Q_m).$$

We consider below the efficacy of TVP for nowcasting using a few of our available variable combinations.

3.4 Relationship of our BMF approach to other nowcasting approaches

As noted in the introduction, our proposed modeling approach falls in the 'partial models' set. It is most similar to the univariate UMIDAS method of Foroni, et al. (2011). Relative to

[5] We have also considered a stationary specification for the evolution of log volatility, an AR(1) model with a coefficient of 0.9. Overall, this model performed worse than the random walk specification.

other partial model approaches, the innovations in our approach include the use of Bayesian shrinkage, the allowance for time variation in coefficients, and the inclusion of stochastic volatility. Bayesian shrinkage permits us to include a potentially large set of indicators, which some evidence (e.g., De Mol, Giannone, and Reichlin (2008)) suggests should permit our model to achieve forecast accuracy comparable to factor models (full system methods). The use of direct-type estimation means we do not need to model explicitly the conditioning variables. Moreover, with the univariate forecasting equation of our approach, we are able to allow for stochastic volatility and time-varying regression coefficients, two possibly important features to improve the nowcasting performance, particularly for density forecasting, mostly neglected so far in this literature. Accordingly, while we recognize the merits of full-system factor model methods emphasized in Banbura, Giannone, Modugno, and Reichlin (2012), our proposed approach offers simplicity, avoids the need for modeling of conditioning variables, and makes time-varying parameters and stochastic volatility easily tractable.

3.5 Indicators used in BMF and BMF-SV models

In applying our proposed models to nowcasting GDP growth, we consider various combinations of monthly indicators, using coincident, leading, and financial indicators. We chose these particular indicators to be broadly representative of major economic and financial indicators, with some eye to timeliness.

The monthly indicators we use are as follows (Section 5 will provide data details):

- **Coincident**: payroll employment ($\Delta \log$); industrial production ($\Delta \log$); real retail sales ($\Delta \log$); housing starts (log); and the ISM index (overall) for manufacturing.
- **Leading**: the ISM index for supplier delivery times; the ISM index for orders; average weekly hours of production and supervisory workers ($\Delta \log$); and new claims for unemployment insurance.
- **Financial**: stock prices as measured by the S&P 500 index ($\Delta \log$); the 10-year Treasury bond yield; and the 3-month Treasury bill rate.

Table 1 details the model specifications (and variable timing) we use, based on the usual publication schedules of the indicators.

We did not engage in a broad search for best indicators or endeavor to make comparisons of these indicators to others found to work well in some studies. Of course, there are a range of others that could be worth considering. For example, if one were producing forecasts in the middle of the month (rather than early in the month as we do), the Federal Reserve Bank of Philadelphia's business survey would be worth considering (as in such studies as Giannone, Reichlin, and Small (2008)).[6] Moreover, in future research, it might also be worth considering indicators reported at a weekly or daily frequency. While our method can easily handle these higher frequencies, we focus our application on monthly indicators, in light of the finding by Banbura, et al. (2012) that higher frequency information does not seem to be especially useful for nowcasting U.S. GDP growth (except perhaps in a continuous monitoring context).

3.6 Priors

We estimate the models with constant volatility using a normal-diffuse prior. As detailed in sources such as Kadiyala and Karlsson (1997), this prior combines a normal distribution for the prior on the regression coefficients with a diffuse prior on the error variance of the regression. For the models with stochastic volatility, we use independent priors for the coefficients (normal distribution) and volatility components (details below). Since the choice of the prior is not dependent on m, in spelling out the prior we drop the index m from the model parameters for notational simplicity.

In all cases, for the coefficient vector β, we use a prior distribution that is normal, with mean 0 (for all coefficients) and variance that takes a diagonal, Minnesota-style form. The prior variance is Minnesota style in the sense that shrinkage increases with the lag (with the quarter, not with the month within the quarter), and in the sense that we impose more shrinkage on the monthly predictors than on lags of GDP growth (for the basic coincident indicator model, loosening up the cross-variable shrinkage didn't improve results). The shrinkage is controlled by three hyperparameters (in all cases, a smaller number means more shrinkage): λ_1, which controls the overall rate of shrinkage; λ_2, which controls the rate of shrinkage on variables relative to GDP; and λ_3, which determines the rate of shrinkage

[6] We ran a few checks (with our early-month model timing) to see if replacing the ISM with the Philadelphia Fed's business survey would improve forecast accuracy. Using the business survey yielded results either similar to or not quite as good as those obtained with the ISM.

associated with longer lags.

At each forecast origin, the prior standard deviation associated with the coefficient on variable $x_{i,j,t-l}$ of $X_{m,t}$, where i denotes the indicator (employment, etc.), j denotes the month within which the quarter at which the indicator has been sampled, and l denotes the lag in quarters, is specified as follows:

$$\mathrm{sd}_{i,j,t-l} = \frac{\sigma_{GDP}}{\sigma_{i,j}} \frac{\lambda_1 \lambda_2}{l^{\lambda_3}}. \tag{3}$$

For coefficients on lag l of GDP, the prior standard deviation is

$$\mathrm{sd}_l = \frac{\lambda_1}{l^{\lambda_3}}. \tag{4}$$

Finally, for the intercept, the prior is uninformative:

$$\mathrm{sd}_{int} = 1000\sigma_{GDP}. \tag{5}$$

In setting these components of the prior, for σ_{GDP} and $\sigma_{i,j}$ we use standard deviations from AR(4) models for GDP growth and $x_{i,j,t}$ estimated with the available sample of data.

In all of our results, the hyperparameters are set at values that may be considered very common in Minnesota-type priors (see, e.g., Litterman (1986)): $\lambda_1 = 0.2$, $\lambda_2 = 0.2$, and $\lambda_3 = 1$. We have run some limited checks (for a few models) to see what hyperparameter settings would be optimal in a real-time RMSE-minimizing sense. To simplify the optimization, we focused on just λ_2. In effect, the parameter λ_1 can be seen as pinning down the rate of shrinkage for the lags of GDP growth in the model, while, given λ_1, λ_2 pins down the rate of shrinkage on the coefficients of the monthly indicators. Specifically, after simply fixing λ_1 at a conventional value of 0.2, we specified a wide grid of values for λ_2, and generated time series of forecasts for each corresponding model estimate (for a limited set of models). We then looked at choosing λ_2 in pseudo-real time to minimize the RMSE of past forecasts, using 5- or 10-year windows. Using the model with both coincident and leading indicators and both current quarter and past quarter values of the indicators in the model, at the first evaluation point, in late 1989, the optimal λ_2 was 0.2. As forecasting moves forward in time, the optimal setting drifted up a bit and then down a bit, before ending the sample at values as high as 1. For simplicity, in all of the results in the paper, we leave λ_2 at 0.2 through all of our analysis. It is possible that the more computationally intensive approach of optimizing shrinkage at each forecast origin could improve forecast accuracy, but in a

VAR context, Carriero, Clark, and Marcellino (2011) find the payoff to optimization over fixed, conventional shrinkage to be small.

In the prior for the volatility-related components of the model, our approach is similar to that used in such studies as Clark (2011), Cogley and Sargent (2005), and Primiceri (2005). For the prior on ϕ, we use a mean of 0.035 and 5 degrees of freedom. For the period 0 value of volatility of each equation i, we use a prior of

$$\underline{\mu}_{\lambda} = \log \hat{\lambda}_{0,OLS}, \ \underline{\Omega}_{\lambda} = 4. \tag{6}$$

To obtain $\log \hat{\lambda}_{0,OLS}$, we use a training sample of 40 observations preceding the estimation sample to fit an AR(4) model to GDP growth.

Finally, in the model specifications that include time-varying parameters, rather than a prior mean and variance on the constant coefficient vector, we need a mean and variance for the period 0 value of the coefficient vector and a prior on the variance-covariance matrix of innovations to coefficients. The mean and variance for the period 0 coefficient vector are set to, respectively, the prior mean and four times the variance used in the constant parameter case (described above). For the prior on $Q_m = \text{var}(n_{m,t})$, we follow Cogley and Sargent (2005), among others, in using an inverse Wishart distribution. We set the degrees of freedom at $\dim(X_{m,t}) + 1$ and the scale matrix equal to (6.25×10^{-5}) times the prior variance matrix used in the constant parameter case.[7]

3.7 Estimation algorithms

The model with constant volatility is estimated with a Gibbs sampler, using the approach for the Normal-diffuse prior and posterior detailed in such studies as Kadiyala and Karlsson (1997). At any given forecast origin, estimation is quite fast, because the forecasting model is a single equation.[8]

The model with stochastic volatility is estimated with a Metropolis-within-Gibbs algorithm, used in such studies as Clark (2011) and Carriero, Clark, and Marcellino (2012). The

[7] We obtained very similar results with a somewhat tighter prior.

[8] Modifying the prior to make it normal-inverse gamma would permit the use of analytical formula for the posterior mean and variance of the coefficient vector.

posterior mean and variance of the coefficient vector are given by

$$\bar{\mu}_\beta = \bar{\Omega}_\beta \left\{ \sum_{t=1}^{T} \lambda_t^{-1} X_{m,t} y_t + \underline{\Omega}_\beta^{-1} \underline{\mu}_\beta \right\} \tag{7}$$

$$\bar{\Omega}_\beta^{-1} = \underline{\Omega}_\beta^{-1} + \sum_{t=1}^{T} \lambda_t^{-1} X_{m,t} X'_{m,t}, \tag{8}$$

where we again omit the m index from the parameters for notational simplicity.

For the models with TVP, we replace the step of the Gibbs sampler that draws the coefficients from a conditional posterior that is normal with a step that uses the Kalman filter and the backward smoothing and simulation algorithm of Durbin and Koopman (2002) to draw time series of the vector of coefficients. Our approach to drawing the time series of coefficients $\beta_{m,t}$ is the same as that described in sources such as Cogley and Sargent (2005), except that we use the backward smoother of Durbin and Koopman (2002) instead of the Carter and Kohn (1994) smoother. For these models, we also add a Gibbs sampler step to draw Q_m from an inverse Wishart distribution, as described in Cogley and Sargent (2005).

In all cases, we estimate the forecasting models with a recursive scheme: the estimation sample expands as forecasting moves forward in time. For a subset of our models, we have also examined forecast accuracy obtained with a rolling estimation scheme. In general, these results showed a rolling approach to be dominated by the recursive approach used in the paper. Given a model specification (in terms of variables and volatility), in most cases using a recursive scheme yielded point and density forecasts more accurate than those obtained using a rolling scheme. However, in the sample ending in 2008, for models with constant volatility, using a rolling scheme for estimation tended to yield average scores modestly higher than those obtained using a recursive scheme for estimation. Still, using stochastic volatility yielded much larger gains in average scores.[9]

In all cases, we obtain forecast distributions by sampling as appropriate from the posterior distribution. For example, in the case of the BMF-SV model and for a given horizon, for each set of draws of parameters, we: (1) simulate volatility for the quarter being forecast using the random walk structure of log volatility; (2) draw shocks to the variable with

[9] For the accuracy of density forecasts, one key to success is capturing in a timely way the fall in volatility associated with the Great Moderation and the rise in volatility associated with the most recent severe recession. For models treating volatility as constant, the rolling sample approach adapts more to the Great Moderation than does a recursive sample approach. However, the rolling sample approach doesn't adapt to the Great Moderation or the crisis nearly as well as does stochastic volatility.

variance equal to the draw of volatility; and (3) use the structure of the model to obtain a draw of the future value (i.e., forecast) of the variable. We then form point forecasts as means of the draws of simulated forecasts and density forecasts from the simulated distribution of forecasts. Conditional on the model, the posterior distribution reflects all sources of uncertainty (latent states, parameters, and shocks over forecast interval).

To conclude, note that, throughout the analysis, we will focus on current-quarter forecasts (corresponding to 1-step ahead forecasts for most of our models). Our method can easily be extended to longer forecast horizons, and we have generated results for horizons of 2 and 4 quarters ahead, but we found very little evidence of predictability at these longer horizons, in line with the nowcasting literature.

4 Competing Nowcasts

We now detail some alternative nowcasts to which we compare those resulting from our BMF and BMF-SV models. We include nowcasts generated by simple AR models, MIDAS-based models, and surveys. AR models are typically tough benchmarks in forecast competitions. MIDAS models are specifically designed to handle mixed frequency data. And survey-based nowcasts pool many predictions, each based on timely information.

4.1 AR models

In our forecast evaluation, in light of evidence in other studies of the difficulty of beating simple AR models for GDP growth, we include forecasts from AR(2) models. The models take the same basic forms given in (1) and (2), with $X_{m,t}$ defined to include just a constant and two lags of GDP growth. In keeping with our real-time setup, we generate four different AR-based forecasts of GDP growth in each quarter t, based on the data available in real time as of the end of the first week of month 1 of quarter t, at the end of the first week of month 2 of quarter t, at the end of the first week of month 3, and at the end of the first week of month 1 of quarter $t+1$. The models based on month 2, month 3, and month 1 of quarter $t+1$ are all conventional AR(2) specifications relating GDP in quarter t to GDP in quarters $t-1$ and $t-2$. For a given quarter, these model estimates and forecasts differ only in that the GDP data available for estimation and forecasting will differ across the months/data vintages. However, the specification of the model based on month 1 of quarter t differs, because GDP growth for period $t-1$ is not yet available. In this case, the model takes a direct multi-step form relating GDP in quarter t to GDP in quarters $t-2$ and $t-3$, and the forecast horizon is in effect 2 quarters, not 1 quarter. In all cases, in light of prior evidence of the success of AR models estimated by least squares, we estimate the AR models with extremely loose priors, so that our Bayesian estimates based on the normal-diffuse prior effectively correspond to least squares estimates.

4.2 MIDAS

4.2.1 The basic MIDAS model

MIDAS models rely on current and lagged high frequency (monthly in our case) indicators to forecast current and future quarterly GDP growth. To describe MIDAS models, we need to introduce some additional notation. Specifically, we denote as before GDP growth as y_t, where $t = 1, 2, 3, ..., T^y$ is a quarterly time index and T^y is the final quarter for which GDP is available. GDP growth can be expressed also as a monthly variable with missing values, so that GDP growth is observable only in $t_m = 3, 6, 9, ..., T_m^y$ where t_m is the monthly time index and $T_m^y = 3T^y$. Therefore, what we want to obtain is the nowcast or forecast of the economic activity h quarters ahead or, equivalently, $h_m = 3h$ months ahead. The monthly stationary indicator is indicated by x_{t_m}, with $t_m = 1, 2, 3, ..., T_m^x$, where T_m^x is the final month for which the indicator is available. Usually monthly indicators are available earlier during the quarter than the GDP release, so generally we condition the forecast on the information available up to month T_m^x, which includes GDP information up to T^y and indicator observations up to T_m^x with $T_m^x \geq T_m^y = 3T^y$.

