ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Brave, Scott A.; Butters, R. Andrew; Justiniano, Alejandro

Working Paper

Forecasting economic activity with mixed frequency Bayesian VARs

Working Paper, No. 2016-05

Provided in Cooperation with:
Federal Reserve Bank of Chicago

Suggested Citation: Brave, Scott A.; Butters, R. Andrew; Justiniano, Alejandro (2016) : Forecasting economic activity with mixed frequency Bayesian VARs, Working Paper, No. 2016-05, Federal Reserve Bank of Chicago, Chicago, IL

This Version is available at:
https://hdl.handle.net/10419/172921

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Federal Reserve Bank of Chicago

Forecasting Economic Activity with Mixed Frequency Bayesian VARs

Scott A. Brave, R. Andrew Butters, and Alejandro Justiniano

May 2016

WP 2016-05

Forecasting Economic Activity with Mixed Frequency Bayesian VARs

Scott A. Brave
Federal Reserve Bank of Chicago

R. Andrew Butters
Indiana University

Alejandro Justiniano*
Federal Reserve Bank of Chicago and Paris School of Economics

May 20, 2016

Abstract

Mixed frequency Bayesian vector autoregressions (MF-BVARs) allow forecasters to incorporate a large number of mixed frequency indicators into forecasts of economic activity. This paper evaluates the forecast performance of MF-BVARs relative to surveys of professional forecasters and investigates the influence of certain specification choices on this performance. We leverage a novel real-time dataset to conduct an out-of-sample forecasting exercise for U.S. real Gross Domestic Product (GDP). MF-BVARs are shown to provide an attractive alternative to surveys of professional forecasters for forecasting GDP growth. However, certain specification choices such as model size and prior selection can affect their relative performance.

JEL Codes: C32, C53, E37
Keywords: mixed frequency, Bayesian VAR, real-time data, nowcasting

*We thank Gianni Amisano, Todd Clark, Giorgio Primiceri, Barbara Rossi, Saeed Zaman, seminar participants at the Advances in Applied Macro Finance and Forecasting Conference, the Euroarea Business Cycle Network Conference at the Norges Bank, the CIRANO Real-Time Workshop, the Federal Reserve Banks of Chicago and Cleveland and the Federal Reserve Board of Governors, and particularly Domenico Giannone for helpful comments. We would also like to thank David Kelley for superb research assistance. The views expressed herein are the authors' and do not necessarily reflect the views of the Federal Reserve Bank of Chicago or the Federal Reserve System. Please address correspondence to: Alejandro Justiniano, Economic Research, Federal Reserve Bank of Chicago, 230 S. La Salle St., Chicago, IL, 60604. E-mail: ajustiniano@frbchi.org. Telephone: (+1)312-322-5900.

1 Introduction

Private sector analysts and policymakers share a need for timely forecasts of economic activity. Typically, these forecasts must blend information from a wide array of sources and collected at different frequencies in order to be both encapsulating and reflective of the most recent events. To better equip forecasters for this challenge, a considerable amount of research has been conducted on developing methods that are able to handle both (i) data observed at different frequencies, as well as (ii) a large number of time series.

A recent addition to this suite of methods is the mixed frequency Bayesian vector autoregression (MF-BVAR) of Schorfheide and Song [2015].[1] Due to their infancy in the forecasting literature, however, much less is known about the predictive performance of MF-BVARs compared to their traditional, that is single frequency counterparts.[2] This paper closes this information gap on two particular dimensions. First, we formally evaluate the real-time performance of MF-BVARs relative to surveys of professional forecasters.[3] Second, we provide an in-depth investigation of how predictive ability is shaped by a set of specification choices

[1] The use of Bayesian methods in forecasting economic activity has a celebrated tradition dating back to Doan et al. [1984] and Litterman [1986], who first documented that Bayesian shrinkage could improve upon the forecast accuracy of a small vector autoregression (VAR). Several others have also documented the superior forecast performance of Bayesian methods over classical factor methods (e.g. Carriero et al. [2011] and Koop [2013]). Factor models, however, require all variables to be stationary. Consequently, it is not always clear what drives these results when the BVAR contains data in levels. Koop [2013] showed that the performance gains of BVARs relative to factor methods holds with a model estimated on stationary variables.

[2] Chauvet and Potter [2013] provide a comprehensive survey of the relative forecasting performance for GDP of other methodologies, including dynamic factor models, Markov switching models, vector autoregressions (VARs), and other more traditional forecasting methods.

[3] Schorfheide and Song [2015] evaluate the relative forecasting performance of a MF-BVAR and the Greenbook forecasts prepared by the staff of the Federal Reserve Board of Governors for FOMC meetings. Because Greenbook forecasts are only publicly available after a five-year delay, their evaluation sample ends in December of 2004, and consequently does not include the Great Recession. They find that, for GDP, the Greenbook forecasts are better than the MF-BVAR at the nowcast and one-quarter ahead horizon, while the MF-BVAR outperforms the Greenbook forecasts at the 2-4 quarters ahead horizons. Contrary to us, they do not formally test the statistical significance of these results, which in their case are based on a considerably smaller number of observations. Similarly, McCracken et al. [2015] compare the forecasts from a blocked Bayesian VAR modeled at the quarterly frequency, but incorporating monthly indicators, to the Survey of Professional Forecasters.

inherent in the implementation of MF-BVARs including model size, choice of prior, data transformations, and lag length.

Our analysis focuses on evaluating nowcasts and medium-term forecasts (up to four quarters ahead) of growth in U.S. real Gross Domestic Product (GDP), where the relevant set of economic indicators (series) is potentially large, observed at different frequencies, and subject to release patterns that are staggered in real-time. Furthermore, we draw comparisons on both a quarter-over-quarter and year-over-year basis in order to disentangle the ability of the MF-BVAR to predict both short and medium-run economic activity.

To gauge the real-time performance of the MF-BVAR, we compare its forecasts to the Blue Chip Consensus and the Survey of Professional Forecasters over the period from the third quarter of 2004 to the third quarter of 2014. The comparison to surveys is greatly enhanced by leveraging a novel real-time dataset: the proprietary archives of the Chicago Fed National Activity Index. This provides us a unique opportunity to replicate the real-time information flow of a number of U.S. macroeconomic indicators.[4] Most of these series serve as critical inputs into the construction of the U.S. National and Income and Product Accounts, the source of GDP, and are commonly used by professional forecasters to inform their predictions. Hence, by leveraging this dataset we are able to assess the real-time informational content of a broader set of indicators for forecasting economic activity than is common in the literature, with some series being used in real-time predictions for the first time.

