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

Andrea Carriero

Queen Mary, University of London

a.carriero@qmul.ac.uk

Todd E. Clark

Federal Reserve Bank of Cleveland

todd.clark@clev.frb.org

Massimiliano Marcellino

Bocconi University, IGIER and CEPR

massimiliano.marcellino@unibocconi.it

February 12, 2014

Abstract

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 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 improves significantly on AR models and performs comparably to survey forecasts. In addition, it provides reliable density forecasts, for which the stochastic volatility specification is quite useful.

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

J.E.L. Classification: E37, C53, C22

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 (2013). 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, 2013) 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 method we propose in this paper is motivated in part by three 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 (2013) 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. 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.

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, we consider both point and density forecasts. It turns out that in terms of point forecasts our proposal improves significantly on AR models and performs comparably to survey forecasts. In addition, it easily provides reliable density forecasts, for which the stochastic volatility specification is quite useful.

To place our proposed approach within the broader nowcasting literature, it is helpful to use the "partial model" (or single equation) methods and "full system" methods classification used by Banbura, et al. (2013). The former type of approach involves specifications focused on the low frequency variable, in which the high frequency explanatory variables are not modeled. In the latter approach, the low and high frequency variables are jointly modeled. Our proposed modeling approach falls in the partial models class.

Among partial model methods, bridge and MIDAS models are most commonly used. Bridge 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. The period t average of each monthly indicator is obtained with data 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 to macroeconomic forecasting by, e.g., Clements and Galvao (2008) and Guerin and Marcellino (2013), 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, a specification labeled unrestricted MIDAS, or U-MIDAS.

Full system methods for nowcasting include factor models and mixed frequency VARs. We refer to the surveys in Banbura et al. (2013) and Foroni, Ghysels and Marcellino (2013) for details and references. Here we only mention a few studies closely related to our proposal. These include: Aastveit et al. (2013), which, in contrast to most of the nowcasting literature, focuses on density forecasts; Chiu, et al. (2011); Ghysels (2012); Schorfheide and Song (2013) and McCracken and Sekhposyan (2012), both of which develop mixed frequency Bayesian VARs; and Marcellino, Porqueddu and Venditti (2012), which introduces a small scale factor model that allows for stochastic volatility in the common and idiosyncratic components, and provides density forecasts.

Relative to the existing partial model and full system 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 often improves the accuracy of forecasts from time series models, and it 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 easily able to allow for stochastic volatility, a feature important to the accuracy of density forecasts, mostly neglected so far in the nowcasting literature (we are also easily able to allow time-varying regression coefficients). The ability to tractably include a

large set of indicators and stochastic volatility in the model gives our approach an advantage over other approaches in the partial model and full system classes. For example, with MIDAS methods, it is computationally difficult to consider more than a few indicators, while with factor model methods, it is computationally difficult to include stochastic volatility unless only a small set of variables are used.

The paper is structured as follows. Because data choices and availability play into our model specification choices, we first present the data in Section 2. Section 3 details our model and estimation method, and Section 4 introduces competing nowcasts. We then present results in Section 5. Finally, we provide some concluding remarks in section 6.

2   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.

To forecast GDP, we consider 12 monthly indicators broadly informative about economic and financial development, selected with some eye to timeliness: 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.

In selecting the set of 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)). 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

4

application on monthly indicators, in light of the finding by Banbura, et al. (2013) that higher frequency information does not seem to be especially useful for nowcasting U.S. GDP growth (except perhaps in a continuous monitoring context).

Of the variables we do use, for those 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. 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. 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.

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.

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. We have also computed results using the first estimate of GDP and obtained qualitatively very similar results.

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

This section details our proposed nowcasting models. To help the discussion flow, we first specify the general model forms in section 3.1 and then in section 3.2 detail the sets of indicators in the model. We conclude by presenting in sections 3.3 and 3.4 the priors and algorithms used in estimation.

3.1   General model forms

Starting with our specification that treats 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 within the quarter. As detailed below, given a set of monthly indicators to be used, there is a different regressor set $X_{m,t}$ (and therefore model) for each month m within the quarter, reflecting data availability.

In the stochastic volatility case, our proposed forecasting model for month m within the

quarter takes the form:

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

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 unreported results, we found that treating log volatility as an AR(1) process with a coefficient of 0.9 slightly reduced forecast accuracy). In a vector autoregressive context, studies such as Clark (2011), Carriero, Clark, and Marcellino (2012), and D'Agostino, Gambetti, and Giannone (2013) have found that this type of stochastic volatility formulation improves the accuracy of both point and density forecasts.

Conceptually, one of the merits of our approach is that the BMF models with constant volatility or stochastic volatility can be easily extended to allow for time-varying parameters (TVP) in the conditional mean, as detailed in Carriero et al. (2013). We do not report the details in this paper since empirically TVP did not improve either point or density forecasts.

The specification of the regressor vector $X_{m,t}$ in the BMF and BMF-SV models 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 $X_{m,t}$ 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 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 are given in the next subsection). At these points in time, the availability of other indicators also varies. As a

consequence, the model specification changes in each month of the quarter, reflecting and accommodating the ragged edge of the data. 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 bridge methods. Bridge 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).

3.2 Indicators used

We report below results for both "large" and "small" versions of the BMF and BMF-SV models. The large version includes a broad set of 12 monthly indicators: payroll employment ($\Delta \log$); industrial production ($\Delta \log$); real retail sales ($\Delta \log$); housing starts (log); 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 ($\Delta \log$); new claims for unemployment insurance; 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. The small version uses just the first five indicators of the 12-variable set, which might be considered primary contemporaneous indicators of economic activity. In the results reported in this paper, we only include in the model values of these variables for the current quarter t (the quarter for which GDP growth is being forecast). However, our general approach easily allows the use of values from previous quarters (while this makes the models even larger, Bayesian shrinkage helps limit the effects of parameter estimation error on forecast accuracy). Indeed, in Carriero et al. (2013) we also report results for models in which the period $t-1$ (previous quarter) values of every variable is also included as a predictor.

8

Both the large and small model specifications all include in $X_{m,t}$ 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.