The MIDAS approach is based on direct forecasting, which requires the specification of different models for different forecasting horizons. The forecast model for horizon $h = h_m/3$ is:

$$y_{t+h} = y_{t_m+h_m} = \beta_0 + \beta_1 b\left(L_m, \theta\right) x_{t_m+w}^{(3)} + \varepsilon_{t_m+h_m}, \tag{9}$$

where y_{t_m} and x_{t_m} are, respectively, GDP growth and the monthly indicator, $x_{t_m}^{(3)}$ is the corresponding skip-sampled monthly indicator, $w = T_m^x - T_m^y$, and $b\left(L_m, \theta\right)$ is the exponential Almon lag:

$$b\left(L_m, \theta\right) = \sum_{k=0}^{K} c\left(k, \theta\right) L_m^k, \quad c\left(k, \theta\right) = \frac{\exp\left(\theta_1 k + \theta_2 k^2\right)}{\sum\limits_{k=0}^{K} \exp\left(\theta_1 k + \theta_2 k^2\right)}. \tag{10}$$

Note that the exponential Almon lag permits the introduction of substantial dynamics into the model without parameter proliferation. In our BMF models parameter proliferation is instead controlled at the estimation stage, by means of shrinkage.

We estimate the MIDAS model using nonlinear least squares (NLS) in a regression of y_{t_m} on $x^{(3)}_{t_m-k}$, yielding coefficients $\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\beta}_0$ and $\widehat{\beta}_1$. Since the model is $h-$dependent, we reestimate it for multi-step forecasts and when new information becomes available.

The forecast is given by:

$$\widehat{y}_{T^y_m+h_m|T^x} = \widehat{\beta}_0 + \widehat{\beta}_1 B\left(L^{1/m}; \widehat{\theta}\right) x^{(3)}_{T^x_m}. \tag{11}$$

As far as the specification is concerned, we use a large variety of initial parameter specifications, compute the residual sum of squares (RSS) from equation (9) and choose the parameter set which gives the smallest residual sum of squares as initial values for the NLS estimation. K in the exponential Almon lag function is fixed at 12, and the parameters are restricted to $\theta_1 < 5$ and $\theta_2 < 0$.

4.2.2 The U-MIDAS model

The adoption of the exponential Almon lag polynomial $b\left(L_m, \theta\right)$ permits the use of many lags of the high frequency indicator without increasing too much the number of parameters. However, it also constrains the shape of the dynamic response of the low frequency target variable, GDP growth for us, to the high frequency indicators. Therefore, when the frequency mismatch is small, as in the case of monthly and quarterly variables, it can be preferable to use an unrestricted polynomial. The resulting model is called unrestricted MIDAS, or U-MIDAS, by Foroni, et al. (2012). Another advantage of U-MIDAS is that it can be estimated by simple OLS.

4.2.3 Introducing an AR term

A natural extension of the basic MIDAS model is the introduction of an autoregressive term. Including the AR dynamics is desirable but not straightforward. Ghysels et al. (2004) show that the introduction of lagged dependent variables creates efficiency losses. Moreover, it would result in the creation of seasonal patterns in the explanatory variables. Therefore, we follow Clements and Galvao (2008) and introduce the AR dynamics as a common factor to rule out seasonal patterns. We estimate the AR-MIDAS, defined as:

$$y_{t_m+h_m} = \beta_0 + \lambda y_{t_m} + \beta_1 b\left(L_m, \theta\right)\left(1 - \lambda L_m^{h_m}\right) x_{t_m+w}^{(3)} + \varepsilon_{t_m+h_m}, \tag{12}$$

where the λ coefficient can be estimated together with the other coefficients by NLS. Even in this case, we follow the procedure described for the MIDAS approach: first compute the RSS from (12), choose the parameters that minimize it, and use them as initial values for the NLS estimation.

An AR term can be also easily included in the U-MIDAS specification, where no additional complications arise and the model can be still estimated by OLS.

Since empirically we find that the use of an AR term improves the forecasting performance of both MIDAS and U-MIDAS, we will always include it.

4.2.4 Dealing with multiple indicators

The extent of the nonlinearity of the MIDAS model increases substantially when more than one indicator is included in the model. This is not an issue for the U-MIDAS specification, but in this case the problem arises because of the increase in the number of parameters. Hence, we will compute results for single indicator MIDAS and UMIDAS specifications, but only report results for pooled forecasts constructed as simple averages of the forecasts based on the coincident, leading, and financial groups of indicators, as specified above in section 4.1. This approach worked well according to Kuzin, Marcellino, and Schumacher (2012). Detailed results for each indicator are available upon request.

4.3 Surveys

We also consider GDP growth nowcasts based on the Survey of Professional Forecasters (SPF), available quarterly, and Blue Chip Consensus, available on a monthly basis, since they are closely monitored by decision makers and typically perform quite well.

We should note that the forecasts from the nowcasting models, Blue Chip, and the SPF reflect information sets that, in terms of timing, should be similar. In particular, the Blue Chip (BC) survey is conducted a few days before publication on the 10th of each month. So it should usually be the case that Blue Chip respondents have available the same information each nowcasting model uses. For example, for month 2 of quarter t, we define the model to use information normally available at the end of the first week of the month, which will include employment and the ISM for month 1 of the quarter. At the time of the Blue Chip survey, that same information would normally be available to participating forecasters. In the case of the SPF forecast, the mid-quarter timing of the survey means that the SPF forecast should only be comparable to the Blue Chip and model forecasts made in month 2 of the quarter (while most comparable, the SPF forecast should normally reflect a little more information than would be available to Blue Chip or the models).

5 Data

We focus on current-quarter forecasting of real GDP (or GNP for some of the sample) in real time. Quarterly real-time data on GDP or GNP are taken from the Federal Reserve Bank of Philadelphia's Real-Time Data Set for Macroeconomists (RTDSM). For simplicity, hereafter "GDP" refers to the output series, even though the measures are based on GNP and a fixed weight deflator for much of the sample.

As indicated in section 3, to forecast GDP, we use a range of monthly variables: payroll employment, industrial production, real retail sales (nominal deflated by the CPI), housing starts, the ISM index (overall) for manufacturing, the ISM index for supplier delivery times, the ISM index for orders, average weekly hours of production and supervisory workers, new claims for unemployment insurance, stock prices as measured by the S&P 500 index, the 10-year Treasury bond yield, and the 3-month Treasury bill rate.

For the variables subject to significant revisions — payroll employment, industrial production, retail sales, and housing starts — we use real-time data, obtained from the RTDSM (employment, industrial production, and housing starts) or the Federal Reserve Bank of St. Louis' ALFRED database (retail sales). For the CPI, we use the 1967-base year CPI available from the BLS rather than a real-time series; Kozicki and Hoffman (2004) show that the 1967 base year series is very similar to real-time CPI inflation. For the other variables, subject to either small revisions or no revision, we simply use the currently available time series, obtained from the Federal Reserve Board's FAME database.

The full forecast evaluation period runs from 1985:Q1 through 2011:Q3 (using period t to refer to a forecast for period t), which involves real-time data vintages from January 1985 through March 2012. For each forecast origin t starting in the first month of 1985:Q1, we use the real-time data vintage t to estimate the forecast models and construct forecasts of GDP growth in the quarter.[10] The starting point of the model estimation sample is always 1970:Q2, the soonest possible given data availability and lags allowed in models.

In light of the potential for the large surprises of the recent sharp recession to alter results, we also report results for a sample ending in 2008:Q2, before the recession became dramatic.

As discussed in such sources as Croushore (2006), Romer and Romer (2000), and Sims (2002), evaluating the accuracy of real-time forecasts requires a difficult decision on what to take as the actual data in calculating forecast errors. The GDP data available today for, say, 1985, represent the best available estimates of output in 1985. However, output as defined and measured today is quite different from output as defined and measured in 1970. For example, today we have available chain-weighted GDP; in the 1980s, output was measured with fixed-weight GNP. Forecasters in 1985 could not have foreseen such changes and the potential impact on measured output. Accordingly, we follow studies such as Clark (2011), Faust and Wright (2009), and Romer and Romer (2000) and use the second available estimates in the quarterly vintages of the RTDSM of GDP/GNP as actuals in evaluating forecast accuracy.[11]

6 Results

This section presents results on the accuracy of point and density forecasts from our proposed BMF and BMF-SV methods relative to the accuracy of forecasts from AR models, MIDAS specifications, the Survey of Professional Forecasters (SPF), and Blue Chip. For the MIDAS, SPF, and Blue Chip forecasts, our comparisons are limited to point forecasts. The section first describes the metrics used and then provides the results. As noted in section 3, we present results for both a full sample of 1985:Q1-2011:Q3 and a pre-crisis sample of 1985:Q1-2008:Q2. The results for the TVP specifications are summarized in the final subsection, in the interest of brevity and since, as we will see, SV is much more relevant than TVP, at

[10] In forming the dataset used to estimate the forecasting models at each point in time, we use the monthly vintages of (quarterly) GDP available from the RTDSM, taking care to make sure the GDP time series used in the regression is the one available at the time the forecast is being formed.

[11] We have also computed results using the first estimate of GDP and qualitatively there are no major changes.

least in our application. Results on overall model fit are provided in the appendix.

6.1 Metrics

To assess the accuracy of point forecasts, we use RMSEs. To facilitate presentation, we report RMSEs for each nowcasting model, Blue Chip, and SPF relative to the AR model with constant volatility. To provide a rough gauge of whether the differences in RMSEs are statistically significantly, we use the Diebold and Mariano (1995)–West (1996) t-statistic for equal MSE, applied to the forecast of each model relative to the benchmark.[12]

For comparing our proposed BMF and BMF-SV forecasts and MIDAS-based forecasts to AR model forecasts, the overlap between each alternative model and the benchmark could in principle complicate inference. Our models of interest do not strictly nest the AR models, because the AR models include 2 lags of GDP growth while the nowcasting models include just 1 lag of GDP growth. But it is possible that the models overlap, in the sense that the true model could be an AR(1) specification. Clark and McCracken (2012) develop forecast tests for potentially overlapping models. Based on forecast performance, it seems unlikely that the AR model and nowcasting models overlap, so we proceed to treat them as being non-nested. The results in West (1996) imply that we can test equal accuracy of point forecasts from non-nested models by computing a simple t-test for equal MSE. Because some of the differences in squared forecast errors have some serial correlation that appears to be of an MA(1) form, we compute the t-statistics with a heteroskedasticity and auto-correlation consistent (HAC) variance, using a rectangular kernel and bandwidth of 1.[13] Computing the t-statistics with the (data-dependent) pre-whitened quadratic spectral HAC estimator of Andrews and Monahan (1992) yields very similar results.

The RMSE, while informative and commonly used for forecast comparisons, is based on the point forecasts only and therefore ignores the rest of the forecast density. The overall calibration of the density forecasts can be measured with log predictive density scores, motivated and described in such sources as Geweke and Amisano (2010). At each forecast origin, we compute the log predictive score using the real-time outcome and the probability

[12]In all cases, we abstract from the corrections to test statistics based on real time forecasts developed in Clark and McCracken (2009), partly for simplicity and partly because the corrections proved not to be very important in the application results of Clark and McCracken (2009).

[13]We also incorporate in the t-statistics the small-sample adjustment of Harvey, Leybourne, and Newbold (1997).

density of the forecast. For all models, we compute the density using an empirical estimate of the forecast density based on 5000 draws of forecasts, a non-parametric density estimator, and a Gaussian kernel. To facilitate model comparisons, we report average log scores for our BMF and BMF-SV models relative to a benchmark AR model with stochastic volatility (AR-SV). To provide a rough gauge of the statistical significance of differences in average log scores, we use the Amisano and Giacomini (2007) t-test of equal means, applied to the log score for each model relative to the AR-SV model. We view the tests as a rough gauge because, for forecasts from estimated models, the asymptotic validity of the Amisano and Giacomini (2007) test requires that, as forecasting moves forward in time, the models be estimated with a rolling, rather than expanding, sample of data. To allow for the potential of some serial correlation in score differences, we compute the t-statistics with a HAC variance estimate obtained with a rectangular kernel and bandwidth of 1 (using the pre-whitened quadratic spectral estimator yields very similar results).

Some researchers have argued that the average cumulative ranked probability score (CRPS) should be preferred to the average log predictive score, for being less sensitive to outliers and for better recognizing forecasts that are reasonably close to, but not exactly at, the outcome. Therefore, we have also computed CRPS results for all models, and make them available upon request. Broadly, these CRPS results are similar to the average log score results we report below, except that we get smaller differences in the pre-crisis sample relative to the full sample.

As further checks on density forecast calibration, we also provide results on the accuracy of interval forecasts and selected results for probability integral transforms (PITs). Motivated in part by central bank interest in forecast confidence intervals and fan charts, recent studies such as Giordani and Villani (2010) have used interval forecasts as a measure of forecast accuracy for macroeconomic density forecasts. We compute results for 70 percent interval forecasts, defined as the frequency with which real-time outcomes for GDP growth fall inside 70 percent highest posterior density intervals estimated in real time for each model. To provide a rough gauge of statistical significance, we include p-values for the null of correct coverage (empirical = nominal rate of 70 percent), based on t-statistics computed with a HAC variance estimate obtained with a rectangular kernel and bandwidth of 1. The p-values provide only a rough gauge of significance in the sense that the theory underlying

Christofferson's (1998) test results abstracts from forecast model estimation — that is, parameter estimation error — while the forecasts for which we provide interval forecast results are obtained from estimated models.

The probability integral transform (PIT) emphasized by Diebold, Tay, and Gunther (1998) provides a more general indicator of the accuracy of density intervals than does an interval forecast coverage rate. For an illustrative set of models, we provide PIT histograms, obtained as decile counts of PIT transforms. For optimal density forecasts at the 1-step horizon, the PIT series would be independent uniform (0,1) random variables. Accordingly, the histograms would be flat. Studies such as Christoffersen and Mazzotta (2005), Clements (2004) and Geweke and Amisano (2010) consider similar measures of density forecasts. To provide some measure of the importance of departures from the iid uniform distribution, we include in the histograms 90% intervals estimated under the binomial distribution (following Diebold, Tay, and Gunther (1998)). These intervals are intended to be only a rough guide to significance of departures from uniformity; more formal testing would require a joint test (for all histogram bins) and addressing the possible effects of model parameter estimation on the large-sample distributions of PITs.