Our analysis delivers three main findings. *First, a moderately sized MF-BVAR provides a favorable alternative to surveys at medium-term forecast horizons.* More specifically, a model including 21 indicators (in levels), delivers forecasts whose performance for the quarterly

[4]Information on the Chicago Fed National Activity Index can be found at Federal Reserve Bank of Chicago [2015]. Brave and Butters [2010, 2014] discuss the use of real-time data for the index in nowcasting U.S. real GDP growth and inflation. McCracken and Ng [2015] summarize a similar real-time macroeconomic database (FRED-MD) maintained by the Federal Reserve Bank of St. Louis. As of December 2015, FRED-MD only provides (monthly) vintage data for 2015 with no disclosed plan to make earlier vintages available in the future.

growth rate of U.S. GDP is either comparable or superior to the surveys two to four quarters ahead. In terms of root mean-squared forecast errors (RMSFE), the MF-BVAR outperforms both surveys at these forecast horizons by roughly 10 to 15 percent, with Diebold-Mariano tests rejecting the null of equal forecast accuracy in most cases for predictions three to four quarters out. Giacomini-White tests of equal conditional predictive ability confirm this finding, and further indicate that the forecast gains with the MF-BVAR were particularly noticeable during the Great Recession, but were not confined to this episode.

Second, both surveys generally record similar or slightly lower RMSFEs for near-term predictions. This is particularly true for the nowcast, with both surveys achieving gains in point forecast accuracy compared to the MF-BVAR, but that are generally not statistically significant. The superior relative performance of surveys in the nowcast is shown to be driven to some extent by issues regarding the timing of information in our real-time dataset. Indeed, we err on the side of caution against providing the model with more information than would have been available to survey participants, often putting it at a distinct disadvantage. Specifically, for some series (depending on their release date within the month), we provide only lagged information to the model, as opposed to the most recent release. This baseline timing assumption is appropriate for the Blue Chip Consensus survey, but puts the model at a distinct disadvantage relative to the Survey of Professional Forecasters given that this survey is conducted later within the month. In fact, in an alternative specification that aligns the information set more in accordance with this survey and what a policymaker or private analyst would face in the interim period between Blue Chip surveys (but before GDP data are available) the MF-BVAR performs considerably better at near-term forecasting. While this alternative timing provides a fairer comparison to the Survey of Professional Forecasters, we view this exercise mainly as confirming the value of the MF-BVAR in processing the real-time flow of information.

Our third main finding is that the MF-BVAR's forecast performance is sensitive to some specification choices, in particular, model size, priors, and working in levels as opposed to growth rates. More specifically, parsing through the variants considered in our analysis we offer the following observations regarding the role of specification choices:

1. **Model size**: Augmenting the dataset with monthly indicators that cover different real expenditure components of GDP beyond those variables commonly used in the literature drastically enhances forecast performance. In terms of RMSFE, the gains range from roughly 10 to 20 percent for the quarterly growth rate of GDP at forecast horizons 0 to 4 quarters ahead, and are even larger for four-quarter changes. Further expanding the model's size by adding quarterly real expenditure components of GDP results in small improvements in forecast accuracy.

2. **Choice of priors**: We consider conjugate Normal-Inverse Wishart priors governed by a small set of hyperparameters. Using the marginal likelihood to select those hyperparameters delivers RMSFEs that are roughly 6 to 16 percent lower relative to using default settings for the priors that are common in the literature.

3. **Levels vs. Growth Rates**: Specifications in levels outperform those in first differences with gains in forecast performance that range from 8 to 14 percent for predictions one quarter ahead and beyond.

4. **Lags**: RMSFEs are fairly insensitive to the choice of lag order in the range of three to seven monthly lags, provided the shrinkage parameters are selected with the marginal likelihood for each lag length.

Overall, our analysis suggests that MF-BVARs can be a valuable tool for the real-time forecasting of U.S. economic activity, even when compared to surveys of professional forecasters.

Previous studies have identified professional surveys as a formidable benchmark with which to judge a model's predictive ability.[5] Therefore, the gains in forecast performance we document relative to them offer confirmation of the value of MF-BVARs. *This comparison to surveys combined with the in-depth analysis of how predictive ability is affected by different specification choices are the two primary contributions of our analysis. These contributions complement the evidence presented by Schorfheide and Song [2015] and provide a further understanding of how MF-BVARs fare in predicting economic activity in real time.*

Additionally, we also confirm Schorfheide and Song [2015]'s finding that in a BVAR context a mixed frequency as opposed to quarterly set-up leads to marked improvements in forecast performance for growth in U.S. GDP. These authors note that forecast gains of the MF-BVAR relative to a traditional BVAR disappear beyond two quarters. Instead, we point out that improvements in predictive accuracy from working in mixed frequency remain large and statistically significant even four quarters out when the focus is on year-over-year (as opposed to quarter-over-quarter) growth rate comparisons, which is commonly the case among policymakers. In addition, our comparison is done with a more general information structure than they considered and involves conditional forecasting due to the staggered nature of data releases.

Our findings also parallel the performance gains from mixed frequency data found in other frameworks as well, including dynamic factor models (Mariano and Murasawa [2003, 2010], Aruoba et al. [2009]), and Mixed Data Sampling (MIDAS) regressions first proposed by Ghysels et al. [2004] and extended to the VAR case by Foroni et al. [2013] and Ghysels [2016]. Similarly, our work contributes to the growing literature investigating the role of specification choices on forecasting performance for traditional, that is single frequency, BVARs. For example, investigations on the effect of model size (Bańbura et al. [2010], Koop [2013]), the

[5] For instance, Chauvet and Potter [2013] find that the Blue Chip Consensus forecasts for U.S. real GDP growth outperform all of the models they consider across the 1 and 2 quarter ahead forecast horizons.

choice of prior (Koop [2013], Carriero et al. [2015], Giannone et al. [2015]), and specification choices more generally (Carriero et al. [2015]) that have been conducted for the traditional BVAR setting. Interestingly, our work for the mixed frequency case finds a somewhat larger sensitivity to some specification choices than what is usually reported in these studies.

The rest of the paper is organized as follows. Section 2 briefly describes our framework for estimating MF-BVARs and discusses data-based methods for eliciting priors. The data used in the analysis and the associated timing of the real-time information flow is then discussed in section 3 along with our method of forecast evaluation. Forecasting results of the MF-BVAR relative to surveys as well alternative specifications are presented in section 4 in addition to the comparison between MF and quarterly BVARs. Finally, section 5 concludes and offers suggestions for future work.

2 Methodology

This section briefly outlines the key elements of our MF-BVAR. In order to accommodate the mixed frequency nature of the dataset, the model is cast within a state-space system. Inference is conducted using a Gibbs sampling procedure to handle the latent values of low frequency variables. Moreover, we rely on Bayesian shrinkage that puts guidance both on the individual dynamics of each series and on the overall co-movement among all the series in the system. Finally, the resulting priors are driven by a low dimensional vector of hyperparameters selected with the marginal data density. The presentation here is purposefully succinct, with ample references to more thorough treatments and some explicit details relegated to the appendix.[6]

[6] For a more comprehensive treatment of state-space methods, see Durbin and Koopman [2012]. For more comprehensive treatments of BVARs, see Karlsson [2013] and Del Negro and Schorfheide [2011].

2.1 State-Space Representation of a MF-BVAR

A state-space framework is a natural representation for an MF-BVAR given the ease with which it can accommodate missing observations and produce forecasts (see for instance Aruoba et al. [2009] and Brave and Butters [2012]). In our application, we use a mix of both quarterly and monthly time series in the state-space, although the methods described here are general enough to handle any type of mixed frequency setting.