As noted above, depending on the month of the quarter the forecast is being formed, exactly what variables are in the large and small BMF and BMF-SV models (that is, in $X_{m,t}$) varies. Table 1 details the model specifications (and variable timing) we use, based on the usual publication schedules of the indicators. Consider, for example, the version of the model used to forecast GDP growth as of month $m = 2$ of the quarter. In this case, reflecting data availability, the small BMF and BMF-SV models include in $X_{m,t}$ the following: a constant, GDP growth in quarter $t - 1$, and employment growth and the ISM index in month 1 of quarter t. As of month $m = 3$ of the quarter, with more data available, the small BMF and BMF-SV models include in $X_{m,t}$ the following: a constant, GDP growth in quarter $t - 1$, employment growth and the ISM index in month 2 of quarter t, and employment growth, the ISM, growth in industrial production, growth in retail sales, and log housing starts in month 1 of quarter t.

Among these specifications, the largest empirical model we consider includes 34 explanatory variables in $X_{m,t}$ (including up to three months of observations within the quarter for 12 different monthly indicators, a constant, and a lag of GDP — see Table 1 for the precise list). While the "small" BMF model is relatively small, it is not small in an absolute sense: depending on the month of the quarter, the model includes in $X_{m,t}$ up to 14 regressors (including up to three months of observations within the quarter for the 5 monthly indicators, a constant, and a lag of GDP growth — see Table 1 for the precise list). With models of these sizes, under 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).

3.3   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 form 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 small BMF 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 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 (while we only consider a lag of 1 in this paper, Carriero, et al. (2013) include results for models with a lag of 2), is specified as follows:

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

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

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

Finally, for the intercept, the prior is uninformative:

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

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 $\lambda_3 = 1$. In Carriero, et al. (2013), we ran some limited checks (for BMF models with different subsets of our 12 monthly indicators) 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. For example, using a model with nine economic 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 moved 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 (2014) find the payoff to optimization over fixed, conventional shrinkage to be small.

Finally, 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.

11

3.4 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.

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 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.

In presenting our results, we focus on forecasts obtained by estimating the forecasting models with a recursive scheme: the estimation sample expands as forecasting moves forward in time. A rolling scheme, under which the size of the estimation sample remains fixed over time but the first observation moves forward in time, is in general less efficient but can be more robust in the presence of changes in regression parameters and (for density forecasts) error variances. Hence, for the BMF models with constant volatility we also report results based on a rolling estimation scheme. However, as we will show below, rolling window estimation of the model is not sufficient to get point and density forecasts that are as good as those obtained with the (recursive) BMF-SV specification (with a gap that is particularly large for density forecasts).

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, 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 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

12

states, parameters, and shocks over forecast interval).

4   Competing Nowcasts

We compare our BMF and BMF-SV nowcasts with those generated from AR models and with survey-based forecasts (which pool many predictions, based on timely information). These are typically tough benchmarks in forecast competitions. The results in Carriero, et al. (2013) also indicate that MIDAS and U-MIDAS specifications can also produce relatively good nowcasts of GDP growth. However, these models are primarily designed for use with small sets of indicators and point forecasts; using large sets of indicators and allowing stochastic volatility to obtain reliable density forecasts is relatively difficult. As a result, we abstract from MIDAS and U-MIDAS forecasts in the comparison.

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.

13

4.2   Surveys

We also consider GDP growth nowcasts based on the Survey of Professional Forecasters (SPF), available quarterly, and the Blue Chip Consensus, available on a monthly basis, since they are closely monitored by decision makers and typically perform quite well. 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   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, the Survey of Professional Forecasters (SPF), and Blue Chip. For the 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 2, we present results for both a full sample of 1985:Q1-2011:Q3 and a pre-crisis sample of 1985:Q1-2008:Q2.

5.1   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.

For comparing our proposed BMF and BMF-SV 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. However, since forecast performance suggests it is unlikely that the AR model and nowcasting models overlap, 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, as we do. To capture some low-order serial correlation, we compute the t-statistics with a heteroskedasticity and auto-correlation consistent (HAC) variance, using a rectangular kernel and bandwidth of 1 and the small-sample adjustment of Harvey, Leybourne, and Newbold (1997).

To assess the accuracy of density forecasts, we use 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 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.

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 possible effects of forecast model estimation — that is, parameter estimation error.

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. 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.

5.2   Point forecasts

To assess the accuracy of point forecasts, Table 2 provides RMSE comparisons of our proposed BMF and BMF-SV nowcasting models, 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 (as noted above, the RMSEs of the AR model can change across months of the quarter, due to a change in the model specification from the first month of the quarter to the second and to GDP data revisions from month to month). The 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. The two panels of the table refer to the periods 1985:Q1-2011Q3 and 1985:Q1-2008Q2.

We can draw six main conclusions from the RMSE results in Table 2. 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 the 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. As a consequence, the accuracy of the nowcasting models relative to the AR baseline increases with the addition of more information on the quarter. For both the small and large BMF and BMF-SV models, the RMSE gain in the full sample results rises from about 7 percent in month 1 to 18 percent in month 3 and 23-25 percent in month 1 of the next quarter. 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. 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.

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 large BMF-SV nowcasting model. 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.

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, though the differences are not statistically significant. The challenge of beating a survey forecast with good nowcasting models is also evident in such studies as Banbura, et al. (2013), 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 (2013) 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 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 (2013) 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 Table 2 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.

Fifth, there are no major differences between the small (again, small is relative — even the small model involves as many as 14 regressors) and large BMF models. The former are slightly better when the sample ends in 2008, the latter over the full sample, suggesting that more information became relevant during the crisis. We have also experimented (in Carriero, et al. (2013)) with different subgroups of the indicators. Perhaps the most interesting finding is that financial indicators, by themselves, do poorly in forecasting current-quarter GDP growth. However, in line with the comparison of results across samples mentioned above, including financial indicators with other indicators helps the models (a little) during the

recent crisis.

Finally, rolling estimation of the BMF models with constant volatilities generates systematically higher RMSEs than recursive estimation of the same models. This finding suggests that the efficiency losses from using a smaller set of observations are larger than the gains from getting partial robustness to possible breaks.

The main message that we can take from the point forecast evaluation is that overall our BMF method is superior to AR model forecasts and comparable 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 and interval forecasts, and, as we will now see, in this context the stochastic volatility specification that we adopt becomes quite relevant.

5.3   Density forecasts: average predictive scores

To assess the calibration of density forecasts, Table 3 provides 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). 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 two panels of the table refer to the periods 1985:Q1-2011Q3 and 1985:Q1-2008Q2.