6.2 Point forecasts

To assess the accuracy of point forecasts, Tables 2 and 3 provide RMSE comparisons of our proposed BMF and BMF-SV nowcasting models, pooled MIDAS, Blue Chip, and the SPF to forecasts from the AR model. To facilitate comparisons, the first row of each table provides the RMSE of the AR model forecast; remaining rows provide the ratio of each forecast's RMSE relative to the AR model's RMSE. A number less than 1 means a given forecast is more accurate than the AR model. The numbers in parentheses are the p-values of two-sided t-statistics for equal MSE.

Before discussing the results, we should note that the forecasts from the nowcasting models, Blue Chip, and the SPF reflect information sets that, in terms of timing, should be similar. In particular, the Blue Chip (BC) survey is conducted a few days before publication on the 10th of each month. So it should usually be the case that Blue Chip respondents have available the same information each nowcasting model uses. For example, for month 2 of quarter t, we define the model to use information normally available at the end of the first week of the month, which will include employment and the ISM for month 1 of the quarter. At the time of the Blue Chip survey, that same information would normally be available to participating forecasters. In the case of the SPF forecast, the mid-quarter timing of the survey means that the SPF forecast should only be comparable to the Blue Chip and model forecasts made in month 2 of the quarter (while most comparable, the SPF forecast should normally reflect a little more information than would be available to Blue Chip or the models).

We can draw five main conclusions from the RMSE results in Tables 2 and 3. First, as might be expected, the accuracy of forecasts from the BMF and BMF-SV models improves as more data on the quarter becomes available, and we move from month 1 to 2, 2 to 3, and 3 to month 1 of next quarter. The gains look a little bigger with the move from month 2 to month 3 than from month 3 to month 1 of the next quarter. For example, in the full sample, the RMSE level (not reported in the tables) for the model with coincident indicators (0 lags, denoting current quarter indicators only) and stochastic volatility falls from 1.873 in month 2 of quarter t to 1.668 in month 3 and to 1.521 in month 1 of quarter $t+1$. As a consequence, the accuracy of the nowcasting models relative to the AR baseline increases with the addition of more information on the quarter. In the case of the model with coincident indicators (0 lags, denoting current quarter indicators only) and stochastic volatility, the RMSE gain in the full sample results rises from about 6 percent in month 1 to 18 percent in month 3 and 25 percent in month 1 of the next quarter. The pooled MIDAS and U-MIDAS forecasts based on either the coincident indicators or all the indicators are comparable to the nowcasting models based on our proposed BMF approach. Somewhat surprisingly, the Diebold-Mariano-West test doesn't often imply the gains to be statistically significant in the full sample, but it does imply more significance in the sample that ends before the depths of the crisis.

Second, in the sample that ends in mid-2008 and thereby avoids the huge forecast errors of the severe recession, our BMF and BMF-SV nowcasting models are often as accurate as or even a bit more accurate (although not significantly so) than Blue Chip, particularly in months 2 and 3 of quarter t. For example, the RMSE ratio of the model with coincident indicators (0 lags, denoting current quarter indicators only) and stochastic volatility is 0.955 in month 1, 0.926 in month 2, and 0.852 in month 3, compared to corresponding ratios for

Blue Chip of 0.963, 0.983, and 0.881. However, from month 3 of quarter t to month 1 of quarter $t + 1$, the Blue Chip forecasts improve in accuracy more so than do the model forecasts. As a result, in month 1 of quarter $t + 1$, the nowcasting models are generally less accurate than Blue Chip, although not dramatically so in the case of the better models. Continuing with the same example, the nowcasting model yields an RMSE ratio of 0.799 for month 1 of quarter $t + 1$, compared to the Blue Chip forecast's ratio of 0.737. The pooled MIDAS and U-MIDAS forecasts are generally not as accurate as some of our better models, except for month 1 of quarter t and although the differences in accuracy are small.

To shed some further light on the performance of the nowcasting models and Blue Chip over time, Figure 1 compares actual quarterly GDP growth (annualized) to point forecasts from Blue Chip and our BMF-SV nowcasting model with coincident indicators. The chart makes clear the improvement in accuracy that occurs with the addition of more data on the quarter — improvement that seems most noticeable around recessions (1990-91, 2001, 2007-2009). It also shows that, over some periods of time, the model is more accurate than Blue Chip, while in others, Blue Chip is more accurate than the model. One period in which Blue Chip fares better is the most recent recession, when Blue Chip did a better job of picking up and projecting unprecedented declines in GDP growth.[14]

Accordingly, the third main conclusion from the RMSE results is that, in the full sample, the nowcasting models are somewhat less accurate than Blue Chip, seemingly due in part to relative performance in the depths of the crisis.[15] The challenge of beating a survey forecast with good nowcasting models is also evident in such studies as Banbura, et al. (2012), who develop a mixed frequency factor model-based forecast that is comparable to, but not quite as good, as SPF in forecasts for 1995-2010.

In light of the evidence in Chauvet and Potter (2012) that the advantage of some time series models over an AR model baseline stems largely from periods of recession, not during economic expansions (or normal times), we have checked the forecast performance of our models during just economic expansions (dropping out observations falling during NBER

[14]In light of these patterns, we have tried using the SPF forecast, available for a longer historical time series than our Blue Chip data, as an additional regressor in our model, with a prior that pushes the model forecasts towards the SPF. However, the resulting model did not yield any major forecast gains.

[15]In unreported results, we computed t-tests of equal accuracy of the nowcasting model forecasts against Blue Chip. According to these tests, for forecasts generated in months 1-3 of quarter t, the differences in the accuracy of model forecasts and BC forecasts are modest enough that they are never statistically significant.

recessions). During expansions, our nowcasting models also forecast more accurately than the AR baseline. That said,in terms of RMSEs, the advantages of the nowcasting models over the AR baseline are somewhat smaller when recessions are excluded than in the full sample. (The expansion versus recession distinction is smaller in density forecast accuracy than in point forecast accuracy.) Overall, the advantages of our models over an AR baseline may be less affected by the expansion versus recession distinction than the models of Chauvet and Potter (2012) were affected because our models exploit more within-the-quarter indicators of economic activity.

Returning to the primary take-aways from our results, a fourth conclusion to draw from Tables 2 and 3 is that including stochastic volatility in our proposed BMF nowcasting model doesn't have much payoff, or cost, in terms of the accuracy of point forecasts. Broadly, for a given variable set included in a nowcasting model, BMF and BMF-SV yield similar RMSE ratios, with the SV version sometimes a little better and other times a little worse.

Finally, no single nowcasting model — that is, no single set of predictors — jumps out as best. Either coincident indicators by themselves or coincident indicators in combination with financial indicators seem to work about as well as anything else, in the full sample. By themselves, financial indicators do poorly in forecasting current-quarter GDP growth. However, including financial indicators with coincident indicators helps the models (a little) during the recent crisis.

The main message that we can take from the point forecast evaluation is that overall our BMF method is comparable to other mixed frequency forecasting approaches and also to survey forecasts, though the surveys performed a bit better during the crisis. However, a major advantage of our approach is that it also easily delivers density forecasts, and, as we will now see, in this context the stochastic volatility specification that we adopt becomes quite relevant.

6.3 Density forecasts: average predictive scores

To assess the calibration of density forecasts, Tables 4 and 5 provide average log score comparisons of our BMF and BMF-SV nowcasting models, taking an AR model with stochastic volatility as the benchmark (since previous research has shown stochastic volatility to improve density accuracy of AR forecasts). Unfortunately, while the SPF includes some density forecast information, it is not directly in a form that would easily permit computing log scores for current-quarter forecasts of GDP growth. The Blue Chip survey doesn't provide any forecast density information. And, density forecasts for pooled MIDAS models have not been derived yet. Accordingly, we focus our density evaluation on forecasts from the BMF and BMF-SV models estimated with simulation methods. To facilitate comparisons, the first row of each table provides the average log score of the AR-SV forecast; remaining rows provide the score of each other model forecast less the benchmark score. Entries greater than 0 mean a given density forecast is more accurate (has a higher score) than the AR-SV baseline. The numbers in parentheses are the p-values of two-sided t-statistics for tests of equality of average log scores.

The main findings are as follows. First, including stochastic volatility in a model considerably improves its average log score. This is true for both the AR model and our BMF nowcasting models. Consider, for example, the model with coincident indicators available early in month 2 of quarter t. The constant volatility version of the model yields an average score that is 15.1% below the AR-SV baseline, while the stochastic volatility version yields a score that is 8.5% above the baseline.

To provide some intuition for the importance of allowing time-varying volatility, Figure 2 reports the estimates of stochastic volatility from an AR model and our BMF-SV nowcasting model that includes coincident and financial indicators (other nowcasting models yield similar results), obtained from the full sample of data available in our last real-time data vintage. The volatility plotted is $\lambda_{m,t}^{0.5}$ from equation (2), $m = 1, 2, 3$, corresponding to the standard deviation of shocks to GDP growth in each model. For the AR-SV model, we report just the posterior median of $\lambda_{m,t}^{0.5}$; for the BMF-SV model, we report the posterior median and the 70 percent credible set. The charts show that time-variation in volatility is considerable for an AR model, reflecting the Great Moderation and a rise in volatility during the recent recession. Including within-quarter monthly indicators tends to dampens the swings in volatility, more so as more months of data within the quarter become available. However, even with the BMF-SV nowcasting model, there continue to be sizable movements in volatility.

The second main finding is that the average log scores of the BMF models improve as more data becomes available for the quarter (i.e., scores are higher for models with 2 months of data than 1 month of data, etc.). As a consequence, some of the nowcasting models with 2 or 3 months of data on the quarter but constant volatility score better than the AR-SV model. However, in the full sample, in only a couple of cases are these gains statistically significant. Moreover, in the pre-crisis sample, nowcasting models with constant volatilities have a harder time beating the AR-SV benchmark. In fact, in this shorter sample, none of these nowcasting specifications are significantly better than the benchmark. However, most of the BMF-SV models continue to beat the AR-SV benchmark.

Third, almost all of the BMF-SV models improve upon the average log score of the baseline AR-SV specification. The gains increase as the nowcasting models get more months of data. In most cases, the gains are statistically significantly, even in the case of month 1 of the quarter.

Finally, as in the case of point forecasts, it is hard to identify a best model. The BMF-SV model with just coincident indicators and 1 lag of indicators (as well as current quarter monthly indicators) might be considered best, or at least as good as any other. This specification also performs well in the in-sample analysis and in point forecasting. Again, by themselves, financial indicators are not very helpful for forecasting. They can help in conjunction with coincident indicators.

6.4 Interval forecasts

As another measure of density forecast accuracy, we consider interval forecasts. For all of our econometric models, Tables 6 and 7 provide coverage rates defined as the frequency with which actual GDP growth falls within 70 percent forecast intervals, along with p-values for the test that empirical coverage equals the 70 percent nominal rate. A number greater (less) than 70 percent means that a given model yields posterior density intervals that are, on average, too wide (too narrow).

The coverage rates in the tables are striking. Models with constant volatilities in all cases yield coverage rates of about 90 percent, which are in all cases significantly different from the nominal rate of 70 percent. Somewhat surprisingly, given the patterns in the score results, coverage doesn't show much tendency to get better with the addition of more data (it does get a little better, but not much). This suggests the improvement in predictive scores that occurs with the addition of months of data is due to improvement in the forecast mean. Regardless, models with stochastic volatility in almost all cases yield coverage rates close enough to 70 percent that they are not statistically different from 70 percent.

To further highlight the importance of SV, in Figures 3 and 4 we report the real-time 70% interval forecasts from a BMF model based on coincident indicators, without (Figure 3) and with SV (Figure 4). Figure 3 confirms that coverage is pretty terrible for models with constant volatilities. As of month 1 of the quarter, for a model with constant volatility, the 70 percent bands are so wide that actual outcomes hardly ever fall outside the bands. With more months of data, the bands narrow some, but it remains the case that actual outcomes rarely fall outside the bands. As Figure 4 indicates, the same indicator model with stochastic volatility yields much narrower bands, and therefore more outcomes that fall outside the 70 percent bands.

6.5 Probability integral transforms (PITs)

As noted above, PITs can be seen as a generalization of coverage rates (across different rates). In the interest of brevity, we provide in Figures 5 and 6 PITs histograms for just the BMF and BMF-SV models that include coincident indicators, but other models (including AR models) would yield a similar conclusion about the role of stochastic volatility. If the forecasting models were properly specified, the PITs would be uniformly distributed, yielding a completely flat histogram.

The PITs histograms yield results in line with the simple coverage comparison of the previous subsection. As Figure 5 indicates, for models with constant volatilities, the PITs have a distinct tent-type shape, which is consistent with forecast distributions that are too dispersed. In the case of the BMF model with coincident indicators and constant volatility, adding more data doesn't seem to materially improve the shape of PITs. This finding provides further evidence that, in the case of models with constant volatilities, the improvement in predictive scores that occurs with the addition of months of data is due to improvement in the forecast mean, not the shape of the distribution. Figure 6 shows that including stochastic volatility in the nowcasting model yields much flatter PITs histograms. So by the PITs measure, too, including stochastic volatility materially improves the calibration of density forecasts.

6.6 Results for models with TVP

In the interest of brevity, we present just a basic set of results for versions of our models allowing TVP. Specifically, for the 1985-2011 sample, Tables 8 and 9 provide RMSE and average log score results, for BMF-TVP and BMV-TVP-SV specifications that include just coincident indicators, just financial indicators, and just leading indicators. To facilitate the evaluation of the effects of introducing TVP, we include in the table the results for the corresponding models with constant parameters (repeating results provided in Tables 2 and 3). As in previous tables, we report ratios of RMSEs relative to the AR model benchmark and scores relative to the score of the AR-SV benchmark.

At least in our application to nowcasting U.S. GDP growth, adding TVP to our BMF and BMF-SV specifications doesn't offer any consistent gains to the accuracy of point or density forecasts. In most, although not all, cases, the RMSE ratio of each model with TVP are a little higher than the RMSE ratio of the corresponding model with constant coefficients. Consider, for example, the model with coincident indicators used to forecast in month 3 of the quarter. With constant coefficients, the model yields a RMSE ratio of 0.823, compared to a ratio of 0.874 for the same model with TVP. Similarly, in most (but not all) cases, the score differentials of the models with TVP are lower than the score differentials of the corresponding constant-coefficient models. Of course, we can't rule out the possibility that TVP might be more helpful in other applications, and one of the advantages of our proposed modeling framework is that it readily permits the inclusion of TVP.

7 Conclusions

We have developed a Bayesian Mixed Frequency method for producing current-quarter forecasts of GDP growth with a (possibly large) range of available within-the-quarter monthly observations of economic indicators, such as employment and industrial production, and financial indicators, such as stock prices and interest rates.