Consider an n-dimensional vector y_t of macroeconomic time series. In general, all the variables in y_t will not be observed in every period. In our setting, an observation could be missing for one of two reasons. First, and most prevalent, any series that is observed at a lower frequency (quarterly) than the base frequency (monthly) will have missing values.[7] Second, given publication delays, even series observed at the base frequency may have missing observations towards the end of the sample. This second type of missing observation is readily handled in state-space models (Durbin and Koopman [2012]).

To accommodate the mixed frequency nature of the time series in y_t, we adopt the convention of Harvey [1989] and model the underlying higher (monthly) frequency movement of each series, stacked in the vector x_t. To match the realized values of those series observed at a lower (quarterly) frequency, the corresponding elements of this vector are then aggregated with the appropriate accumulator depending on each series' temporal aggregation properties (see section A.1 in the appendix). For instance, the observed level of GDP corresponds to the three-month average of the corresponding element of x_t.

The vector x_t is assumed to follow a vector autoregression of order p, given by

$$x_t = c + \Phi_1 x_{t-1} + ... + \Phi_p x_{t-p} + \epsilon_t; \quad \epsilon_t \sim i.i.d.N(0, \Sigma). \tag{1}$$

[7] We adopt the convention of placing quarterly observations in the third month of each quarter.

where each Φ_l is an n-dimensional square matrix containing the coefficients associated with lag l. This (monthly) VAR can be written in companion form and combined with a measurement equation for y_t to deliver a state-space model given by

$$y_t = Z_t s_t \tag{2}$$

$$s_t = C_t + T_t s_{t-1} + R_t \epsilon_t. \tag{3}$$

The vector of observables, y_t, is defined as above, while the state vector, s_t, is given by

$$s_t' = \left[x_t',\ \ldots,\ x_{t-p}',\ \Psi_t'\right],$$

which includes both lags of the time series at the base (monthly) frequency and Ψ_t, a vector of accumulators. Each accumulator maintains the appropriate combination of current and past x_t's to preserve the temporal aggregation of the lower (quarterly) frequency time series.

As for the system matrices, the initial n rows of each transition matrix T_t concatenate the coefficients associated with each lag $\Phi = [\Phi_1, \Phi_2, ..., \Phi_p]$. Notice that even if the VAR parameters are assumed to be time-invariant (as in our empirical analysis), the state-space system matrices are indexed by t due to the deterministic time variation required in calculating the accumulators, Ψ_t. The remaining entries of this matrix correspond to ones and zeros to preserve the lag structure, or some scaled replication of the coefficients to build an accumulator. The VAR intercepts sit at the top of C_t, while scaled versions of intercepts are in rows associated with each accumulator. The rest of C_t has zeros. Finally, each R_t corresponds to the natural selection matrix augmented to accommodate the additional accumulator variables in the state.[8]

[8] Brave et al. [2015] provide a more complete description of the transformation from the standard VAR

The matrix Z_t is comprised solely of selection rows made up of zeros and ones. Its row dimension will vary over time due to the changing dimensionality of the observables. In particular, every three periods when the quarterly time series are observed it will have the full n selection rows. For the remaining months in which only monthly time series are observed, a subset of these selection rows will be included. Furthermore, towards the end of the sample not all of the monthly indicators will be available, depending on their release schedule; and, hence, a further subset of the selection rows will not be used (see section A.1 in the appendix).

2.2 Gibbs Sampling Procedure

With the model cast in a state-space framework we wish to estimate the full set of parameters and latent states given by $\Theta = \{\Phi, c, \Sigma, \{x_t, \Psi_t\}_{t=1}^T\}$. Denoting the history of data in the estimation sample through time $t \leq T$ as $Y_{1:t}$, inference on Θ concerns the VAR parameters, $\{\Phi, c, \Sigma\}$, the latent monthly variables $\{x_t\}_{t=1}^T$ (of the quarterly time series as well as any missing monthly variables), and the accumulators $\{\Psi\}_{t=1}^T$ conditional on $Y_{1:T}$. To conduct inference, Schorfheide and Song [2015] propose a two-block Gibbs sampler that, conditional on a pre-sample $Y_{-p+1:0}$ used to initialize the lags, generates draws from the conditional posterior distributions:

$$P(\Phi, c, \Sigma | X_{0:T}, Y_{-p+1:T}) \tag{4}$$

and

$$P(X_{0:T} | Y_{-p+1:T}, \Phi, c, \Sigma), \tag{5}$$

where with a slight abuse of notation we stack $\{x_t, \Psi_t\}_{t=1}^T$ into the matrix $X_{0:T}$.

system to the augmented state-space system that incorporates accumulators.

The first density, given in (4), is the posterior of the VAR parameters conditional on all data and the latent variables. With a suitable choice of priors, sampling from this distribution reduces to taking a draw from a straightforward multivariate regression. The second density, given in (5), corresponds to the smoothed estimates of the latent variables. A draw from this distribution is obtained via the simulation smoother of Durbin and Koopman [2012]. Hence, the estimation of the MF-BVAR iterates between taking draws from these two conditional posterior distributions.

2.3 Shrinkage through Dummy Observations

To overcome the curse of dimensionality in (4), we impose prior information regarding the parameters of the VAR.[9] Generally speaking, the priors we use combine a slightly modified version of the well-known Minnesota prior (Litterman [1986]) with a set of priors that guide the sum of autoregressive lags as well as the co-persistence of the variables in the model (see Del Negro and Schorfheide [2011] or Karlsson [2013] for detailed treatments and Carriero et al. [2015] for an analysis of their effects on forecast accuracy in the case of single frequency VARs).

The four hyperparameters that govern these priors are collected in the vector Λ. Let $\Phi_l(i,j)$ denote the coefficient of the l-th lag of the j-th variable in the i-th variable's equation given by equation (1). The first two entries of Λ specify the prior beliefs summarized by

$$E[\Phi_l(i,j)] = \begin{cases} \delta_i & j=i, l=1 \\ 0 & \text{otherwise} \end{cases} \qquad V[\Phi_l(i,j)] = \begin{cases} \left(\frac{1}{\lambda_1 l^{\lambda_2}}\right)^2 & j=i \\ \left(\frac{1}{\lambda_1 l^{\lambda_2}} \frac{\sigma_i}{\sigma_j}\right)^2 & \text{otherwise} \end{cases} \tag{6}$$

which shrink the VAR system toward independent random walks when δ_i=1 or white noise

[9] The benchmark model includes 21 variables and 3 lags and consequently requires 64 coefficients to be estimated for each variable on a maximum sample size of 475 monthly observations.

when δ_i=0.[10]

The overall tightness of the prior is controlled by λ_1, and is subsequently referred to as the **tightness**. As $\lambda_1 \to \infty$, the posterior distribution is dominated by the prior; while, conversely, as $\lambda_1 \to 0$ the posterior coincides with the OLS estimates of the VAR. The second element of the prior is λ_2, the **decay** hyperparameter, which governs the rate at which coefficients at distant lags are shrunk further toward zero. For the intercepts, c, we adopt a fairly diffuse prior as is customary in Bayesian VARs.