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 small BMF model 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 large BMF-SV nowcasting model, 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 and BMF-SV 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, these gains are rarely statistically significant. Moreover, in the pre-crisis sample, nowcasting models with constant volatilities have a harder time beating the AR-SV benchmark.

Third, both 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. The large model is better for short horizons, the small for longer horizons. However, results in Carriero, et al. (2013) for different subgroups of the indicators indicate that, financial indicators, by themselves (as opposed to in conjunction with other indicators, as in our large model), are not very helpful for density forecasting.

Finally, for density forecasting, rolling estimation of the BMF model with constant volatility sometimes (not always) improves on the accuracy of the recursively generated forecasts from the same model, but falls short of the recursively estimated model with stochastic volatility. Consider forecasts from the small BMF model from month 3 of the quarter. Relative to the AR-SV baseline, the recursively estimated BMF-SV model has a score differential of 19.5 percent, compared to a score differential of 1.9 percent for the rolling window version of the BMF forecast and -0.7 percent for the recursive version of the BMF forecast. This finding suggests that, in alternative model formulations such as MIDAS, it

would be necessary to incorporate stochastic volatility — which would be difficult to do given the nonlinear regression problem involved in MIDAS — to achieve material gains in density accuracy.

5.4 Interval forecasts

As another measure of density forecast accuracy, we consider interval forecasts. For all of our econometric models, Table 4 provides 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 two panels of the table refer to the periods 1985:Q1-2011Q3 and 1985:Q1-2008Q2.

The coverage rates in the tables are striking. Recursive estimation of models with constant volatilities in all cases yields 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 across months of the quarter (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.

However, estimating the same models (BMF with constant volatilities) with a rolling window of observations yields better coverage rates, as much as 10 percentage points lower than the rates obtained with recursively estimated models. But a simple rolling window approach is not enough to yield correct coverage: coverage rates for the rolling scheme versions of the BMF models are all (with one exception) statistically different from 70 percent.

Correct coverage is achieved by including stochastic volatility in our BMF specification. Our models with stochastic volatility in all cases yield coverage rates close enough to 70 percent that they are not statistically different from 70 percent (at the 5% significance level).

To further illustrate the importance of SV, in Figures 3 and 4 we report the real-time 70% interval forecasts from the large BMF model, without (Figure 3) and with SV (Figure

4). Figure 3 confirms that coverage is pretty terrible for models with constant volatilities (estimated recursively). 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 model with stochastic volatility yields much narrower bands, and therefore more outcomes that fall outside the 70 percent bands.

5.5   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 large BMF and BMF-SV models (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. 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. Hence, by the PITs measure, too, including stochastic volatility materially improves the calibration of density forecasts.

6   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

22

also consider versions of the model with stochastic volatility, while most of the existing approaches assumed that the variance is constant. Similarly, models with time-varying regression coefficients (with or without stochastic volatility), can be easily used but empirically are less promising, at least in U.S. data (see Carriero et al. (2013)).

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 improves significantly on AR models and performs comparably to 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.

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.

References

[1] Aastveit, K., Gerdrup, K., and Jore, A.S. (2013), Nowcasting GDP in Real-Time: A Density Combination Approach, *Journal of Business and Economic Statistics*, forthcoming.

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

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

[5] 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.

[6] Banbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2013), Nowcasting and the Real Time Data Flow, in *Handbook of Economic Forecasting, Volume 2*, Elliott, G., and Timmermann, A. (eds). Amsterdam: North Holland.

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

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

[9] Carriero, A., Clark, T., and Marcellino, M. (2014), Bayesian VARs: Specification Choices and Forecasting Performance, *Journal of Applied Econometrics,* forthcoming.

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

[11] Carriero, A., Clark, T., and Marcellino, M. (2013), Real-Time Nowcasting with a Bayesian Mixed Frequency Model with Stochastic Volatility, CEPR WP 9312.

[12] Chauvet, M., and Potter, S. (2013), Forecasting Output, in *Handbook of Economic Forecasting, Volume 2*, Elliott, G., and Timmermann, A. (eds). Amsterdam: North Holland.

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

[14] 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.

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

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

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

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] D'Agostino, A., Gambetti, L. and Giannone, D. (2013), Macroeconomic Forecasting and Structural Change, *Journal of Applied Econometrics* 28, 82-101.

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

[24] 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.

[25] 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.

[26] 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.

[27] Faust, J., and Wright, J. (2013), Inflation Forecasting, in *Handbook of Economic Forecasting, Volume 2*, Elliott, G., and Timmermann, A. (eds). Amsterdam: North Holland.

[28] Foroni, C., Ghysels, E., and Marcellino, M. (2013), Mixed-Frequency Vector Autoregressive Models, in Fomby, T., Kilian, L and Murphy, A. (eds), *Advances in Econometrics: VAR Models in Macroeconomics, Financial Econometrics, and Forecasting — New Developments and Applications*.

[29] Foroni, C., Marcellino, M., and Schumacher, C. (2012), U-MIDAS: MIDAS Regressions with Unrestricted Lag Polynomials, *Journal of the Royal Statistical Society*, forthcoming.

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

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

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

[33] 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.

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

[35] Guerin, P., and Marcellino, M. (2013), Markov Switching MIDAS Models, *Journal of Business and Economic Statistics*, 31(1), 45-56.

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

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

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

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

[40] 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.

[41] Marcellino, M., Porqueddu, M., and Venditti, F. (2012), Short-Term GDP Forecasting with a Mixed Frequency Dynamic Factor Model with Stochastic Volatility, Banca d'Italia, WP 896.

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

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

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

[45] Schorfheide, F. and Song, D. (2013), Real-Time Density Forecasting with a Mixed Frequency VAR, manuscript, University of Pennsylvania.