In light of existing evidence of time variation in the variances of shocks to GDP, we also consider versions of the model with stochastic volatility, while most of the existing approaches assumed that the variance is constant. Similarly, we introduce models with time-varying regression coefficients (with or without stochastic volatility), while the latter are treated as constant in most of the literature on mixed frequency models.

We use Bayesian methods to estimate the model, in order to facilitate providing shrinkage on the (possibly large set of) model estimates and conveniently generate predictive densities. Most prior nowcasting research has focused on the accuracy of point forecasts of GDP growth. Instead, we consider both point and density forecasts.

Empirically, we provide results on the accuracy of nowcasts of real-time GDP growth in the U.S. from 1985 through 2011. In terms of point forecasts, our proposal is comparable to alternative econometric methods and survey forecasts, and yields further evidence on the usefulness of intra-quarter information. Moreover, our approach provides reliable density and interval forecasts, for which the stochastic volatility specification is quite useful. Instead, parameter time variation yields no additional gains, at least in our application.

Our proposed approach could be extended in several directions, such as using higher frequency information. It could be also applied to nowcast other relevant economic variables, such as components of GDP, the inflation rate, or fiscal indicators. We leave these interesting extensions for future research.

8 Appendix

To assess more generally how the competing models fit the full sample of data, we follow studies such as Geweke and Amisano (2010) in using 1-step ahead predictive likelihoods. The predictive likelihood is closely related to the marginal likelihood: the marginal likelihood can be expressed as the product of a sequence of 1-step ahead predictive likelihoods. In our model setting, the predictive likelihood has the advantage of being simple to compute. For model M_i, the log predictive likelihood is defined as

$$\log \mathrm{PL}(\mathrm{M}_i) = \sum_{t=t_0}^{T} \log p(y_t^o | y^{(t-1)}, M_i), \tag{13}$$

where y_t^o denotes the observed outcome for the data vector y in period t, $y^{(t-1)}$ denotes the history of data up to period $t-1$ (note that the model is estimated and the forecast generated using only data up through $t-1$), and $p(\cdot)$ denotes the predictive density.[16] In computing the predictive density, we use an empirical estimate of the forecast density based on 5000 draws of forecasts, a non-parametric density estimator, and a Gaussian kernel. In computing the log predictive likelihood, we sum the log values over the period 1985:Q1 through 2011:Q3 (or through 2008:Q2), computed with our real-time data vintages.

Appendix Tables 1 and 2 provide the sums of log predictive likelihoods for each model, for the full sample and the pre-crisis sample, respectively. The first row gives the fit of the AR model with stochastic volatility, taken as the baseline for model fit (as noted above, the model fit for the AR models differs by month due to differences in data vintages and, for the month 1 case, data availability). The remaining rows give the difference between the likelihood of the indicated BMF or BMF-SV model and the baseline likelihood. A positive number means the indicated model fits better than the baseline. In log terms, a difference in likelihood of a few points corresponds to a large difference in model probabilities.

The following comments are worth making. First, stochastic volatility significantly improves the fit of an AR model (by more than 26 log points). Second, stochastic volatility also significantly improves the fit of all the BMF-SV nowcasting models relative to constant volatility BMF counterparts. However, the gains to stochastic volatility are somewhat more modest with the nowcasting models than the AR models. Third, in month 1 of the quarter,

[16] For the AR forecasting models based on the data available in the first month of the quarter, for which in quarter t only data through $t-2$ are available, the forecast horizon is actually 2 quarters.

the nowcasting models have modest advantages over an AR model. The BMF nowcasting models fit the data modestly better (by about 6-7 log points) than the AR model with constant volatility. The BMF-SV nowcasting models fit the data somewhat better (4 to 14 log points) than the AR model with SV. Fourth, the fit of each nowcasting model consistently improves as more months of data become available, more so in moving from month 2 to 3 and from month 3 to month 1 of the next quarter than in moving from month 1 to month 2. Finally, the pre-crisis sample yields results similar to those for the full sample, with the difference that advantages of the nowcasting models over the AR models are modestly larger in the pre-crisis sample than the full sample. The best BMF-SV models are also similar across samples and typically include the coincident indicators, possibly combined with the financial indicators.

References

[1] Aastveit, K., Gerdrup, K., and Jore, A.S. (2011), Nowcasting GDP in Real-Time: A Density Combination Approach, mimeo, Norges Bank.

[2] Amisano, G., and Giacomini, R. (2007), Comparing Density Forecasts via Weighted Likelihood Ratio Tests, *Journal of Business and Economic Statistics* 25, 177-190.

[3] Andrews, D.W.K., and Monahan, J.C. (1992), An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator, *Econometrica* 60, 953-966.

[4] Baffigi, A., Golinelli, R., and Parigi, G. (2004), Bridge Models to Forecast the Euro Area GDP, *International Journal of Forecasting* 20, 447-460.

[5] Banbura, M., Giannone, D., and Reichlin, L. (2010), Large Bayesian Vector Autoregressions, *Journal of Applied Econometrics* 25, 71-92.

[6] Banbura, M., Giannone, D., and Reichlin, L. (2011), Nowcasting, in *Oxford Handbook on Economic Forecasting*, Clements, M.P., and Hendry D.F. (eds). Oxford University Press: Oxford.

[7] Banbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2012), Nowcasting and the Real Time Data Flow, in *Handbook of Economic Forecasting*, Elliott, G., and Timmermann, A. (eds). Forthcoming.

[8] Bencivelli, L., Marcellino, M., and Moretti, G.L. (2012), Selecting predictors by Bayesian model Averaging in Bridge Models, manuscript, Banca d'Italia.

[9] Camacho, M., Perez-Quiros, G. (2010), Introducing the EURO-STING: Short Term INdicator of Euro Area Growth, *Journal of Applied Econometrics* 25, 663-694.

[10] Carriero, A., Kapetanios, G., and Marcellino M. (2011), Forecasting Large Datasets with Bayesian Reduced Rank Multivariate Models, *Journal of Applied Econometrics* 26, 735-761.

[11] Carriero, A., Clark, T., and Marcellino, M. (2011), Bayesian VARs: Specification choices and forecasting performance, CEPR WP 8273.

[12] Carriero, A., Clark, T., and Marcellino, M. (2012), Common Drifting Volatility in Large Bayesian VARs, CEPR WP 8894.

[13] Carter, C.K., and Kohn, R. (1994), On Gibbs Sampling for State Space Models, *Biometrika* 81, 541-553.

[14] Chauvet, M., and Potter, S. (2012), Forecasting Output, in *Handbook of Economic Forecasting*, Elliott, G., and Timmermann, A. (eds). Forthcoming.

[15] Chen, W., Anderson, B.D., Deistler, M., and Filler, A. (2012), Properties of Blocked Linear Systems, *Automatica*, 48, 2520-2525.

[16] Chiu, C.W., Eraker, B., Foerster, A.T., Kim, T.B., and Seoane, H.D. (2011), Estimating VARs Sampled at Mixed or Irregular Spaced Frequencies: A Bayesian Approach, Federal Reserve Bank of Kansas City Research Working Paper 11-11.

[17] Christoffersen, P.F. (1998), Evaluating Interval Forecasts, *International Economic Review* 39, 841-862.

[18] Christoffersen, P.F., and Mazzotta, S (2005), The Accuracy of Density Forecasts from Foreign Exchange Options, *Journal of Financial Econometrics* 3, 578-605.

[19] Clark, T.E. (2011), Real-Time Density Forecasts from BVARs with Stochastic Volatility, *Journal of Business and Economic Statistics* 29, 327-341.

[20] Clark, T.E., and McCracken, M.W. (2009), Tests of Equal Predictive Ability With Real-Time Data, *Journal of Business and Economic Statistics* 27, 441-454.

[21] Clark, T.E., and McCracken, M.W. (2012), Tests of Equal Forecast Accuracy for Overlapping Models, manuscript, Federal Reserve Bank of Cleveland.

[22] Clements, M.P. (2004), Evaluating the Bank of England Density Forecasts of Inflation, *Economic Journal* 114, 844-866.

[23] Clements, M.P., and Galvao, A.B. (2008), Macroeconomic Forecasting with Mixed-Frequency Data: Forecasting US Output Growth, *Journal of Business and Economic Statistics* 26, 546-554.

[24] Cogley, T., and Sargent, T. (2005), Drifts and Volatilities: Monetary Policies and Outcomes in the post-WWII US, *Review of Economic Dynamics* 8, 262-302.

[25] Croushore, D. (2006), Forecasting with Real-Time Macroeconomic Data, in *Handbook of Economic Forecasting*, Elliott, G., Granger, C., and Timmermann, A. (eds). North Holland: Amsterdam.

[26] Croushore, D., and Stark, T. (2001), A Real-Time Data Set for Macroeconomists, *Journal of Econometrics* 105, 111-30.

[27] De Mol, C., Giannone, D., and Reichlin, L. (2008), Forecasting Using a Large Number of Predictors: Is Bayesian Shrinkage a Valid Alternative to Principal Components? *Journal of Econometrics* 146, 318-328.

[28] D'Agostino, A., Gambetti, L. and Giannone, D. (2012), Macroeconomic Forecasting and Structural Change, *Journal of Applied Econometrics*, forthcoming.

[29] Diebold, F.X., and Mariano, R. (1995), Comparing Predictive Accuracy, *Journal of Business and Economic Statistics* 13, 253-263.

[30] Diebold, F.X., Gunther, T.A., and Tay, A.S. (1998), Evaluating Density Forecasts with Applications to Financial Risk Management, *International Economic Review* 39, 863-883.

[31] Diron, M. (2008), Short-Term Forecasts Of Euro Area Real GDP Growth: An Assessment Of Real-Time Performance Based On Vintage Data, *Journal of Forecasting* 27, 371-390.

[32] Durbin, J., and Koopman, S.J. (2002), A Simple and Efficient Simulation Smoother for State Space Time Series Analysis, *Biometrika* 89, 603-615.

[33] Faust, J., and Wright, J. (2009), Comparing Greenbook and Reduced Form Forecasts using a Large Realtime Dataset, *Journal of Business and Economic Statistics* 27, 468-479.

[34] Faust, J., and Wright, J. (2012), Inflation Forecasting, in *Handbook of Economic Forecasting*, Elliott, G., and Timmermann, A. (eds). Forthcoming.

[35] Foroni, C., Ghysels, E., and Marcellino, M. (2012), Mixed-Frequency Vector Autoregressive Models, mimeo, EUI.

[36] Foroni, C., Marcellino, M., and Schumacher, C. (2012), U-MIDAS: MIDAS Regressions with Unrestricted Lag Polynomials, CEPR Discussion Paper 8828.

[37] Foroni, C., and Marcellino, M. (2012), Mixed-Frequency Structural Models: Identification, Estimation, and Policy Analysis, mimeo, EUI.

[38] Frale, C., Marcellino, M., Mazzi, G.L., and Proietti, T. (2011), EUROMIND: A Monthly Indicator of the Euro Area Economic Conditions, *Journal of the Royal Statistical Society*, Series A, 174, 439-470.

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

[40] Ghysels, E., Santa-Clara, P., and Valkanov, R. (2004), The MIDAS Touch: MIxed DAta Sampling Regression Models, mimeo, University of North Carolina.

[41] Ghysels, E. (2012), Macroeconomics and the Reality of Mixed Frequency Data, manuscript, University of North Carolina.

[42] Giannone, D., Reichlin, R., and Small, D. (2008), Nowcasting GDP and Inflation: The Real-Time Informational Content of Macroeconomic Data Releases, *Journal of Monetary Economics* 55, 665-676.

[43] Giordani, P., and Villani, M. (2010), Forecasting Macroeconomic Time Series with Locally Adaptive Signal Extraction, *International Journal of Forecasting* 26, 312-325.

[44] Guerin, P., and Marcellino, M. (2012), Markov Switching MIDAS Models, *Journal of Business and Economic Statistics*, forthcoming.

[45] Jacquier, E., Polson, N., and Rossi, P. (1994), Bayesian Analysis of Stochastic Volatility Models, *Journal of Business and Economic Statistics* 12, 371-418.

[46] Harvey, D., Leybourne, S. and Newbold, P. (1997), Testing the Equality of Prediction Mean Squared Errors, *International Journal of Forecasting* 13, 281-291.

[47] Kadiyala, K., and Karlsson, S. (1997), Numerical Methods for Estimation and Inference in Bayesian VAR-Models, *Journal of Applied Econometrics* 12, 99-132.

[48] Kozicki, S., and Hoffman, B. (2004), Rounding Error: A Distorting Influence on Index Data, *Journal of Money, Credit and Banking* 36, 319-e38.

[49] Kuzin, V., Marcellino, M., and Schumacher, C. (2011), MIDAS vs. Mixed-Frequency VAR: Nowcasting GDP in the Euro Area, *International Journal of Forecasting* 27, 529-542.

[50] Kuzin, V., Marcellino, M., and Schumacher, C. (2012), Pooling Versus Model Selection for Nowcasting GDP with Many Predictors: Empirical Evidence for Six Industrialized Countries, *Journal of Applied Econometrics*, forthcoming.

[51] Litterman, R. (1986), Forecasting with Bayesian Vector Autoregressions: Five Years of Experience, *Journal of Business and Economic Statistics* 4, 25-38.

[52] Mariano, R., and Murasawa, Y. (2003), A New Coincident Index Of Business Cycles Based on Monthly And Quarterly Series, *Journal of Applied Econometrics* 18, 427-443.

[53] Marcellino, M., Stock, J.H., and Watson, M.W. (2006), A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series, *Journal of Econometrics* 135, 499-526.

[54] Marcellino, M., and Schumacher, C. (2010), Factor-MIDAS for Now- and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP, *Oxford Bulletin of Economics and Statistics* 72, 518-550.

[55] Marcellino, M., Porqueddu, M., and Venditti, F. (2012), Short-Term GDP Forecasting with a Mixed Frequency Dynamic Factor Model with Stochastic Volatility, working paper, Banca d'Italia.

[56] McCracken, M.W., and Sekhposyan, T. (2012), Real-Time Forecasting with a Large, Mixed Frequency, Bayesian VAR, manuscript, Federal Reserve Bank of St. Louis.

[57] Primiceri, G. (2005), Time Varying Structural Vector Autoregressions and Monetary Policy, *Review of Economic Studies* 72, 821-852.

[58] Rodriguez, A., and Puggioni, G. (2010), Mixed Frequency Models: Bayesian Approaches to Estimation and Prediction, *International Journal of Forecasting* 26, 293-311.

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

[60] Sims, C. (2002), The Role of Models and Probabilities in the Monetary Policy Process, *Brookings Papers on Economic Activity* 2, 1-40.

[61] Song, D., and Schorfheide, F. (2012), Real-Time Density Forecasting with a Mixed Frequency VAR, manuscript, University of Pennsylvania.