This specification differs from Litterman [1986] in two respects. First, parameters on "own" versus "other" variable lags are treated symmetrically, since this is required to maintain conjugacy with a multivariate Normal-Inverse Wishart prior.[11] .Second, rather than assuming that Σ is known, the prior is chosen such that, in expectation, it coincides with the variances of residuals obtained from individual AR regressions as is customary in the literature. We adopt the approach of Bańbura et al. [2010] and Del Negro and Schorfheide [2011] by implementing this prior using a set of artificial observations that are appended to the data.

Forecast performance has been shown to improve with two additional priors concerning the persistence and co-persistence of the variables in the VAR. These additional priors are designed to prevent initial transients and deterministic components from explaining an implausible share of the long-run variability in the system (Sims and Zha [1998], Sims [2000]). The first form of shrinkage is usually known as the **sum of coefficients** prior, and expresses the belief that the sum of own lag autoregressive coefficients for each individual variable should be one. This is governed by λ_3, with larger values implying (as above) a tighter prior. The

[10] We allow the centers, δ_i, to differ from 0 or 1 for some variables in levels who are persistent but would not seem to be described by a random walk with drift. More precisely, we include the center δ_i's for those variables among the elements of Λ that are selected via the marginal data density procedure described in the next section. However, this delivered essentially identical results to just setting the δ_i's to either 1 or 0. Results for these alternative specifications are available from the authors upon request.

[11] More specifically, by treating variables symmetrically the variance of the prior has a Kroenecker product structure with the innovation variance Σ.

second form of shrinkage is known as the **co-persistence** prior and reflects beliefs that if the sum of all VAR coefficients is close to an identity matrix, then the intercepts should be small (or conversely, if the intercepts are not close to zero, then the VAR is stationary). The strength with which this prior is imposed is increasing in λ_4. In both cases, a hyperparameter set to zero corresponds to the exclusion of that prior from the system, while approaching infinity corresponds to a restriction of the system that strictly adheres to the prior. The set of dummy observations that implements the four forms of shrinkage we consider are described in Appendix A.2.

2.4 Selection of Hyperparameters via the Marginal Data Density

The hyperparameters controlling the priors are chosen to maximize the marginal likelihood, $P(Y_{0:T}|\Lambda)$, such that

$$\Lambda^\star = \text{argmax } P(Y_{0:T}|\Lambda),$$

which is closely tied to a model's one-step ahead out-of-sample performance (Geweke [2001]). Under conjugate, or flat priors, the marginal data density, $P(Y_{0:T}|\Lambda)$, is available analytically in the case of single frequency BVARs.[12] This allows using optimization routines to find $\Lambda^\star$ quickly and efficiently (Giannone et al. [2015]).

However, with mixed frequency data and, hence, latent variables, the marginal data density must be approximated to account for the unknown states and the restrictions imposed on them by temporal aggregation. This can be done using the output of the Gibbs sampler and the modified harmonic mean estimator [Schorfheide and Song, 2015]. Unfortunately, computational considerations then prevent using optimization techniques and force the use of sparse

[12] See equation (7.15) in Del Negro and Schorfheide [2011].

grids over which to evaluate each possible combination of the hyperparameters.

Since part of our investigation concerns the sensitivity of forecast performance to hyperparameters, we gauge the general patterns of the marginal likelihood by way of an approximation. More specifically, the latent quarterly variables are first recursively interpolated using a procedure akin to the general augmented distributed lag (ADL) framework of Proietti [2006] that is described in Appendix A.3. This is done using related monthly series not included in the VAR for each quarterly variable.[13] Conditional on these interpolated series, and combined with the monthly indicators, one proceeds to optimize the marginal likelihood of this generated dataset–which is known in closed form–using numerical methods. We found this approximation to be effective in characterizing the contours of the marginal likelihood to different hyperparameters, as shown in Appendix A.4, and to detect possible identification issues. Once an informed grid is set up based on this initial exploration, the Gibbs sampler is run for all possible combinations of the grid elements. The modified harmonic mean is then used to infer the correct $P(Y_{0:T}|\Lambda)$ for each combination and the one with the highest marginal data density is selected.[14]

3 Data, Forecast Timing, and Evaluation

In this section, we briefly describe the data used to estimate our MF-BVAR and the methods used to evaluate its forecasts of GDP growth. Below, we briefly outline (i) the salient features of the vintage data used in estimation, (ii) our forecast timing convention across vintages in comparison to the Blue Chip Consensus (BCC) and Survey of Professional Forecasters (SPF) surveys, as well as (iii) how we make statistical evaluations of the forecast performance of our

[13] In the case of GDP, we have experimented with Manufacturing Industrial Production or (nominal) Personal Income, neither of which is collinear with the indicators included in the VAR.

[14] We have checked that posterior contours of the correct density resemble, qualitatively, those obtained with the interpolation procedure, but differ, as expected, in magnitudes. See section A.4 for a more detailed discussion.

MF-BVAR relative to these surveys.

3.1 Real-time Data

To evaluate forecast performance, we use real-time vintage data for each time series. To construct this dataset, we rely principally on the Federal Reserve Bank of Chicago's proprietary archives of the Chicago Fed National Activity Index (CFNAI), augmented with vintage data from Haver Analytics and the St. Louis and Philadelphia Federal Reserve Banks' ALFRED database and Real-time Dataset for Macroeconomists, respectively.[15] The broad scope of our real-time dataset allows us to estimate models of varying size, including specifications that encompass a large number of monthly series indicators (or series) that are commonly used by professional forecasters when predicting U.S. GDP. As such, it represents an ideal, and previously untapped, dataset with which to evaluate how MF-BVARs compare with surveys of professional forecasters in predicting GDP growth.

The archives are stored at the time of production of the CFNAI, near the middle of each month. Real-time (unrevised) vintages are available on a regular basis going back to 2004 and contain monthly time series for 85 U.S. macroeconomic indicators starting in 1967. The indicators included in the CFNAI archives span measures of production and income; employment, unemployment, and hours; personal consumption and housing; and sales, orders, and inventories.[16] This list encompasses indicators found in the literature on the real-time measurement of U.S. business conditions (e.g. Aruoba et al. [2009]), as well as many of those that appear in the U.S. Conference Board's Business Cycle Indicators. Moreover, several of these indicators serve as critical inputs into the construction of the U.S. National Income and Product Accounts, the source of GDP.

[15] Meanwhile, survey forecasts were obtained from the Haver Analytics BLUECHIP and SURVEY databases, respectively.

[16] For more detail on the data series included in the CFNAI, see the background information available at Federal Reserve Bank of Chicago [2015].