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

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

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

Table 1. Specifications of BMF models of GDP growth

	predictors in model for:			
	month 1	*month 2*	*month 3*	*month 1*
model	*quarter t*	*quarter t*	*quarter t*	*quarter t+1*
1. AR	two lags of GDP growth	two lags of GDP growth	two lags of GDP growth	two lags of GDP growth
	estimated up thru t-2	estimated up thru t-1	estimated up thru t-1	estimated up thru t-1
2. small BMF and BMF-SV	GDP(t-2)	GDP(t-1)	GDP(t-1)	GDP(t-1)
	emp(months 1-3 of t-1)	emp(month 1 of t)	emp(months 1-2 of t)	emp(months 1-3 of t)
	ISM(months 1-3 of t-1)	ISM(month 1 of t)	ISM(months 1-2 of t)	ISM(months 1-3 of t)
	IP(months 1-2 of t-1)		IP(month 1 of t)	IP(months 1-2 of t)
	RS(months 1-2 of t-1)		RS(month 1 of t)	RS(months 1-2 of t)
	starts(months 1-2 of t-1)		starts(month 1 of t)	starts(months 1-2 of t)
3. large BMF and BMF-SV	GDP(t-2)	GDP(t-1)	GDP(t-1)	GDP(t-1)
	emp(months 1-3 of t-1)	emp(month 1 of t)	emp(months 1-2 of t)	emp(months 1-3 of t)
	ISM(months 1-3 of t-1)	ISM(month 1 of t)	ISM(months 1-2 of t)	ISM(months 1-3 of t)
	IP(months 1-2 of t-1)	supdel(month 1 of t)	IP(month 1 of t)	IP(months 1-2 of t)
	RS(months 1-2 of t-1)	orders(month 1 of t)	RS(month 1 of t)	RS(months 1-2 of t)
	starts(months 1-2 of t-1)	hours(month 1 of t)	starts(month 1 of t)	starts(months 1-2 of t)
	supdel(months 1-3 of t-1)	stprice(month 1 of t)	supdel(months 1-2 of t)	supdel(months 1-3 of t)
	orders(months 1-3 of t-1)	tbill(month 1 of t)	orders(months 1-2 of t)	orders(months 1-3 of t)
	hours(months 1-3 of t-1)	tbond(month 1 of t)	hours(months 1-2 of t)	hours(months 1-3 of t)
	claims(months 1-2 of t-1)		claims(month 1 of t)	claims(months 1-2 of t)
	stprice(months 1-3 of t-1)		stprice(months 1-2 of t)	stprice(months 1-3 of t)
	tbill(months 1-3 of t-1)		tbill(months 1-2 of t)	tbill(months 1-3 of t)
	tbond(months 1-3 of t-1)		tbond(months 1-2 of t)	tbond(months 1-3 of t)

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

(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. $t+1$
1985:Q1-2011:Q3				
AR	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
AR-SV	0.999 (0.922)	1.007 (0.227)	1.007 (0.342)	1.010 (0.191)
small BMF	0.932 (0.167)	0.916 (0.102)	0.823 (0.120)	0.770 (0.091)
large BMF	0.932 (0.210)	0.872 (0.141)	0.800 (0.207)	0.770 (0.175)
small BMF, rolling	1.000 (0.989)	0.984 (0.120)	0.890 (0.064)	0.838 (0.053)
large BMF, rolling	1.002 (0.966)	0.952 (0.079)	0.903 (0.322)	0.850 (0.241)
small BMF-SV	0.936 (0.248)	0.906 (0.098)	0.815 (0.108)	0.749 (0.071)
large BMF-SV	0.925 (0.351)	0.879 (0.234)	0.816 (0.246)	0.771 (0.173)
1985:Q1-2008:Q2				
AR	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
AR-SV	1.002 (0.829)	1.005 (0.498)	1.010 (0.312)	1.010 (0.309)
small BMF	0.960 (0.050)	0.941 (0.015)	0.859 (0.018)	0.816 (0.005)
large BMF	0.979 (0.567)	0.914 (0.024)	0.892 (0.144)	0.864 (0.062)
small BMF, rolling	1.003 (0.859)	0.981 (0.196)	0.890 (0.027)	0.830 (0.004)
large BMF, rolling	0.992 (0.692)	0.950 (0.068)	0.920 (0.194)	0.867 (0.041)
small BMF-SV	0.955 (0.159)	0.926 (0.045)	0.852 (0.025)	0.799 (0.003)
large BMF-SV	1.000 (0.995)	0.941 (0.292)	0.914 (0.268)	0.870 (0.070)

Notes: See Table 1 and sections 3 and 4 for the definition of the models. The equal forecast accuracy test is described in section 5.1. The reported RMSEs reflect GDP growth defined in annualized percentage terms.

Table 3. Average log scores relative to AR-SV benchmark
(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. $t+1$
1985:Q1-2011:Q3				
AR-SV	-2.210	-2.144	-2.134	-2.123
AR	-0.245 (0.000)	-0.258 (0.000)	-0.264 (0.000)	-0.269 (0.000)
small BMF	-0.177 (0.015)	-0.151 (0.009)	-0.007 (0.916)	0.045 (0.493)
large BMF	-0.145 (0.031)	-0.091 (0.186)	0.070 (0.416)	0.094 (0.251)
small BMF, rolling	-0.185 (0.039)	-0.119 (0.032)	0.019 (0.649)	0.081 (0.065)
large BMF, rolling	-0.141 (0.020)	-0.087 (0.113)	0.027 (0.610)	0.085 (0.208)
small BMF-SV	0.018 (0.558)	0.085 (0.010)	0.195 (0.002)	0.279 (0.000)
large BMF-SV	0.127 (0.002)	0.126 (0.017)	0.182 (0.065)	0.227 (0.020)
1985:Q1-2008:Q2				
AR-SV	-2.091	-2.049	-2.049	-2.047
AR	-0.307 (0.000)	-0.312 (0.000)	-0.310 (0.000)	-0.307 (0.000)
small BMF	-0.251 (0.000)	-0.208 (0.000)	-0.055 (0.275)	-0.002 (0.962)
large BMF	-0.216 (0.000)	-0.155 (0.002)	0.003 (0.959)	0.034 (0.535)
small BMF, rolling	-0.125 (0.018)	-0.108 (0.020)	0.012 (0.796)	0.082 (0.083)
large BMF, rolling	-0.106 (0.036)	-0.067 (0.141)	0.023 (0.629)	0.083 (0.087)
small BMF-SV	0.028 (0.229)	0.079 (0.022)	0.178 (0.002)	0.259 (0.000)
large BMF-SV	0.088 (0.018)	0.103 (0.029)	0.150 (0.024)	0.182 (0.019)

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 5.1. The reported scores reflect GDP growth defined in annualized percentage terms.