[62] Stock, J.H., and Watson, M.W. (1989), A Probability Model of the Coincident Economic Indicators, in *Leading Economic Indicators*, Lahiri, K., and Moore, G.H. (eds). Cambridge University Press: New York.

[63] Wallis, K. (1986), Forecasting with an Econometric Model: The 'Ragged Edge' Problem, *Journal of Forecasting* 5, 1-13.

[64] West, K.D. (1996), Asymptotic Inference about Predictive Ability, *Econometrica* 64, 1067-84,

[65] Zadrozny, P.A. (1988), Gaussian-Likelihood of Continuous-Time ARMAX Models when Data are Stocks and Flows at Different Frequencies, *Econometric Theory* 4, 108-124.

Table 1. Specifications of BMF models of GDP growth

	predictors in model for:			
model	*month 1* *quarter t*	*month 2* *quarter t*	*month 3* *quarter t*	*month 1* *quarter t+1*
1. AR	two lags of GDP growth estimated up thru t-2	two lags of GDP growth estimated up thru t-1	two lags of GDP growth estimated up thru t-1	two lags of GDP growth estimated up thru t-1
2. coinc.0lags	GDP(t-2) emp(months 1-3 of t-1) ISM(months 1-3 of t-1) IP(months 1-2 of t-1) RS(months 1-2 of t-1) starts(months 1-2 of t-1)	GDP(t-1) emp(month 1 of t) ISM(month 1 of t)	GDP(t-1) emp(months 1-2 of t) ISM(months 1-2 of t) IP(month 1 of t) RS(month 1 of t) starts(month 1 of t)	GDP(t-1) emp(months 1-3 of t) ISM(months 1-3 of t) IP(months 1-2 of t) RS(months 1-2 of t) starts(months 1-2 of t)
3. coinc.1lag	NA	same as in (2), plus: emp(months 1-3 of t-1) ISM(months 1-3 of t-1) IP(months 1-3 of t-1) RS(months 1-3 of t-1) starts(months 1-3 of t-1)	same as in (2), plus: emp(months 1-3 of t-1) ISM(months 1-3 of t-1) IP(months 1-3 of t-1) RS(months 1-3 of t-1) starts(months 1-3 of t-1)	same as in (2), plus: emp(months 1-3 of t-1) ISM(months 1-3 of t-1) IP(months 1-3 of t-1) RS(months 1-3 of t-1) starts(months 1-3 of t-1)
4. fin.0lags	GDP(t-2) stprice(months 1-3 of t-1) tbill(months 1-3 of t-1) tbond(months 1-3 of t-1)	GDP(t-1) stprice(month 1 of t) tbill(month 1 of t) tbond(month 1 of t)	GDP(t-1) stprice(months 1-2 of t) tbill(months 1-2 of t) tbond(months 1-2 of t)	GDP(t-1) stprice(months 1-3 of t) tbill(months 1-3 of t) tbond(months 1-3 of t)
5. fin.1lag	NA	same as in (4), plus: stprice(months 1-3 of t-1) tbill(months 1-3 of t-1) tbond(months 1-3 of t-1)	same as in (4), plus: stprice(months 1-3 of t-1) tbill(months 1-3 of t-1) tbond(months 1-3 of t-1)	same as in (4), plus: stprice(months 1-3 of t-1) tbill(months 1-3 of t-1) tbond(months 1-3 of t-1)
6. lead.0lags	GDP(t-2) supdel(months 1-3 of t-1) orders(months 1-3 of t-1) hours(months 1-3 of t-1) claims(months 1-2 of t-1)	GDP(t-1) supdel(month 1 of t) orders(month 1 of t) hours(month 1 of t)	GDP(t-1) supdel(months 1-2 of t) orders(months 1-2 of t) hours(months 1-2 of t) claims(month 1 of t)	GDP(t-1) supdel(months 1-3 of t) orders(months 1-3 of t) hours(months 1-3 of t) claims(months 1-2 of t)
7. lead.1lag	NA	same as in (6), plus: supdel(months 1-3 of t-1) orders(months 1-3 of t-1) hours(months 1-3 of t-1) claims(months 1-3 of t-1)	same as in (6), plus: supdel(months 1-3 of t-1) orders(months 1-3 of t-1) hours(months 1-3 of t-1) claims(months 1-3 of t-1)	same as in (6), plus: supdel(months 1-3 of t-1) orders(months 1-3 of t-1) hours(months 1-3 of t-1) claims(months 1-3 of t-1)
8. coinc.fin.0lags	predictors of (2) and (4)	predictors of (2) and (4)	predictors of (2) and (4)	predictors of (2) and (4)
9. coinc.fin.1lag	predictors of (3) and (5)	predictors of (3) and (5)	predictors of (3) and (5)	predictors of (3) and (5)
10. coinc.lead.0lags	predictors of (2) and (6)	predictors of (2) and (6)	predictors of (2) and (6)	predictors of (2) and (6)
11. coinc.lead.1lag	predictors of (3) and (7)	predictors of (3) and (7)	predictors of (3) and (7)	predictors of (3) and (7)
12. coinc.lead.fin.0lags	predictors of (2), (4), and (6)	predictors of (2), (4), and (6)	predictors of (2), (4), and (6)	predictors of (2), (4), and (6)

Notes:
1. All models include a constant.
2. Variables are defined as follows: employment (emp); ISM manufacturing index (ISM); industrial productdion (IP); retail sales (RS); housing starts (starts); ISM index of supplier delivery times (supdel); ISM index of new orders (orders); average weekly hours worked (hours); new claims for unemployment insurance (claims); S&P index of stock prices (stprice); 3-month Treasury bill rate (tbill); and 10-year Treasury bond (tbond).
3. The variable transformations are given in section 3.

Table 2. RMSEs relative to AR benchmark, 1985:Q1-2011:Q3
(RMSE for AR, RMSE ratios for all others)
(p-values of equal MSEs in parentheses)

forecast	month 1 qrtr. t	month 2 qrtr. t	month 3 qrtr. t	month 1 qrtr. $t+1$
ARbaseline	2.213	2.066	2.046	2.029
BC	0.831 (0.251)	0.845 (0.305)	0.757 (0.133)	0.622 (0.034)
SPF	NA	0.803 (0.206)	NA	NA
ARbaseline.SV	0.999 (0.922)	1.007 (0.227)	1.007 (0.342)	1.010 (0.191)
coinc.0lags	0.932 (0.167)	0.916 (0.102)	0.823 (0.120)	0.770 (0.091)
coinc.1lag	NA	0.877 (0.169)	0.805 (0.121)	0.743 (0.082)
coinc.fin.0lags	0.929 (0.071)	0.907 (0.093)	0.790 (0.116)	0.762 (0.120)
coinc.fin.1lag	NA	0.838 (0.103)	0.771 (0.118)	0.741 (0.109)
coinc.lead.0lags	0.908 (0.164)	0.867 (0.124)	0.793 (0.154)	0.752 (0.122)
coinc.lead.1lag	NA	0.836 (0.178)	0.784 (0.158)	0.738 (0.114)
coinc.lead.fin.0lags	0.932 (0.210)	0.872 (0.141)	0.800 (0.207)	0.770 (0.175)
fin.0lags	1.030 (0.451)	1.035 (0.326)	1.030 (0.353)	1.023 (0.456)
fin.1lag	NA	1.031 (0.375)	1.027 (0.419)	1.018 (0.579)
lead.0lags	0.959 (0.156)	0.920 (0.102)	0.844 (0.111)	0.805 (0.120)
lead.1lag	NA	0.905 (0.118)	0.837 (0.124)	0.799 (0.132)
coinc.0lags.SV	0.936 (0.248)	0.906 (0.098)	0.815 (0.108)	0.749 (0.071)
coinc.1lag.SV	NA	0.874 (0.196)	0.791 (0.114)	0.716 (0.067)
coinc.fin.0lags.SV	0.900 (0.110)	0.888 (0.116)	0.806 (0.158)	0.758 (0.120)
coinc.fin.1lag.SV	NA	0.844 (0.198)	0.783 (0.168)	0.744 (0.132)
coinc.lead.0lags.SV	0.902 (0.202)	0.868 (0.164)	0.807 (0.179)	0.753 (0.125)
coinc.lead.1lag.SV	NA	0.857 (0.276)	0.793 (0.181)	0.746 (0.131)
coinc.lead.fin.0lags.SV	0.925 (0.350)	0.879 (0.234)	0.816 (0.246)	0.771 (0.173)
fin.0lags.SV	0.996 (0.863)	1.001 (0.946)	1.001 (0.950)	0.987 (0.632)
fin.1lag.SV	NA	0.998 (0.947)	0.990 (0.743)	0.979 (0.571)
lead.0lags.SV	0.949 (0.130)	0.897 (0.083)	0.835 (0.135)	0.809 (0.157)
lead.1lag.SV	NA	0.877 (0.117)	0.833 (0.164)	0.802 (0.171)
MIDAS, coinc.	0.855 (0.139)	0.864 (0.252)	0.783 (0.111)	0.728 (0.069)
MIDAS, lead	0.876 (0.103)	0.878 (0.157)	0.860 (0.154)	0.843 (0.130)
MIDAS, fin.	0.984 (0.682)	0.965 (0.469)	1.028 (0.365)	1.022 (0.412)
MIDAS, all	0.859 (0.053)	0.849 (0.100)	0.815 (0.054)	0.786 (0.050)
U-MIDAS, coinc.	0.883 (0.236)	0.862 (0.252)	0.787 (0.138)	0.723 (0.072)
U-MIDAS, lead	0.912 (0.271)	0.889 (0.221)	0.878 (0.217)	0.866 (0.194)
U-MIDAS, fin.	1.033 (0.324)	0.995 (0.878)	0.981 (0.639)	1.007 (0.831)
U-MIDAS, all	0.894 (0.113)	0.858 (0.113)	0.800 (0.070)	0.773 (0.044)

Notes: See Table 1 and sections 3 and 4 for the definition of the models. The equal forecast accuracy test is described in section 6.1. The reported RMSEs reflect GDP growth defined in annualized percentage terms.

Table 3. RMSEs relative to AR benchmark, 1985:Q1-2008:Q2

(RMSE for AR, RMSE ratios for all others)

(p-values of equal MSEs in parentheses)

forecast	month 1 qrtr. t	month 2 qrtr. t	month 3 qrtr. t	month 1 qrtr. $t+1$
ARbaseline	1.820	1.758	1.745	1.733
BC	0.963 (0.665)	0.983 (0.824)	0.881 (0.123)	0.737 (0.001)
SPF	NA	0.930 (0.331)	NA	NA
ARbaseline.SV	1.002 (0.829)	1.005 (0.498)	1.010 (0.312)	1.010 (0.309)
coinc.0lags	0.960 (0.050)	0.941 (0.015)	0.859 (0.018)	0.816 (0.005)
coinc.1lag	NA	0.923 (0.025)	0.857 (0.021)	0.805 (0.003)
coinc.fin.0lags	0.942 (0.091)	0.918 (0.003)	0.848 (0.012)	0.836 (0.016)
coinc.fin.1lag	NA	0.889 (0.003)	0.847 (0.021)	0.830 (0.015)
coinc.lead.0lags	0.954 (0.102)	0.914 (0.041)	0.869 (0.067)	0.841 (0.034)
coinc.lead.1lag	NA	0.916 (0.099)	0.878 (0.101)	0.839 (0.030)
coinc.lead.fin.0lags	0.979 (0.567)	0.914 (0.024)	0.892 (0.144)	0.864 (0.062)
fin.0lags	1.001 (0.979)	1.022 (0.626)	1.026 (0.517)	1.022 (0.576)
fin.1lag	NA	1.015 (0.740)	1.021 (0.619)	1.013 (0.766)
lead.0lags	0.994 (0.613)	0.953 (0.067)	0.913 (0.031)	0.896 (0.039)
lead.1lag	NA	0.950 (0.069)	0.917 (0.061)	0.903 (0.072)
coinc.0lags.SV	0.955 (0.159)	0.926 (0.045)	0.852 (0.025)	0.799 (0.003)
coinc.1lag.SV	NA	0.922 (0.105)	0.848 (0.028)	0.788 (0.002)
coinc.fin.0lags.SV	0.921 (0.048)	0.914 (0.028)	0.869 (0.065)	0.842 (0.029)
coinc.fin.1lag.SV	NA	0.919 (0.184)	0.884 (0.126)	0.855 (0.057)
coinc.lead.0lags.SV	0.950 (0.257)	0.921 (0.143)	0.888 (0.144)	0.848 (0.046)
coinc.lead.1lag.SV	NA	0.955 (0.504)	0.896 (0.175)	0.856 (0.078)
coinc.lead.fin.0lags.SV	1.000 (0.993)	0.941 (0.292)	0.914 (0.268)	0.870 (0.070)
fin.0lags.SV	0.992 (0.839)	1.016 (0.611)	1.024 (0.364)	1.014 (0.585)
fin.1lag.SV	NA	1.015 (0.686)	1.014 (0.680)	1.008 (0.832)
lead.0lags.SV	0.980 (0.234)	0.933 (0.045)	0.913 (0.097)	0.912 (0.145)
lead.1lag.SV	NA	0.930 (0.122)	0.927 (0.220)	0.921 (0.219)
MIDAS, coinc.	0.902 (0.139)	0.953 (0.252)	0.857 (0.111)	0.806 (0.069)
MIDAS, lead	0.919 (0.103)	0.919 (0.157)	0.915 (0.154)	0.898 (0.130)
MIDAS, fin.	1.009 (0.682)	1.003 (0.469)	1.028 (0.365)	1.044 (0.412)
MIDAS, all	0.889 (0.053)	0.905 (0.100)	0.864 (0.054)	0.844 (0.050)
U-MIDAS, coinc.	0.972 (0.236)	0.969 (0.252)	0.882 (0.138)	0.809 (0.072)
U-MIDAS, lead	0.977 (0.271)	0.940 (0.221)	0.936 (0.217)	0.925 (0.194)
U-MIDAS, fin.	1.044 (0.324)	1.029 (0.878)	1.004 (0.639)	1.021 (0.831)
U-MIDAS, all	0.943 (0.113)	0.921 (0.113)	0.865 (0.070)	0.831 (0.044)

Notes: See Table 1 and sections 3 and 4 for the definition of the models. The equal forecast accuracy test is described in section 6.1. The reported RMSEs reflect GDP growth defined in annualized percentage terms.