Table 1 summarizes the 21 indicators used in our benchmark model. There are 15 monthly indicators, including series capturing production (industrial production and capacity utilization), employment (hours worked), and consumer spending (personal consumption expenditures and retail sales). There are five quarterly indicators in addition to GDP, each capturing one of its components (i.e. Business Fixed Investment, Personal Consumption Expenditures on Nondurable Goods, Exports, Imports, and Government Consumption Expenditures and Gross Investment).[17] For specifications in levels, variables are transformed to logs unless they are already expressed as percentage rates, in which case they are divided by one hundred to retain a comparable scale. In contrast, for specifications in growth rates the transformation used is one hundred times the log difference or the difference in percentage rates.[18]

3.2 Forecast Timing: Baseline and Alternative

To keep track of forecast timing, we label forecast origins as R1, R2, or R3 according to the last available GDP release (i.e. first, second, or third release as labeled by the Bureau of Economic Analysis) at the time a forecast was made. This convention best facilitates keeping track of the information set available to professional forecasters. The first release of any quarter's GDP comes out at the very end of the first month following the end of the quarter, the second release at the end of the second month, and so on. For example, the first release of first quarter's GDP is published at the end of April, the second release in May, and so on.

To clarify the labeling of forecast origins and timing of surveys it is instructive to detail the GDP release and forecast (nowcast) timing using second quarter GDP as an example. At the end of April the first release of the previous quarter's GDP (Q1) is published, thus making first quarter GDP information available to participants in the May Blue Chip survey,

[17] We exclude quarterly Residential Investment and Changes in the Valuation of Inventories to avoid a multicollinearity problem in our benchmark specification.

[18] A data appendix describing the construction of each variable and the source of its vintage data is available from the authors upon request.

which is always conducted on the first two business days of the month. We label this forecast origin R_1 and proceed to generate the first nowcast for second quarter GDP and projections for horizons beyond. The second release of first quarter GDP is published at the end of May, making it available for respondents of the June Survey, and indexes forecast origin R_2. This is the jumping point for our second nowcast of second quarter GDP. Our third and final nowcast at forecast origin R_3 corresponds to the July survey and includes the third release of first quarter GDP.[19] The same pattern applies to the forecast origin and nowcasts for other quarters and is further summarized in figure 8 of the appendix.

Given the vagaries of the release schedules of the monthly indicators in our dataset, it is not entirely clear what the information set of the Blue Chip forecasters includes at the time of each survey. This ambiguity predominantly involves the monthly indicators typically released on the first two business days of the month when the survey is conducted. To err on the side of caution, we adopt a *Baseline Timing* assumption in which for series released after the first two business days of the month we use only "lagged vintage" data. That is for these indicators only the previously available vintage (e.g. the previous month's release) is in our information set rather than the most recent one. Following this *Baseline Timing* assumption, industrial production, capacity utilization, the inventory-sales ratio, retail sales, and hours worked use lagged vintage data to avoid giving the MF-BVAR more information than the Blue Chip forecasters (see third column of table 1). However, with this *Baseline Timing* assumption, we are potentially handicapping the models, relative to the Blue Chip forecasters, not only by limiting the choice of variables (due to the requirement of real-time availability), but also by restricting the timeliness of the content for some of the series used in the model. For instance, we cannot rule out that in some cases the Blue Chip forecasters may

[19] Forecast origin R3 is the first month of the "next" quarter in calendar time (e.g. July is the first month of the third quarter). Hence, the "nowcast" for GDP in this instance should more accurately be described as a backcast, while the one-quarter ahead forecast might be more reflective of a nowcast.

have also had the Employment Situation report in hand at the time they made their forecast, such that the current month's release of hours worked would also be in their information set.

This informational handicap is more evident when compared to the Survey of Professional Forecasters (SPF), which has similar participants to the Blue Chip survey but is conducted only once per quarter roughly in the middle of the second month of each quarter. The *Baseline Timing* assumption puts our MF-BVAR at a clear disadvantage relative to the SPF, since respondents to this survey have access to the latest Employment Situation report, hence hours worked, and likely several of the remaining monthly indicators that are vintage-lagged as well. As such, the MF-BVAR's performance results against the SPF under the *Baseline Timing* assumption should be viewed as quite conservative. Furthermore, these results also serve as a useful robustness check against having endowed the model with more timely information than what was likely available to the Blue Chip survey participants as well.

For comparison, we also present results which use all of the data available roughly in the middle of each month (at the time of the construction of the CFNAI). Under this *Alternative Timing* assumption, we drop the approach of using lagged vintage data for the series typically released between the beginning and middle of the month. This *Alternative Timing* assumption serves two purposes. First, it provides a fairer comparison of the MF-BVAR results to the SPF. Second, although not as informative of a comparison of the model's forecast performance relative to the Blue Chip survey, it helps to gauge how additional information within the month helps with predictive accuracy. That is, because the results of the Blue Chip survey are made publicly available only once each month, and only for a fee, one could view this comparison as reflecting the broader value of MF-BVARs in forming expectations of macroeconomic activity consistent with the real-time data flow through the middle of each month.

3.3 Forecast Evaluations

For the purpose of evaluation, we judge both the survey and MF-BVAR forecasts against the third real-time release of GDP.[20] Predicted growth rates are obtained at each iteration of the Gibbs simulator by generating trajectories and forecasts recursively. That is, for each parameter draw we simulate a history of states and, when computing forecasts, also generate shocks from the current estimate of the error variance. Under our *Baseline Timing* assumption, none of the indicators for the current month are observed, with some even having missing observations for the last three periods due to the variation in publication lags across the series in the model (see the last column of table 1 for a summary of these publication lags).[21]

Quarter-over-quarter and year-over-year (i.e. four-quarter) growth rates of GDP are obtained by combining the monthly trajectories into a quarterly forecast with the appropriate combination of the accumulators. Based on the moments (mean, median, standard deviations) of these draws, the corresponding prediction errors for each vintage and horizon are then constructed. Results in the next section are reported by forecast horizon using the median prediction errors of the model and averaging across forecast origins (R1, R2 or R3, as described in section 3.2).

Forecast accuracy is assessed by root mean-squared forecast errors,

$$RMSFE_m^h = \sqrt{\frac{1}{N_h}\sum_{v=1}^{N_h}(\Delta y^{q,f}_{T_v+h|T_v} - \Delta y^{q,o}_{T_v+h})^2},$$

[20] Several alternative releases exist in which to judge the forecasts. For example, Schorfheide and Song [2015] report their forecasting results relative to the final vintage of GDP in their sample. We chose to report the results relative to the third release as we feel that it more closely aligns with professional forecasters' objectives. Results for alternative releases are available upon request.