Table 4. Coverage rates, nominal 70%
(p-values of correct coverage in parentheses)

forecast	month 1 qrtr. t	month 2 qrtr. t	month 3 qrtr. t	month 1 qrtr. $t+1$
1985:Q1-2011:Q3				
AR	0.925 (0.000)	0.944 (0.000)	0.944 (0.000)	0.944 (0.000)
AR-SV	0.720 (0.653)	0.692 (0.851)	0.720 (0.653)	0.729 (0.502)
small BMF	0.925 (0.000)	0.935 (0.000)	0.925 (0.000)	0.916 (0.000)
large BMF	0.925 (0.000)	0.925 (0.000)	0.869 (0.000)	0.897 (0.000)
small BMF, rolling	0.822 (0.001)	0.841 (0.000)	0.813 (0.003)	0.813 (0.003)
large BMF, rolling	0.794 (0.016)	0.841 (0.000)	0.785 (0.033)	0.766 (0.106)
small BMF-SV	0.748 (0.259)	0.729 (0.502)	0.748 (0.259)	0.757 (0.171)
large BMF-SV	0.673 (0.552)	0.673 (0.552)	0.626 (0.116)	0.720 (0.653)
1985:Q1-2008:Q2				
AR	0.947 (0.000)	0.957 (0.000)	0.968 (0.000)	0.957 (0.000)
AR-SV	0.723 (0.614)	0.691 (0.859)	0.723 (0.614)	0.745 (0.323)
small BMF	0.947 (0.000)	0.947 (0.000)	0.936 (0.000)	0.926 (0.000)
large BMF	0.947 (0.000)	0.947 (0.000)	0.872 (0.000)	0.894 (0.000)
small BMF, rolling	0.840 (0.000)	0.872 (0.000)	0.840 (0.000)	0.840 (0.000)
large BMF, rolling	0.830 (0.001)	0.862 (0.000)	0.809 (0.008)	0.798 (0.019)
small BMF-SV	0.777 (0.076)	0.745 (0.323)	0.745 (0.323)	0.755 (0.215)
large BMF-SV	0.670 (0.541)	0.702 (0.964)	0.638 (0.216)	0.713 (0.786)

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 5.1.
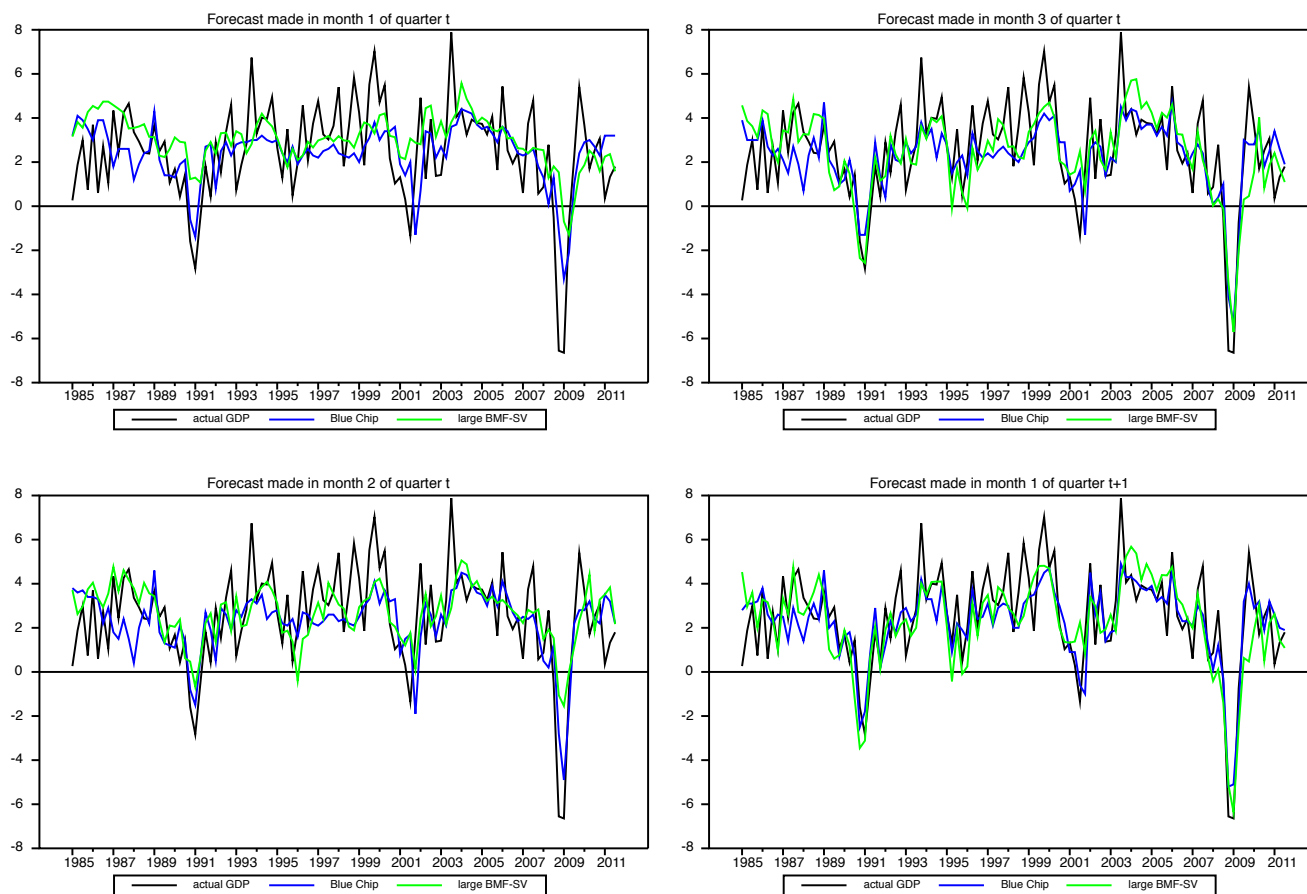
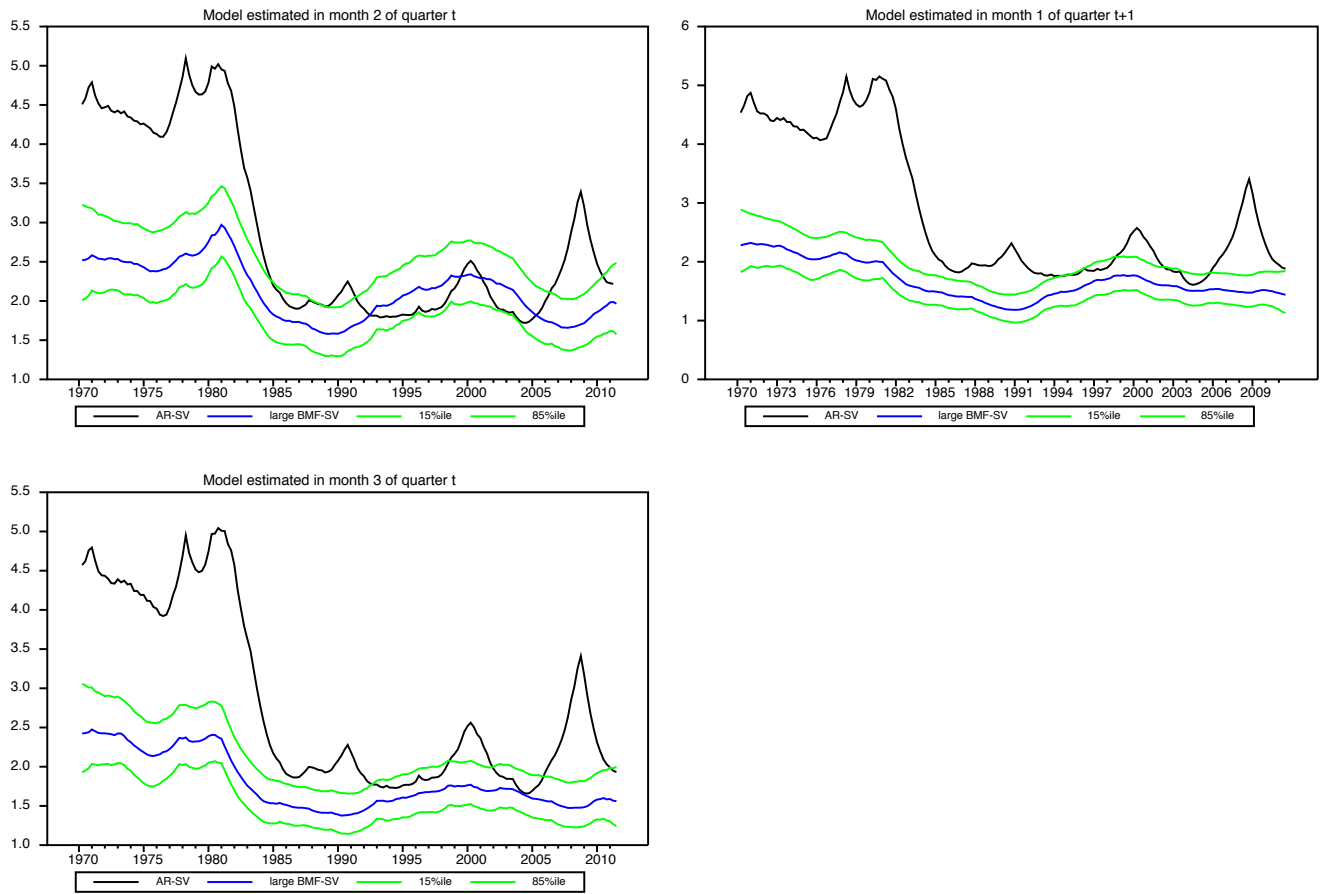Figure 1: Real-time point forecasts from Blue Chip and large BMF-SV model