Table 4. Average log scores relative to AR-SV benchmark, 1985:Q1-2011:Q3

(Score for AR-SV, differences in scores for all others)

(p-values of equal mean scores in parentheses)

forecast	month 1 qrtr. t	month 2 qrtr. t	month 3 qrtr. t	month 1 qrtr. $t+1$
ARbaseline.SV	-2.210	-2.144	-2.134	-2.123
ARbaseline	-0.245 (0.000)	-0.258 (0.000)	-0.264 (0.000)	-0.269 (0.000)
coinc.0lags	-0.177 (0.015)	-0.151 (0.009)	-0.007 (0.916)	0.045 (0.493)
coinc.1lag	NA	-0.108 (0.115)	0.019 (0.777)	0.081 (0.251)
coinc.fin.0lags	-0.162 (0.012)	-0.133 (0.026)	0.032 (0.670)	0.061 (0.404)
coinc.fin.1lag	NA	-0.062 (0.387)	0.067 (0.393)	0.100 (0.193)
coinc.lead.0lags	-0.162 (0.032)	-0.108 (0.114)	0.071 (0.373)	0.105 (0.181)
coinc.lead.1lag	NA	-0.032 (0.680)	0.104 (0.203)	0.139 (0.080)
coinc.lead.fin.0lags	-0.145 (0.031)	-0.091 (0.186)	0.070 (0.416)	0.094 (0.251)
fin.0lags	-0.229 (0.000)	-0.260 (0.000)	-0.258 (0.000)	-0.266 (0.000)
fin.1lag	NA	-0.252 (0.000)	-0.241 (0.000)	-0.242 (0.000)
lead.0lags	-0.201 (0.003)	-0.192 (0.002)	-0.110 (0.112)	-0.062 (0.374)
lead.1lag	NA	-0.168 (0.009)	-0.076 (0.273)	-0.022 (0.760)
coinc.0lags.SV	0.018 (0.558)	0.085 (0.010)	0.195 (0.002)	0.279 (0.000)
coinc.1lag.SV	NA	0.137 (0.005)	0.221 (0.003)	0.315 (0.000)
coinc.fin.0lags.SV	0.102 (0.001)	0.106 (0.000)	0.205 (0.011)	0.247 (0.003)
coinc.fin.1lag.SV	NA	0.193 (0.003)	0.208 (0.031)	0.262 (0.010)
coinc.lead.0lags.SV	0.098 (0.004)	0.126 (0.012)	0.201 (0.013)	0.271 (0.001)
coinc.lead.1lag.SV	NA	0.137 (0.108)	0.152 (0.156)	0.077 (0.751)
coinc.lead.fin.0lags.SV	0.125 (0.002)	0.126 (0.017)	0.182 (0.065)	0.227 (0.020)
fin.0lags.SV	0.001 (0.973)	-0.011 (0.651)	-0.020 (0.523)	-0.009 (0.815)
fin.1lag.SV	NA	-0.013 (0.657)	-0.009 (0.788)	-0.004 (0.918)
lead.0lags.SV	0.047 (0.036)	0.078 (0.019)	0.114 (0.018)	0.137 (0.017)
lead.1lag.SV	NA	0.110 (0.008)	0.141 (0.010)	0.179 (0.009)

Notes: See Table 1 and sections 3 and 4 for the definition of the models. The average log score and the equal forecast accuracy test are described in section 6.1. The reported scores reflect GDP growth defined in annualized percentage terms.

Table 5. Average log scores relative to AR-SV benchmark, 1985:Q1-2008:Q2

(Score for AR-SV, differences in scores for all others)

(p-values of equal mean scores in parentheses)

forecast	month 1 qrtr. t	month 2 qrtr. t	month 3 qrtr. t	month 1 qrtr. $t+1$
ARbaseline.SV	-2.091	-2.049	-2.049	-2.047
ARbaseline	-0.307 (0.000)	-0.312 (0.000)	-0.310 (0.000)	-0.307 (0.000)
coinc.0lags	-0.251 (0.000)	-0.208 (0.000)	-0.055 (0.275)	-0.002 (0.962)
coinc.1lag	NA	-0.173 (0.000)	-0.035 (0.492)	0.024 (0.646)
coinc.fin.0lags	-0.226 (0.000)	-0.188 (0.000)	-0.029 (0.575)	0.002 (0.975)
coinc.fin.1lag	NA	-0.133 (0.010)	-0.000 (0.997)	0.033 (0.514)
coinc.lead.0lags	-0.243 (0.000)	-0.177 (0.000)	0.009 (0.871)	0.042 (0.461)
coinc.lead.1lag	NA	-0.107 (0.037)	0.037 (0.508)	0.071 (0.198)
coinc.lead.fin.0lags	-0.216 (0.000)	-0.155 (0.002)	0.003 (0.959)	0.034 (0.535)
fin.0lags	-0.273 (0.000)	-0.304 (0.000)	-0.298 (0.000)	-0.300 (0.000)
fin.1lag	NA	-0.294 (0.000)	-0.280 (0.000)	-0.275 (0.000)
lead.0lags	-0.274 (0.000)	-0.256 (0.000)	-0.182 (0.001)	-0.133 (0.013)
lead.1lag	NA	-0.235 (0.000)	-0.149 (0.004)	-0.096 (0.070)
coinc.0lags.SV	0.028 (0.229)	0.079 (0.022)	0.178 (0.002)	0.259 (0.000)
coinc.1lag.SV	NA	0.121 (0.003)	0.199 (0.003)	0.286 (0.000)
coinc.fin.0lags.SV	0.088 (0.002)	0.102 (0.001)	0.187 (0.003)	0.218 (0.001)
coinc.fin.1lag.SV	NA	0.153 (0.001)	0.159 (0.044)	0.203 (0.023)
coinc.lead.0lags.SV	0.074 (0.010)	0.102 (0.034)	0.176 (0.007)	0.227 (0.001)
coinc.lead.1lag.SV	NA	0.085 (0.255)	0.097 (0.322)	-0.014 (0.959)
coinc.lead.fin.0lags.SV	0.086 (0.019)	0.103 (0.029)	0.150 (0.024)	0.182 (0.019)
fin.0lags.SV	-0.008 (0.775)	-0.018 (0.520)	-0.012 (0.710)	0.010 (0.732)
fin.1lag.SV	NA	-0.022 (0.512)	0.003 (0.933)	0.012 (0.741)
lead.0lags.SV	0.031 (0.192)	0.056 (0.106)	0.076 (0.077)	0.086 (0.066)
lead.1lag.SV	NA	0.077 (0.051)	0.096 (0.037)	0.113 (0.030)

Notes: See Table 1 and sections 3 and 4 for the definition of the models. The average log score and the equal forecast accuracy test are described in section 6.1. The reported scores reflect GDP growth defined in annualized percentage terms.

Table 6. Coverage rates, nominal 70%, 1985:Q1-2011:Q3
(p-values of correct coverage in parentheses)

forecast	month 1 qrtr. t	month 2 qrtr. t	month 3 qrtr. t	month 1 qrtr. $t+1$
ARbaseline	0.925 (0.000)	0.944 (0.000)	0.944 (0.000)	0.944 (0.000)
ARbaseline.SV	0.720 (0.653)	0.692 (0.851)	0.720 (0.653)	0.729 (0.502)
coinc.0lags	0.925 (0.000)	0.935 (0.000)	0.925 (0.000)	0.916 (0.000)
coinc.1lag	NA	0.916 (0.000)	0.907 (0.000)	0.897 (0.000)
coinc.fin.0lags	0.925 (0.000)	0.925 (0.000)	0.916 (0.000)	0.907 (0.000)
coinc.fin.1lag	NA	0.935 (0.000)	0.907 (0.000)	0.879 (0.000)
coinc.lead.0lags	0.925 (0.000)	0.916 (0.000)	0.907 (0.000)	0.907 (0.000)
coinc.lead.1lag	NA	0.916 (0.000)	0.860 (0.000)	0.897 (0.000)
coinc.lead.fin.0lags	0.925 (0.000)	0.925 (0.000)	0.869 (0.000)	0.897 (0.000)
fin.0lags	0.944 (0.000)	0.944 (0.000)	0.944 (0.000)	0.944 (0.000)
fin.1lag	NA	0.935 (0.000)	0.944 (0.000)	0.935 (0.000)
lead.0lags	0.935 (0.000)	0.944 (0.000)	0.925 (0.000)	0.925 (0.000)
lead.1lag	NA	0.944 (0.000)	0.935 (0.000)	0.916 (0.000)
coinc.0lags.SV	0.748 (0.259)	0.729 (0.502)	0.748 (0.259)	0.757 (0.171)
coinc.1lag.SV	NA	0.757 (0.171)	0.701 (0.983)	0.701 (0.983)
coinc.fin.0lags.SV	0.729 (0.502)	0.729 (0.502)	0.729 (0.502)	0.729 (0.502)
coinc.fin.1lag.SV	NA	0.664 (0.427)	0.636 (0.168)	0.645 (0.236)
coinc.lead.0lags.SV	0.757 (0.171)	0.701 (0.983)	0.710 (0.816)	0.710 (0.816)
coinc.lead.1lag.SV	NA	0.654 (0.322)	0.579 (0.012)	0.589 (0.020)
coinc.lead.fin.0lags.SV	0.692 (0.851)	0.673 (0.552)	0.626 (0.116)	0.720 (0.653)
fin.0lags.SV	0.785 (0.033)	0.748 (0.259)	0.748 (0.259)	0.729 (0.502)
fin.1lag.SV	NA	0.794 (0.016)	0.738 (0.369)	0.757 (0.171)
lead.0lags.SV	0.785 (0.033)	0.794 (0.016)	0.813 (0.003)	0.785 (0.033)
lead.1lag.SV	NA	0.757 (0.171)	0.748 (0.259)	0.729 (0.502)

Notes: See Table 1 and sections 3 and 4 for the definition of the models. The coverage rate and the test of correct coverage are described in section 6.1.

Table 7. Coverage rates, nominal 70%, 1985:Q1-2008:Q2

(p-values of correct coverage in parentheses)

forecast	month 1 qrtr. t	month 2 qrtr. t	month 3 qrtr. t	month 1 qrtr. $t+1$
ARbaseline	0.947 (0.000)	0.957 (0.000)	0.968 (0.000)	0.957 (0.000)
ARbaseline.SV	0.723 (0.614)	0.691 (0.859)	0.723 (0.614)	0.745 (0.323)
coinc.0lags	0.947 (0.000)	0.947 (0.000)	0.936 (0.000)	0.926 (0.000)
coinc.1lag	NA	0.936 (0.000)	0.915 (0.000)	0.904 (0.000)
coinc.fin.0lags	0.957 (0.000)	0.947 (0.000)	0.926 (0.000)	0.915 (0.000)
coinc.fin.1lag	NA	0.957 (0.000)	0.915 (0.000)	0.894 (0.000)
coinc.lead.0lags	0.947 (0.000)	0.947 (0.000)	0.915 (0.000)	0.915 (0.000)
coinc.lead.1lag	NA	0.936 (0.000)	0.862 (0.000)	0.904 (0.000)
coinc.lead.fin.0lags	0.947 (0.000)	0.947 (0.000)	0.872 (0.000)	0.894 (0.000)
fin.0lags	0.968 (0.000)	0.968 (0.000)	0.968 (0.000)	0.968 (0.000)
fin.1lag	NA	0.957 (0.000)	0.968 (0.000)	0.957 (0.000)
lead.0lags	0.947 (0.000)	0.957 (0.000)	0.957 (0.000)	0.957 (0.000)
lead.1lag	NA	0.957 (0.000)	0.957 (0.000)	0.936 (0.000)
coinc.0lags.SV	0.777 (0.076)	0.745 (0.323)	0.745 (0.323)	0.755 (0.215)
coinc.1lag.SV	NA	0.777 (0.076)	0.713 (0.786)	0.691 (0.859)
coinc.fin.0lags.SV	0.755 (0.215)	0.766 (0.133)	0.745 (0.323)	0.734 (0.457)
coinc.fin.1lag.SV	NA	0.691 (0.859)	0.638 (0.216)	0.638 (0.216)
coinc.lead.0lags.SV	0.787 (0.040)	0.734 (0.457)	0.702 (0.964)	0.702 (0.964)
coinc.lead.1lag.SV	NA	0.660 (0.411)	0.564 (0.008)	0.585 (0.025)
coinc.lead.fin.0lags.SV	0.691 (0.859)	0.702 (0.964)	0.638 (0.216)	0.713 (0.786)
fin.0lags.SV	0.787 (0.040)	0.755 (0.215)	0.755 (0.215)	0.734 (0.457)
fin.1lag.SV	NA	0.809 (0.008)	0.745 (0.323)	0.777 (0.076)
lead.0lags.SV	0.809 (0.008)	0.819 (0.003)	0.830 (0.001)	0.798 (0.019)
lead.1lag.SV	NA	0.787 (0.040)	0.766 (0.133)	0.734 (0.457)

Notes: See Table 1 and sections 3 and 4 for the definition of the models. The coverage rate and the test of correct coverage are described in section 6.1.

Table 8. RMSEs relative to AR benchmark, 1985:Q1-2011:Q3
BMF models with TVP
BMF models with TVP

forecast	month 1 qrtr. t	month 2 qrtr. t	month 3 qrtr. t	month 1 qrtr. $t+1$
Models with constant coefficients				
coinc.0lags	0.932 (0.167)	0.916(0.102)	0.823 (0.120)	0.770 (0.091)
fin.0lags	1.030 (0.451)	1.035(0.326)	1.030 (0.353)	1.023 (0.456)
lead.0lags	0.959 (0.156)	0.920(0.102)	0.844 (0.111)	0.805 (0.120)
coinc.0lags.SV	0.936 (0.248)	0.906(0.098)	0.815 (0.108)	0.749 (0.071)
fin.0lags.SV	0.996 (0.863)	1.001(0.946)	1.001 (0.950)	0.987 (0.632)
lead.0lags.SV	0.949 (0.130)	0.897(0.083)	0.835 (0.135)	0.809 (0.157)
Models with TVP				
TVP.coinc.0lags	0.925 (0.493)	0.950(0.601)	0.874 (0.264)	0.831 (0.157)
TVP.fin.0lags	1.036 (0.227)	1.021(0.483)	1.016 (0.656)	1.009 (0.827)
TVP.lead.0lags	0.934 (0.341)	0.915(0.435)	0.909 (0.533)	0.882 (0.439)
TVP.coinc.0lags.SV	0.963 (0.712)	0.916(0.374)	0.882 (0.286)	0.835 (0.164)
TVP.fin.0lags.SV	1.010 (0.783)	1.029(0.352)	1.035 (0.366)	1.012 (0.772)
TVP.lead.0lags.SV	1.020 (0.793)	0.941(0.540)	0.945 (0.679)	0.891 (0.449)

Notes: See Table 1 and sections 3 and 4 for the definition of the models. The equal forecast accuracy test is described in section 6.1. The reported RMSEs reflect GDP growth defined in annualized percentage terms.