[21] For series that have missing observations toward the end of the estimation sample, the expected value of the shocks for these indicators conditional on all the data from the estimation sample is not equal to zero. We simulate the corresponding shock process for these indicators by taking draws from the simulation smoother of Durbin and Koopman [2012].

for each BVAR specification m and horizon h, with $\Delta y^{q,o}_{T_v+h}$ the observed quarterly growth rate from the third release of GDP, $\Delta y^{q,f}_{T_v+h}$ the corresponding forecast, with v indexing data vintages, T_v the last non missing observation in that vintage, and N_h corresponding to the number of out-of-sample observations (vintages) available for each forecast horizon. Comparisons relative to surveys are easier to interpret in terms of gains in RMSFE. Therefore, we report

$$100\left(1-\frac{RMSFE^h_b}{RMSFE^h_s}\right), \tag{7}$$

with $RMSFE^h_b$ and $RMSFE^h_s$ corresponding to the benchmark MF-BVAR specification and surveys, respectively. With the RMSFE of the surveys in the denominator, positive values indicate percentage improvements in predictive accuracy with the MF-BVAR relative to the professional forecasters.

The statistical significance of any differences in unconditional predictive ability is assessed with a one-sided Diebold and Mariano [1995] test of equal mean-squared forecast error consistent with the sign of the percentage gain. To this test, we incorporate a small-sample size correction and calculate p-values using both standard Normal and Student's t critical values as recommended by Harvey et al. [1997]. Heteroskedastic and autocorrelation-consistent variances are constructed for this purpose using the Bartlett kernel with lag length set equal to four months for the nowcast and an additional three months for each forecast horizon (i.e. 4, 7, 10, 13, and 16 months, for horizons 0, 1, 2, 3, and 4 quarters ahead, respectively).

Following Giacomini and White [2006], we also report conditional tests of predictive ability. These tests are based on predicted differences in squared forecast errors conditional on a constant and one quarterly (three-month) lag. A predicted difference in squared forecast errors significantly less than zero indicates that the MF-BVAR conditionally outperformed a particular survey forecast in that period. To gauge the degree to which this was true, we

compute the share of forecasts for which the predicted difference in squared forecasts errors between the MF-BVAR and the surveys was negative. These results provide a rough sense of the real-time reliability of the MF-BVAR relative to the surveys, as they embed both the unconditional nature of the Diebold-Mariano test (which is equivalent to evaluating the predicted average difference in squared forecast errors) and the conditional nature of the predictive ability of past performance. We adopt the view of Carriero et al. [2015] that this test serves as a rough gauge of the statistical significance of the performance differences, as the properties of the test are derived under a fixed window estimation scheme while we use recursive samples instead.

4 Empirical Results

Having laid out the estimation framework for the MF-BVAR and discussed how we evaluate forecasts we now provide a detailed analysis of predictive performance. Section 4.1 first lays out our benchmark MF-BVAR specification, the evaluation sample and additional details regarding the estimation. We then present results in three parts. Section 4.2, compares real-time forecasts against the surveys of professional forecasters using this benchmark specification. We begin with results relative to the Blue Chip survey under the *Baseline Timing* assumption regarding the flow of information. To illustrate the MF-BVAR's ability to process the real-time information between surveys, we then compare these findings with those using our *Alternative Timing* assumption. As a robustness check, we finally contrast these results with a comparison of the MF-BVAR to the Survey of Professional Forecasters under both timing assumptions. In section 4.3, we turn to the sensitivity of the forecast performance of the MF-BVAR to several specification choices including (i) model size, (ii) the choice of hyperparameters, (iii) whether the model is in levels or growth rates, and (iv) lag length.

Finally, section 4.4 revisits how the performance of the benchmark MF-BVAR specification compares to a more traditional quarterly frequency BVAR. For this comparison, we present results within the context of both our real-time out-of-sample exercise as well as a longer pseudo out-of-sample experiment. In all of these cases, we draw comparisons on both a quarter-over-quarter and year-over-year growth rate basis in order to further disentangle the relative ability of the MF-BVARs to predict both short and medium-run movements in GDP.

4.1 Benchmark specification and evaluation sample

Our real-time out-of-sample forecasting exercise runs from the third quarter of 2004 through the third quarter of 2014 using recursive samples. The beginning of the sample is imposed by the availability of the CFNAI archives.[22] This results in an evaluation sample of 123 forecasts, each corresponding to a different real-time data vintage. Our sample is shorter than some others in the literature, but has the advantage of being able to document the real-time relevance of many data series previously unavailable for analysis.[23]

Our benchmark model comprises the twenty one series in Table 1, all in levels, and the lag length of the MF-BVAR is set to three. Regarding the choice of hyperparameters, as in Giannone et al. [2015], we specify prior distributions shown in the first five columns of table 2.[24] The 90 percent probability bands implied by these priors (sixth column in that table) encompass settings commonly found in the literature for persistent variables (equal to 5 for

[22] Conducting the forecasting exercise using a rolling window of 11 years (132 months) leads to comparable or slightly worse forecast performance. Results are available from the authors upon request.

[23] For the four-quarter ahead forecast horizon, we lose 15 out-of-sample observations, leading to a total sample size of 108. We drop one quarter during this period that coincided with the federal government shutdown in the third quarter of 2013. The shutdown delayed the release of a number of economic indicators including GDP and, hence, resulted in a delayed release schedule for the CFNAI which would have given the MF-BVAR an information advantage. Results are almost identical if this quarter is included.

[24] The notation of Giannone et al. [2015] for the hyperparameters corresponds to the the inverse of ours, such that an overall tightness of 4 in our context is equal to 1/4 in their case. We have experimented with estimating the inverse of our hyperparameters, as in their paper, and obtain broadly similar results provided the priors are adjusted accordingly to represent the same broad coverage of the hyperparameter domain.

the tightness and 1 for all other hyperparameters), while also allowing for smaller values (i.e. less shrinkage). The hyperparameters are chosen using only the first vintage in our real-time dataset and kept fixed thereafter.[25] The last column of table 2 reports the estimates of $\Lambda^\star$ for our benchmark MF-BVAR in levels. The optimal hyperparameters in our case are broadly in line with usual settings for models in levels, except for the sum of coefficients prior, λ_3, which is close to zero. The implications of this estimate for predictive accuracy is discussed later on.

Conditional on these hyperparameters, for every vintage the Gibbs sampler (described in section 2) is used to estimate the MF-BVAR. The first vintage covers the sample period from January of 1974 through July of 2004, with the initial three years of data used to elicit a prior for the initial unobserved states conditional on the prior means of the VAR parameters (see section 2 for further details).[26] For this first vintage, we initialize the Gibbs sampler using 24 parallel chains of 4,000 draws with a burn in phase of 2,000 iterations. For each subsequent vintage, the mean of the posterior density of VAR parameters from this initial exploration is used as an initialization, with the first 2,000 draws discarded and the remaining 2,000 retained. For our benchmark specification, this real-time evaluation requires 967 computer hours using a 2015 workstation with 24 cores using dual Xeon 2.5 GHz processors and performing all computations in Matlab with parallelization. We have checked, nonetheless, that chains with a larger number of draws deliver almost identical results, particularly for predictive accuracy.[27]

[25] The first part of our two-step procedure provides an initial guess for Λ^* that can be easily updated, say, every 6 or 12 vintages. However, performing a fine grid search around this guess in the second-step is computationally quite intensive. In general, updating the hyperparameters every two years produced fairly similar results to keeping them fixed and, hence, given the computational implications, we only report results under the scenario of holding these hyperparameters fixed. Forecast performance results where the hyperparameters are updated every 12 vintages are available from the authors upon request.