Figure 2: Volatility ($\lambda_{m,t}^{0.5}$ of equation (3)) estimates from AR-SV and large BMF-SV models, last vintage of data

34

70% Interval forecasts of GDP growth, 1985:Q1-2011:Q3
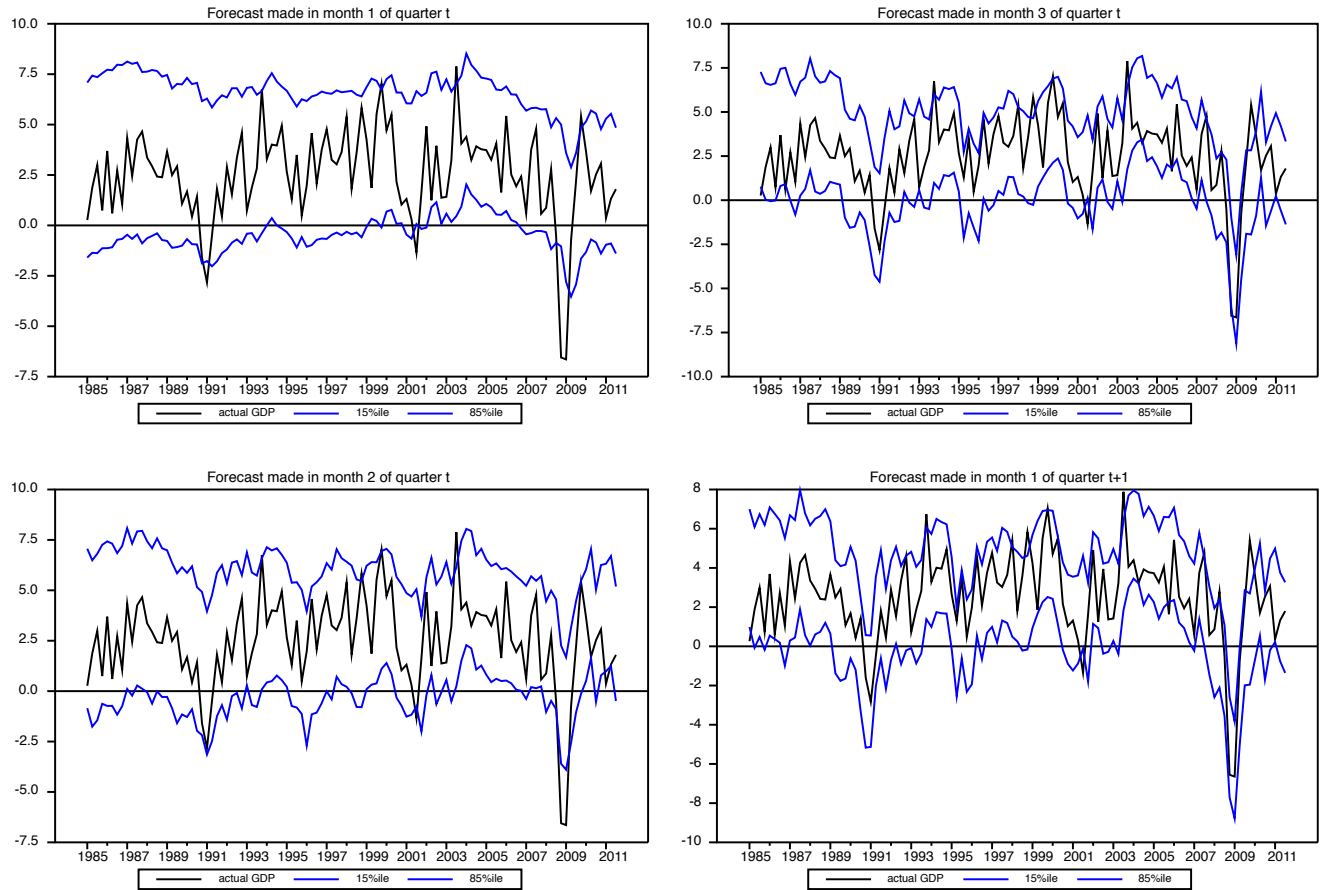### *Model: large BMF*