Table 9. Average log scores relative to AR-SV benchmark, 1985:Q1-2011:Q3
BMF models with TVP
(p-values of equal mean scores in parentheses)

forecast	month 1 qrtr. t	month 2 qrtr. t	month 3 qrtr. t	month 1 qrtr. $t+1$
Models with constant coefficients				
coinc.0lags	-0.177 (0.015)	-0.151(0.009)	-0.007 (0.916)	0.045 (0.493)
fin.0lags	-0.229 (0.000)	-0.260(0.000)	-0.258 (0.000)	-0.266 (0.000)
lead.0lags	-0.201 (0.003)	-0.192(0.002)	-0.110 (0.112)	-0.062 (0.374)
coinc.0lags.SV	0.018 (0.558)	0.085(0.010)	0.195 (0.002)	0.279 (0.000)
fin.0lags.SV	0.001 (0.973)	-0.011(0.651)	-0.020 (0.523)	-0.009 (0.815)
lead.0lags.SV	0.047 (0.036)	0.078(0.019)	0.114 (0.018)	0.137 (0.017)
Models with TVP				
TVP.coinc.0lags	-0.179 (0.058)	-0.084(0.230)	0.017 (0.787)	0.039 (0.516)
TVP.fin.0lags	-0.218 (0.000)	-0.248(0.000)	-0.242 (0.000)	-0.239 (0.000)
TVP.lead.0lags	-0.263 (0.003)	-0.147(0.067)	-0.118 (0.167)	-0.147 (0.062)
TVP.coinc.0lags.SV	0.031 (0.638)	0.086(0.112)	0.116 (0.057)	0.154 (0.012)
TVP.fin.0lags.SV	-0.012 (0.666)	-0.026(0.311)	-0.010 (0.737)	0.014 (0.678)
TVP.lead.0lags.SV	-0.088 (0.177)	0.023(0.688)	-0.016 (0.839)	0.001 (0.991)

Notes: See Table 1 and sections 3 and 4 for the definition of the models. The equal forecast accuracy test is described in section 6.1. The reported RMSEs reflect GDP growth defined in annualized percentage terms.

Appendix Table 1. Sums of log predictive likelihoods, 1985:Q1-2011:Q3

(Sum for AR-SV benchmark, difference in LPL in all others)

model	month 1 qrtr. t	month 2 qrtr. t	month 3 qrtr. t	month 1 qrtr. $t+1$
ARbaseline	-26.223	-27.629	-28.265	-28.810
ARbaseline.SV	-236.461	-229.425	-228.348	-227.110
coinc.0lags	-18.946	-16.201	-0.739	4.828
coinc.1lag	NA	-11.596	2.072	8.619
coinc.fin.0lags	-17.292	-14.277	3.407	6.568
coinc.fin.1lag	NA	-6.681	7.160	10.688
coinc.lead.0lags	-17.316	-11.522	7.587	11.183
coinc.lead.1lag	NA	-3.401	11.175	14.898
coinc.lead.fin.0lags	-15.517	-9.743	7.514	10.090
fin.0lags	-24.456	-27.819	-27.558	-28.446
fin.1lag	NA	-26.987	-25.782	-25.940
lead.0lags	-21.541	-20.499	-11.760	-6.643
lead.1lag	NA	-18.018	-8.164	-2.352
coinc.0lags.SV	1.940	9.106	20.827	29.839
coinc.1lag.SV	NA	14.674	23.605	33.699
coinc.fin.0lags.SV	10.928	11.349	21.885	26.473
coinc.fin.1lag.SV	NA	20.629	22.275	28.057
coinc.lead.0lags.SV	10.492	13.436	21.538	29.000
coinc.lead.1lag.SV	NA	14.621	16.317	8.212
coinc.lead.fin.0lags.SV	13.403	13.437	19.499	24.253
fin.0lags.SV	0.097	-1.192	-2.160	-0.986
fin.1lag.SV	NA	-1.439	-0.987	-0.459
lead.0lags.SV	5.070	8.392	12.240	14.697
lead.1lag.SV	NA	11.742	15.085	19.129

Notes: See Table 1 for the definition of the models. SV indicates BMF specifications with stochastic volatility. The log predictive likelihood is defined in section 6.1. The reported likelihoods reflect GDP growth defined in annualized percentage terms.

Appendix Table 2. Sums of log predictive likelihoods, 1985:Q1-2008:Q2

(Sum for AR-SV benchmark, difference in LPL in all others)

model	month 1 qrtr. t	month 2 qrtr. t	month 3 qrtr. t	month 1 qrtr. $t+1$
ARbaseline	-28.832	-29.353	-29.095	-28.903
ARbaseline.SV	-196.511	-192.646	-192.606	-192.392
coinc.0lags	-23.630	-19.544	-5.213	-0.230
coinc.1lag	NA	-16.292	-3.288	2.230
coinc.fin.0lags	-21.254	-17.653	-2.724	0.151
coinc.fin.1lag	NA	-12.472	-0.021	3.104
coinc.lead.0lags	-22.829	-16.675	0.857	3.916
coinc.lead.1lag	NA	-10.098	3.453	6.701
coinc.lead.fin.0lags	-20.269	-14.523	0.272	3.176
fin.0lags	-25.681	-28.589	-28.056	-28.208
fin.1lag	NA	-27.592	-26.360	-25.804
lead.0lags	-25.775	-24.046	-17.136	-12.505
lead.1lag	NA	-22.124	-14.024	-9.069
coinc.0lags.SV	2.593	7.416	16.710	24.325
coinc.1lag.SV	NA	11.402	18.668	26.872
coinc.fin.0lags.SV	8.270	9.560	17.549	20.504
coinc.fin.1lag.SV	NA	14.369	14.912	19.107
coinc.lead.0lags.SV	6.984	9.587	16.536	21.374
coinc.lead.1lag.SV	NA	8.001	9.151	-1.288
coinc.lead.fin.0lags.SV	8.120	9.681	14.114	17.107
fin.0lags.SV	-0.778	-1.680	-1.085	0.909
fin.1lag.SV	NA	-2.037	0.265	1.097
lead.0lags.SV	2.914	5.242	7.132	8.128
lead.1lag.SV	NA	7.268	9.042	10.591

Notes: See Table 1 for the definition of the models. SV indicates BMF specifications with stochastic volatility. The log predictive likelihood is defined in section 6.1. The reported likelihoods reflect GDP growth defined in annualized percentage terms.

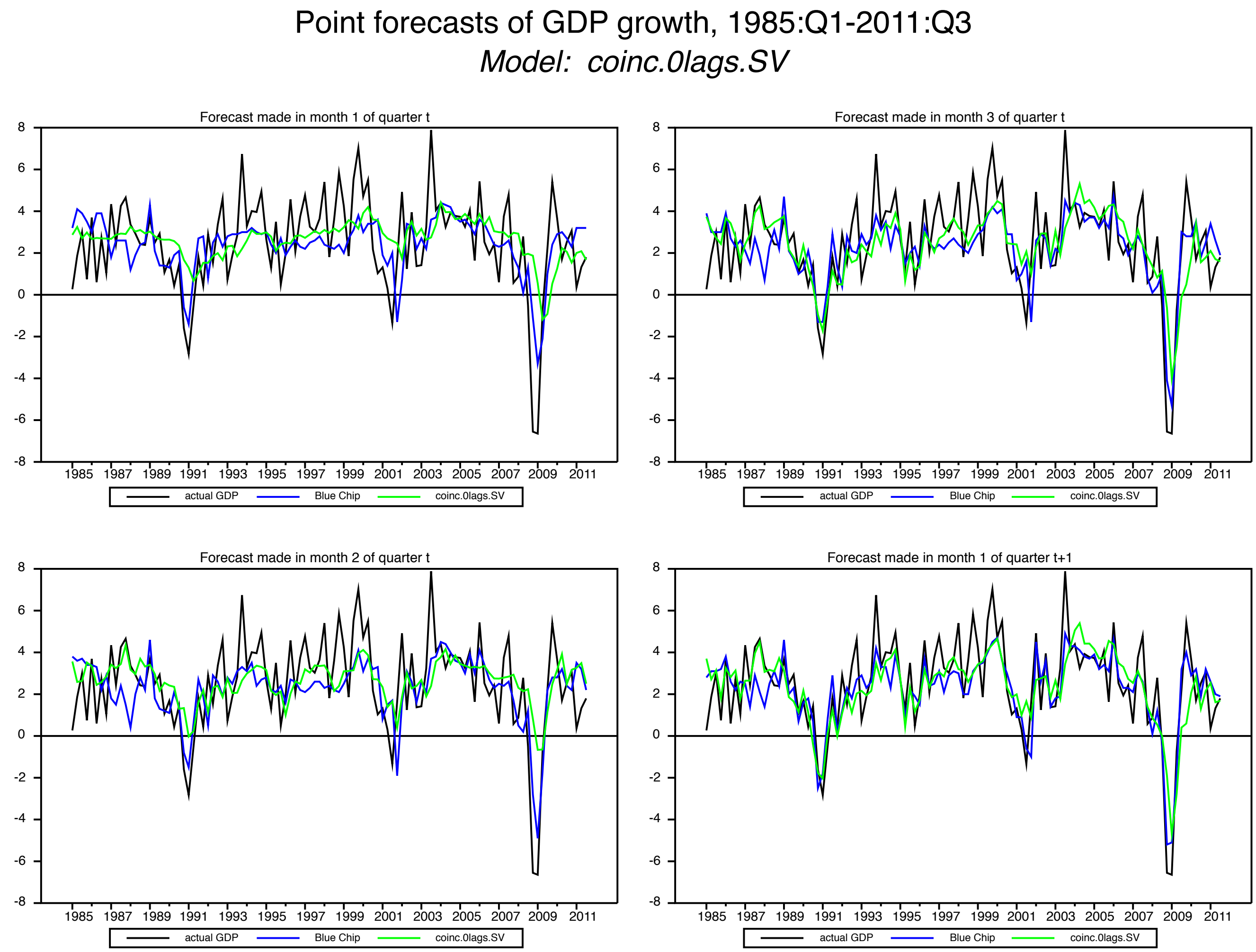

Figure 1: Real-time point forecasts from Blue Chip and BMF-SV model with coincident indicators

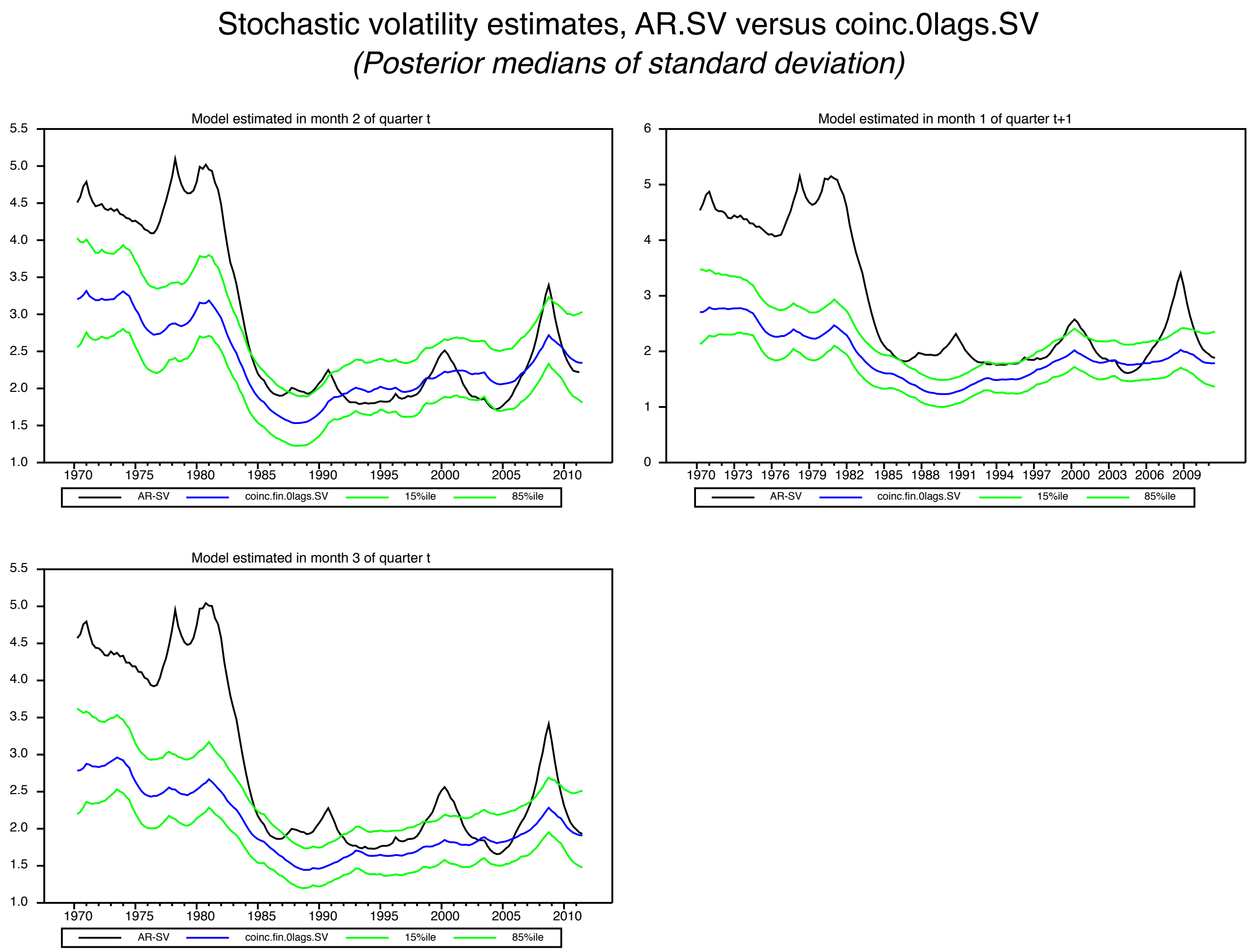

Figure 2: Volatility ($\lambda_{m,t}^{0.5}$ of equation (3)) estimates from selected models, last vintage of data