[26] More precisely, the first 5 months of 1973 are used to obtain mean values for the dummy priors, while data from June 1973 through December 1976 are used to run the Kalman filter using the prior mean of the VAR parameters. The resulting mean and variance for the state in December 1976 provide the initialization for the Kalman filtering step of the simulation smoother. This procedure is repeated, over the same sample period, for each data vintage to account for possible historical revisions or other changes to the data.

[27] Coding the filter and smoothing recursions in MEX files resulted in considerable computational gains. As a benchmark, for the filtering and smoothing of a time series that included 333 time periods and 25 state variables, the MEX version exhibited computational gains over the traditional Matlab version of 55 percent.

4.2 Comparisons to Professional Forecasters

Blue Chip Consensus

Figure 1 shows RMSFE percentage gains for our benchmark MF-BVAR specification relative to the Blue Chip survey under our conservative *Baseline Timing* assumption regarding the information available to survey respondents (see section 3.2). Reported are both unconditional gains (left panels) as well as the percentage of conditional forecasts with a lower RMSE (right panels) for both quarter-over-quarter (q/q) (top panels) and year-over-year (y/y) (bottom panels) growth rates of GDP. For each panel, forecasts are pooled across all forecast origins with the horizontal axis corresponding to the forecast horizon in quarters.

The key insight from this figure is that *medium*-run predictions from our MF-BVAR outperform the Blue Chip mean forecast in real-time over our sample period. Looking first at unconditional RMSFEs, the MF-BVAR delivers gains as large as 10 to 15 percent for quarter-over-quarter growth rates at the three and four quarter horizons (top left panel). For year-over-year growth rates, the performance improvements at these forecast horizons are even larger (bottom left). Furthermore, across both growth rate comparisons, the gains at three to four quarters ahead are statistically significant for standard confidence levels of one-sided Diebold-Mariano tests.[28]

For shorter forecast horizons, our MF-BVAR under-performs relative to the Blue Chip survey. However, in every such instance, the unconditional RMSFE gain of the survey mean forecast is small and not statistically different from zero. This is particularly evident in the nowcast (0 quarters ahead) where the Blue Chip mean forecast improves upon our MF-BVAR by less than 5 percent for both q/q and y/y growth rates (left panels).

[28] Figure 1 and subsequent figures report statistical significance using standard Normal critical values and the small-sample size correction recommended by Harvey et al. [1997]. Results with Student's t critical values were qualitatively similar and do not significantly change the inferences shown here.

These patterns holds as well when considering conditional RSMFE results (right panels). The number of predictions with lower conditional RMSFE than the Blue Chip mean forecast increases from less than 50% in the nowcast to roughly 80% at longer forecast horizons for both q/q and y/y growth rate comparisons. However, in this case, a statistically significant improvement of our MF-BVAR relative to the Blue Chip is achieved only for y/y growth rates at forecast horizons from two to four quarters ahead.[29]

Additional information is obtained by looking at the predicted differences in squared forecast errors based on past model performance, as suggested by Giacomini and White [2006]. These are reported in figure 2 for the year-over-year growth rate of GDP. In this case a negative value corresponds to a predicted squared forecast error difference favoring the MF-BVAR relative to the Blue Chip mean forecast. Interestingly, these plots suggest that the model's gains accrued mostly during the Great Recession, but were not limited to this episode, as revealed by the large percentage of predictions at nearly every forecast horizon with lower conditional RMSFE. This figure also reports the p-values from the Giacomini-White test of equal conditional forecast accuracy, which is rejected in favor of the MF-BVAR particularly for predictions 3 and 4 quarters out.

Blue Chip Consensus with Alternative Timing

The conservative approach to the flow of information used above lends itself to two separate but closely related questions. First, how effective is the MF-BVAR at incorporating the real-time flow of information within a month? And, would there be any value to a principal (e.g. a policymaker) in using the MF-BVAR forecasts updated with this data relative to the last

[29] To provide further context, the RSMFE results for q/q growth rates suggest that the MF-BVAR outperforms the Blue Chip survey by roughly 30 basis points at the four quarter horizon, while for the nowcast the Blue Chip survey outperforms the MF-BVAR by about 10 basis points (both in terms of annualized growth). For y/y growth rates, the RSMFE results suggest that the MF-BVAR outperforms the Blue Chip survey by about 40 basis points at the four quarter horizon, while for the nowcast the Blue Chip survey outperforms the MF-BVAR by roughly 5 basis points.

survey? To answer both of these questions, we re-estimate our MF-BVAR using all of the vintage data in the model available through the middle of each month.

Results under this *Alternative Timing* assumption are displayed in figure 3, whose structure mirrors that of the *Baseline Timing* in figure 1 to facilitate comparisons. The first thing to note from this figure is that the performance of our MF-BVAR at medium-term forecast horizons is remarkably robust in terms of both the magnitude of the RMSFE gains and statistical significance relative to the Blue Chip mean forecasts. This is true in terms of both unconditional and conditional predictive ability. As expected, the real-time data flow within each month matters considerably for model performance at shorter forecast horizons. In particular, the additional information from data releases through the middle of each month significantly improves the MF-BVAR's nowcast and one-quarter ahead predictions relative to the Blue Chip mean forecasts. In fact, for both growth rates and at all horizons the MF-BVAR outperforms this survey. Furthermore, the MF-BVAR delivers a forecast with lower conditional RMSFE than the Blue Chip mean forecast in 80% or more of cases at all forecast horizons, with statistical significance achieved for y/y growth rates at two to four quarters ahead.

Clearly, forecast comparisons with this *Alternative Timing* assumption are unfair if one wishes to understand how the MF-BVAR fares relative to the Blue Chip survey in real-time. Instead, this exercise demonstrates that the model incorporates the newly available information in a manner that improves forecast accuracy. More specifically, the MF-BVAR proves to be particularly effective in updating near-term forecasts with incoming data between Blue Chip survey releases. Another way in which to see this is to consider again the predicted differences in squared forecast errors based on past model performance for the year-over-year growth rate of GDP. Under our *Alternative Timing* assumption, the MF-BVAR delivers a nowcast with lower conditional RMSFE than the Blue Chip mean forecast in 100% of cases

(right panels in figure 1).

Survey of Professional Forecasters

Next, we compare our MF-BVAR to the Survey of Professional Forecasters (SPF) under both timing assumptions. Here, the comparison is over fewer forecast origins given the SPF's structure of only producing one forecast per quarter (taken near the middle of the second month of the quarter). Given our *Baseline Timing* assumption, SPF forecasters have more real-time information available to them than our MF-BVAR contains, particularly the Employment Situation report. In contrast, our *Alternative Timing* assumption, while not perfect, should closely replicate the SPF information set by nature of the fact that it aligns much more closely with the SPF survey dates.