Figure 3: Real-time interval forecasts from large BMF model

70% Interval forecasts of GDP growth, 1985:Q1-2011:Q3
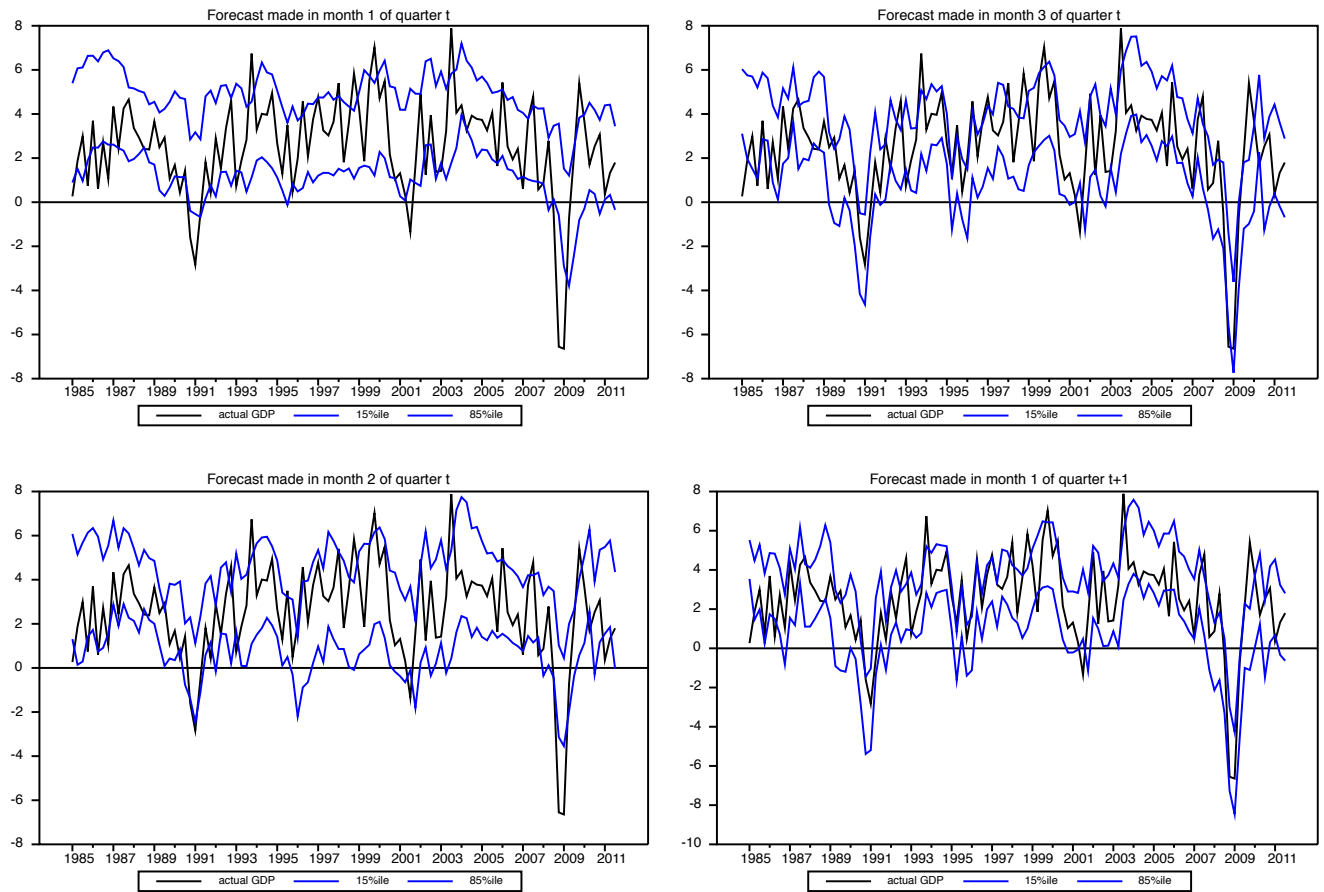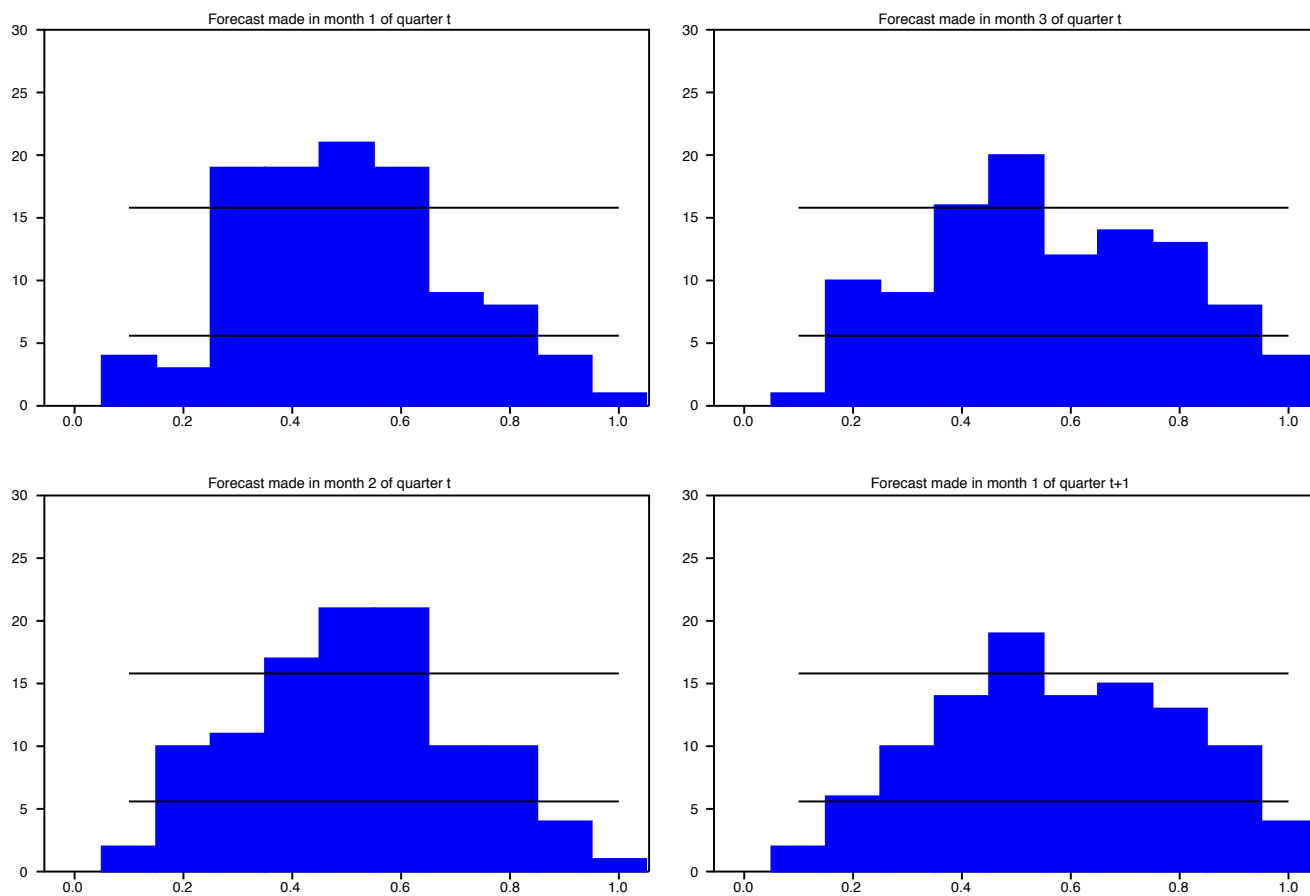## *Model: large BMF-SV*