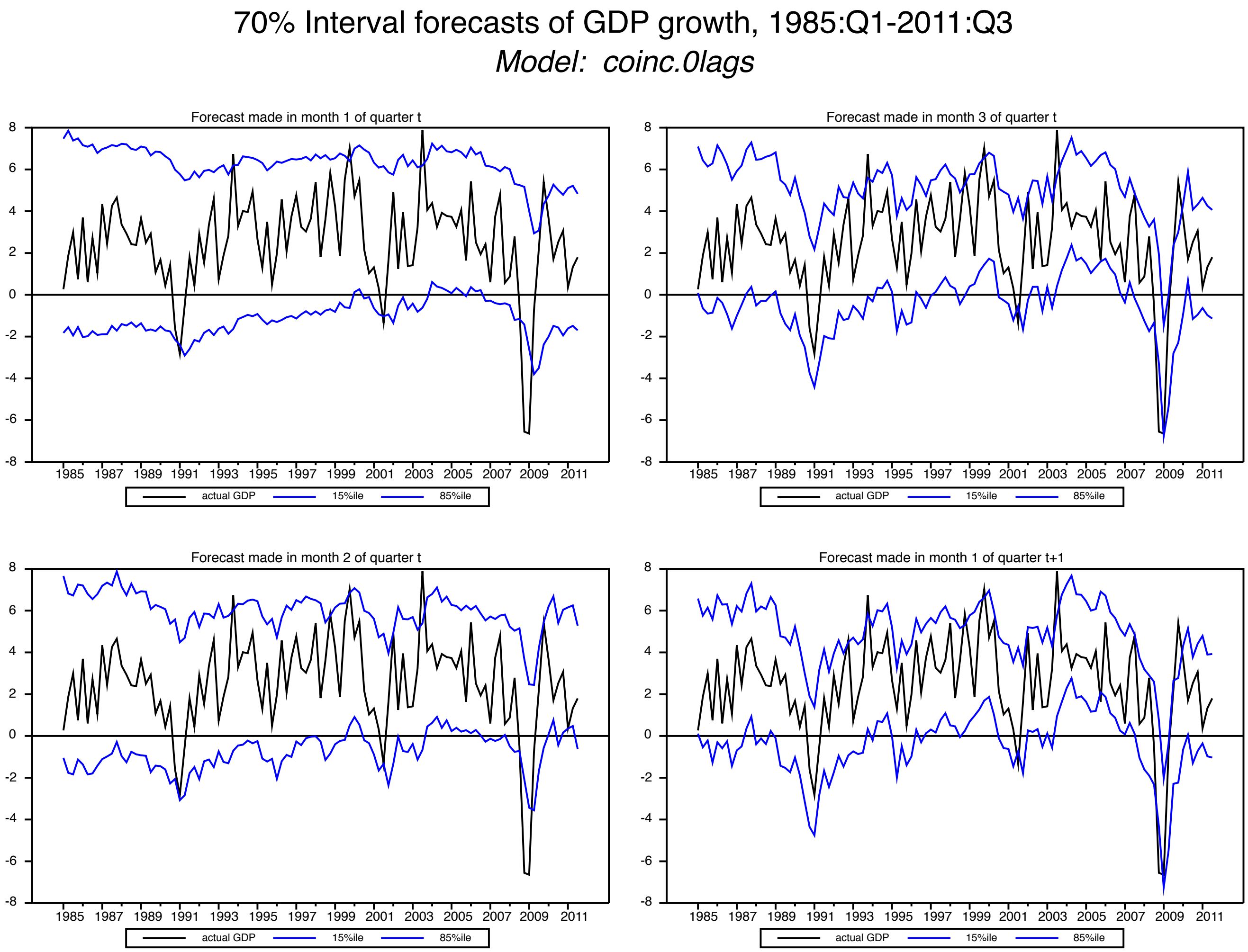

Figure 3: Real-time interval forecasts from BMF model with coincident indicators

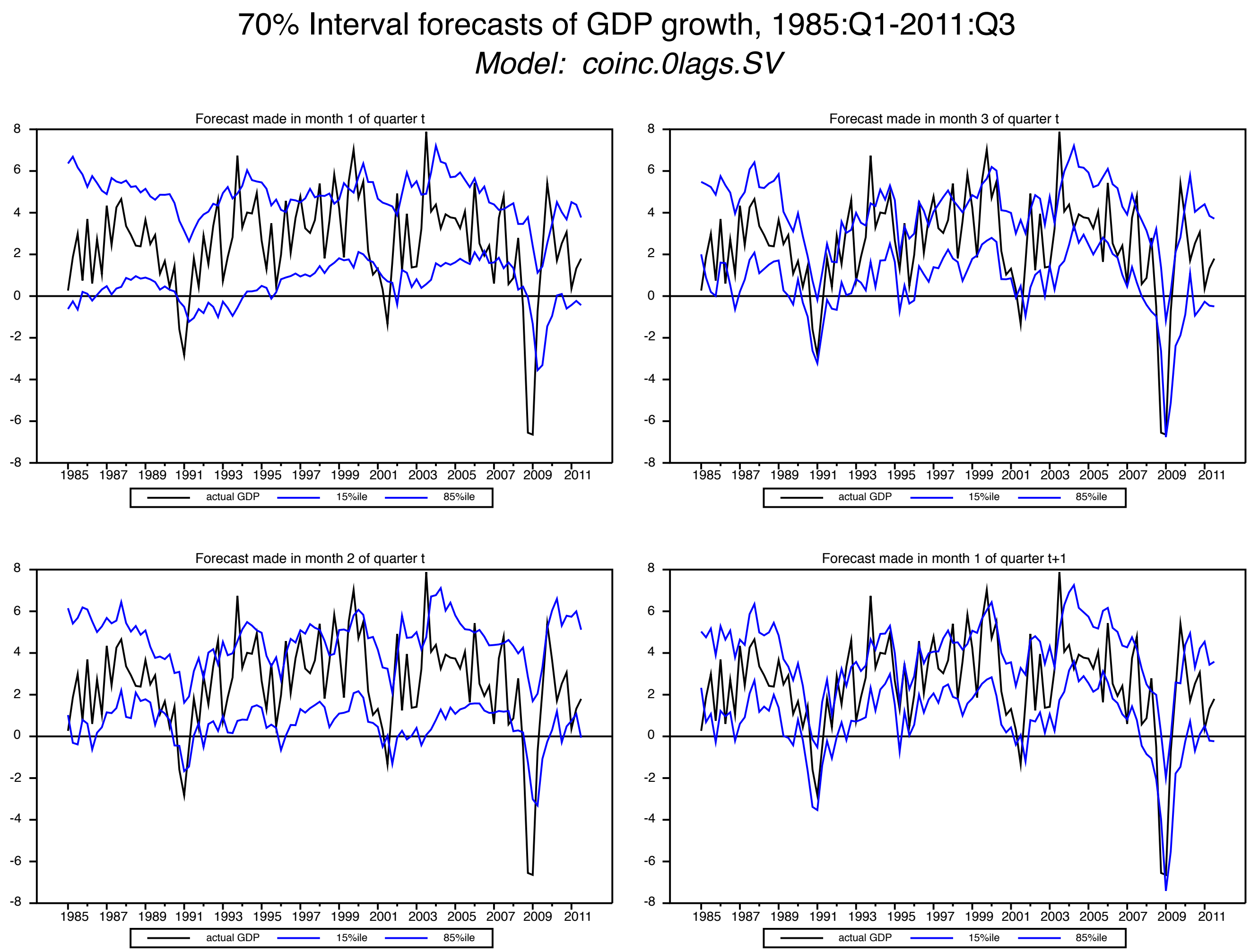

Figure 4: Real-time interval forecasts from BMF-SV model with coincident indicators

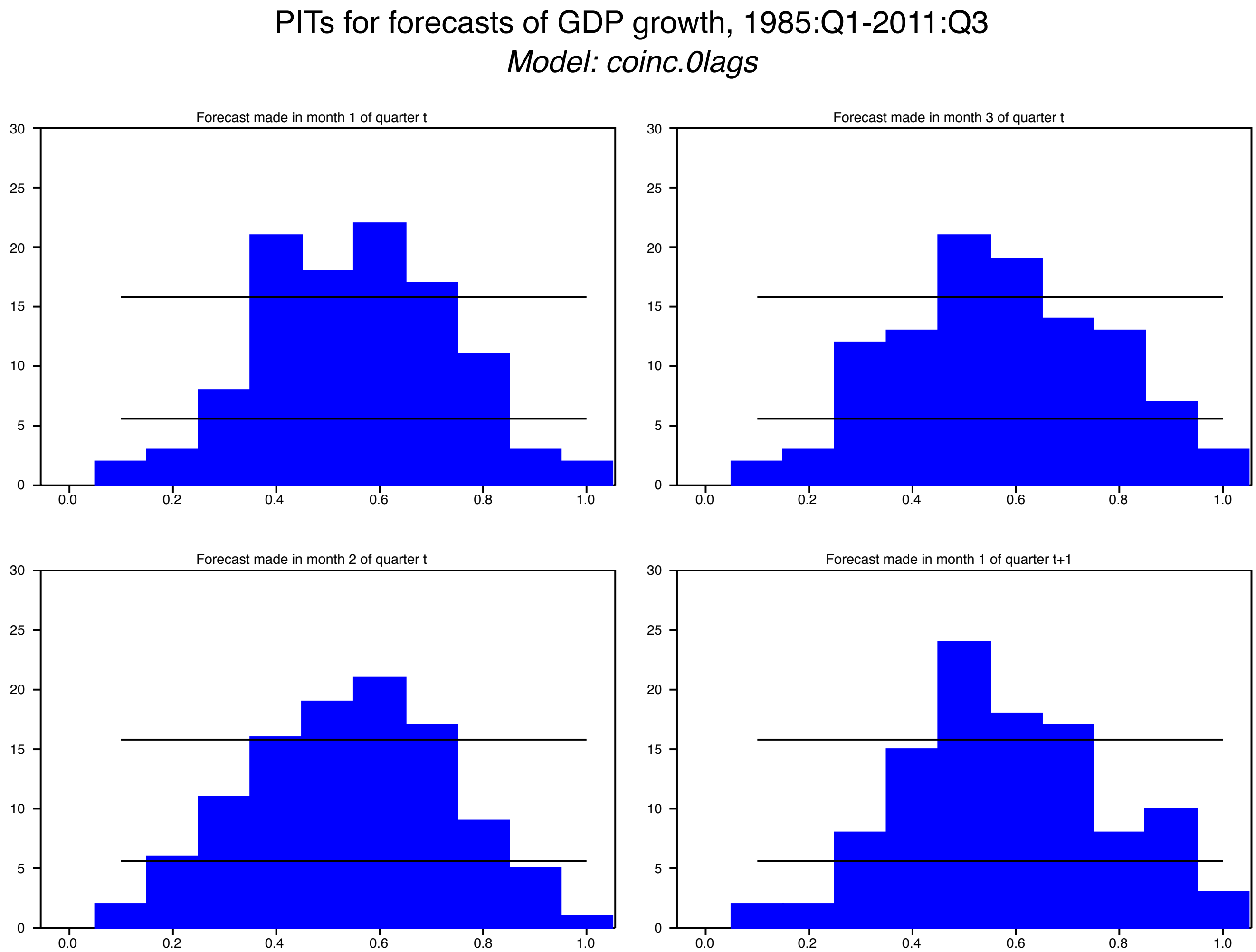

Figure 5: PITs histograms, BMF model with coincident indicators

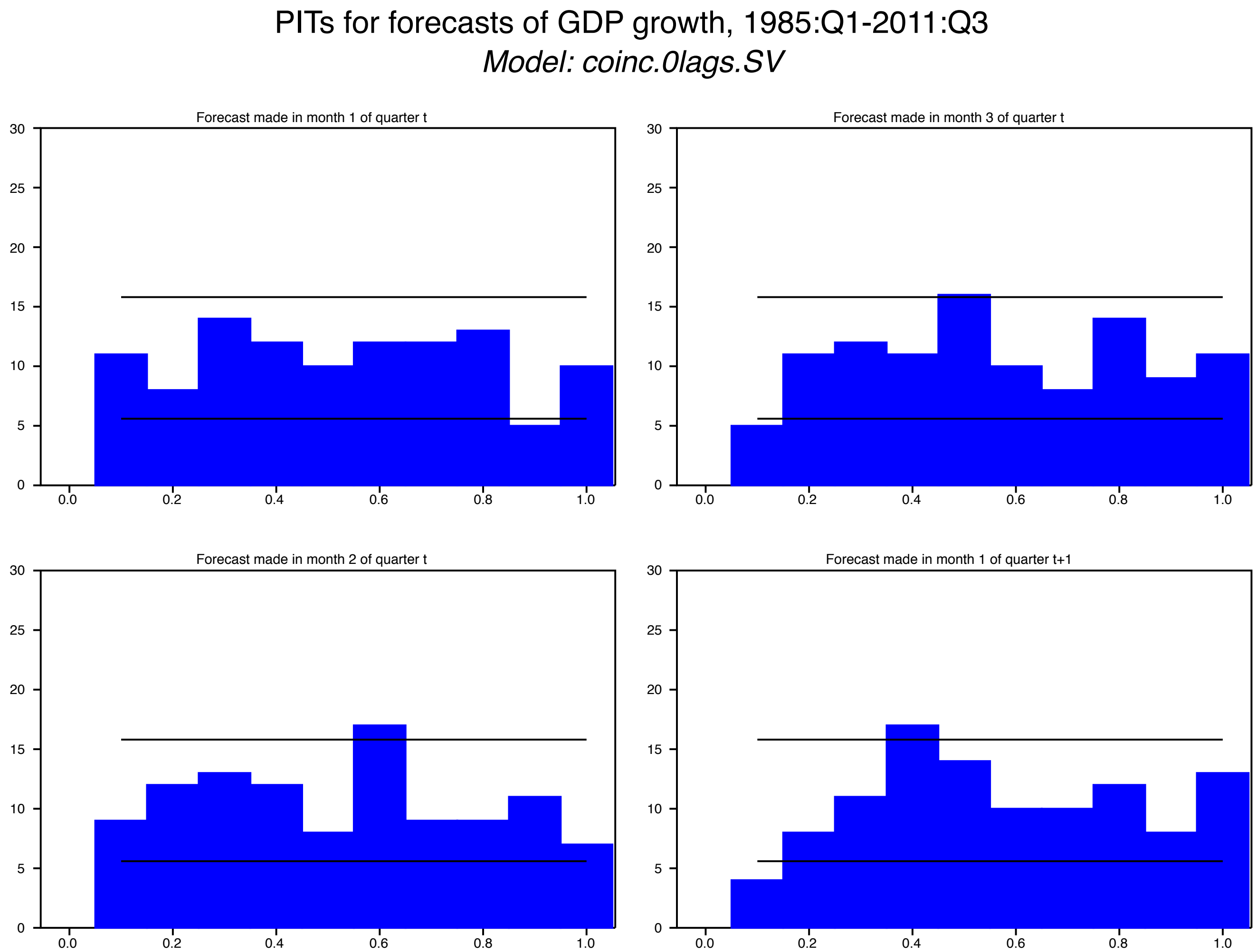

Figure 6: PITs histograms, BMF-SV model with coincident indicators