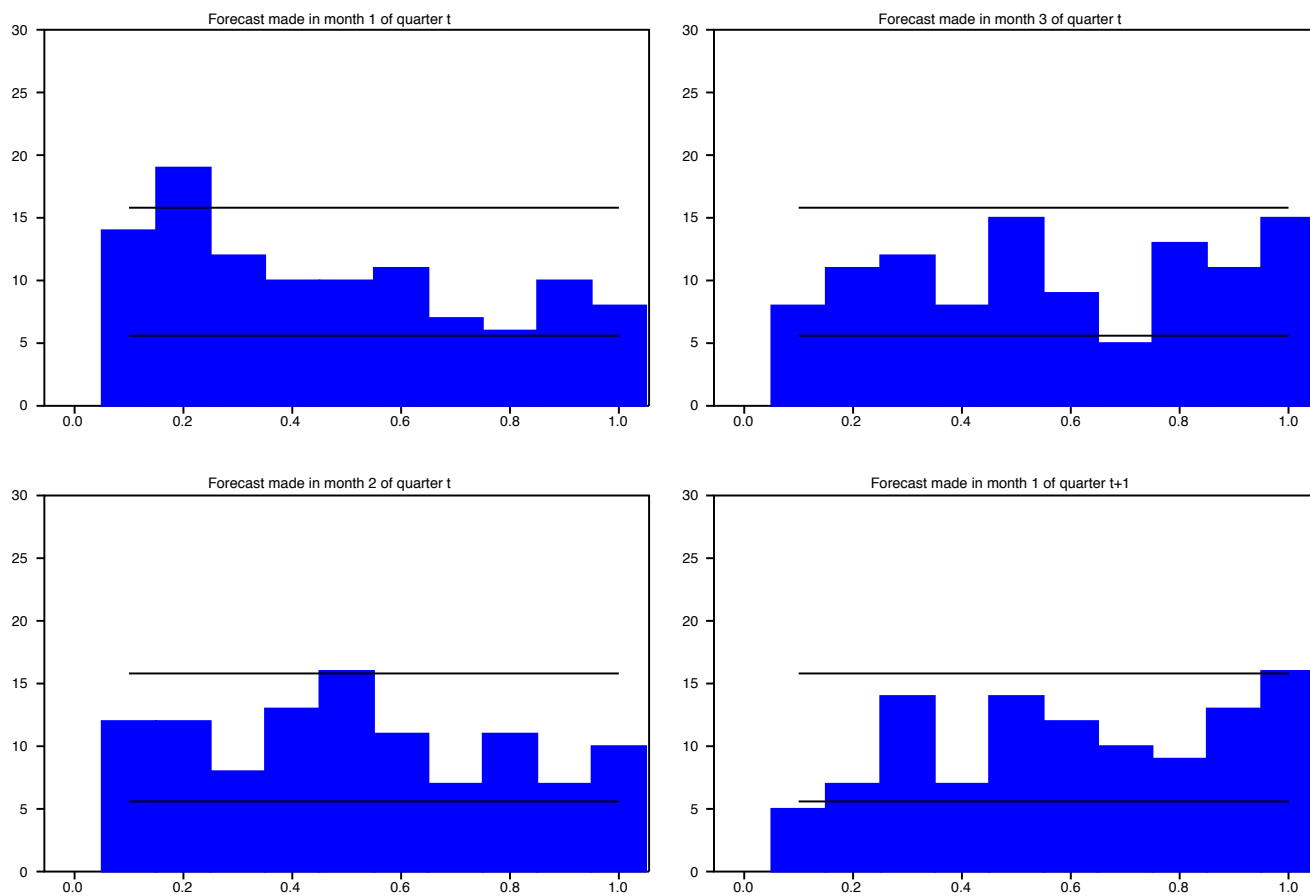Figure 4: Real-time interval forecasts from large BMF-SV model

Figure 5: PITs histograms, large BMF model

Figure 6: PITs histograms, large BMF-SV model