Figure 4 presents the forecast performance for our MF-BVAR compared to the SPF's median forecast under both timing assumptions. Not surprisingly, the results are broadly similar to those reported in comparison to the Blue Chip survey across both timing assumptions and for both q/q and y/y growth rates. For the q/q growth rates, the MF-BVAR outperforms the SPF at every forecast horizon under the *Baseline Timing* except for the nowcast. Under the *Alternative Timing* assumption, however, gains are recorded across all horizons. For both timing assumptions, the gains are statistically significant at longer horizons (e.g. 3 and 4 quarters out).

For y/y growth rates, the MF-BVAR compares favorably to the SPF at longer horizons regardless of the assumptions on the flow of information. The only instance where the relative performance of the MF-BVAR falls short of the SPF comes under the *Baseline Timing* at the nowcast and one quarter ahead horizons.[30] As explained, due to differences in the information

[30]The SPF, unlike Blue Chip, provides a forecast of the previous quarter's revised level which we use when constructing the current quarter's growth rate forecast for SPF. The Blue Chip survey simply gives the current quarter's growth rate.

set we would expect this discrepancy to shrink considerably under the *Alternative Timing* assumption, which is exactly what can be seen when comparing the left and right panels.

Consequently, the patterns of forecast performance of our MF-BVAR relative to the SPF are similar to that versus the Blue Chip survey, with any differences at short horizons most likely accentuated by the greater informational disadvantage of the MF-BVAR under our *Baseline Timing* assumption. More importantly, pooling across the Blue Chip and SPF, the favorable performance of the MF-BVAR relative to surveys of professional forecasters two quarters and beyond does not appear to be explained by differences in information sets across surveys.

4.3 Specifications and Forecast Accuracy

Motivated by the favorable comparison of our benchmark MF-BVAR to surveys of professional forecasters, we now explore how sensitive the model's performance is to various specification choices. We tackle in turn the issues of model size, the choice of priors, whether variables are specified in levels or growth rates, and lag length. For each alternative specification, we measure the gains in RMSFE for our benchmark specification. That is, we report improvements in RMSFE as in equation (7) with our benchmark MF-BVAR in the numerator, such that positive values correspond to improvements compared to a particular alternative specification. As for the statistical significance of any RMSFE differences, strictly speaking, due to the encompassing nature of some specifications, the Diebold-Mariano tests do not apply. However, motivated by the Monte Carlo evidence reported by Clark and McCracken [2011a,b], we take the conservative approach of Carriero et al. [2012] in reporting one-sided test results. Finally, all specifications (including the benchmark of course) are estimated under our *Baseline Timing* assumption, over the same sample, and with the same methods that underlie the results of the previous section.

Model size

Model size has been shown to be an important dimension of forecast performance in single frequency BVARs (Bańbura et al. [2010], Koop [2013], Chauvet and Potter [2013]). Optimal model size depends on the relative benefits of incorporating more information from additional indicators versus the costs of estimating additional parameters. The analysis of this section addresses this issue in the MF-BVAR context. Specifically, we answer two questions: (i) how does a model with only a few commonly used and timely indicators perform relative to our benchmark MF-BVAR? And, (ii) what is the relative value of the additional monthly versus quarterly series contained within our benchmark MF-BVAR specification?

To answer the first question, we consider a *small-scale* model that includes a subset of the monthly variables from our benchmark specification. In addition to quarterly GDP, this model retains Industrial Production, monthly Personal Consumption Expenditures, hours worked, and the ISM Manufacturing Purchasing Managers Index. These four series are among the most commonly referenced indicators of U.S. economic activity, with hours worked encompassing both the extensive and intensive margins of employment fluctuations. Furthermore, these series are among the most timely indicators available each month and, therefore, do not suffer from long availability lags, as is the case with some of the other series in our benchmark specification.

To answer the second question, we consider a *medium-scale* model which builds off from the small-scale model by also including the additional monthly variables from our benchmark specification. More specifically, we add to the small-scale model real manufacturing and trade sales, real manufacturing and trade inventories, real manufacturer's orders of core capital goods, capacity utilization, the ratio of total business inventories and sales, real non-residential private construction spending, real public construction spending, real retail sales, real personal income less transfers, real exports of goods, and real imports of goods, all of which are monthly.

Gains in RMSFE (unconditional) for our benchmark MF-BVAR relative to the small-scale and medium-scale models are shown in table 3 for both q/q and y/y growth rates at forecast horizons 0–4 quarters ahead. Focusing on the small-scale model (first two columns), a clear pattern across these results is evident. The information contained in the additional variables included in our benchmark specification significantly enhances forecast performance. This improvement is evident both across all forecast horizons as well as types of growth rates. Broadly speaking, our benchmark MF-BVAR outperforms the small-scale model for q/q growth rates of GDP by roughly 15 percent on average across horizons, with the gains being statistically significant at standard confidence levels. Percentage gains for y/y growth rates are generally even larger.[31]

Next, we examine the relative performance of our benchmark MF-BVAR to the medium-scale model (last two columns). Two observations emerge from this comparison. First, our benchmark MF-BVAR generally outperforms the medium-scale model across both types of growth rates and forecast horizons, but to a much smaller degree than it does relative to the small-scale model. Gains in RMSFE range from roughly 1-2 percent for q/q growth rates and 4-8 percent for y/y growth rates, and are negligible in both instances for the nowcast. Second, the remaining quarterly indicators in the benchmark specification generally improve forecast performance, particularly for predictions one quarter ahead and beyond, but gains are considerably smaller than when adding the monthly series to the small-scale model. To draw this conclusion, we note that the medium and benchmark specification differ solely in the presence of the quarterly series. As such, Table 3 reveals significant gains in expanding the number of monthly variables (comparison to small) and more muted gains with additional quarterly data (comparison to medium). We interpret these results as suggesting that the

[31] Tests of equal predictive ability comparing the small-scale model to the surveys of professional forecasters overwhelmingly reject the null of equal forecast accuracy in favor of the surveys for all horizons and both q/q and y/y growth rates.

favorable performance of our benchmark specification relative to the surveys of professional forecasters stems in large part from the information embedded in the additional monthly indicators contained within the medium-scale model.

Priors

As outlined in section 2.4, our estimation strategy includes a data-driven methodology for selecting hyperparameters centered on maximizing the marginal data density. As discussed in Geweke [2001], this approach leads to superior one-step ahead prediction performance. However, how this approach performs along different forecast horizons is less clear. We assess how important the choice of priors is by documenting the gains/losses across all horizons from picking standard (default) values from the literature.

To this end, an alternative specification (with the same variables as our benchmark specification) is estimated with the following hyperparameters

$$\lambda_1 = 5; \;\; \lambda_2 = \lambda_3 = \lambda_4 = 1, \tag{8}$$

which are the default values used in Carriero et al. [2015] and Giannone et al. [2015] in the context of traditional, i.e. single frequency, BVARs. The last two columns of table 4 show the RMSFE gains with our benchmark and reveal that a data-driven method for choosing hyperparameters yields considerable improvements in forecast accuracy. Averaging across horizons, the hyperparameters chosen with the marginal data density improve RMSFEs by 20 percent for y/y growth rates, with gains in the 6 to 16 range for q/q growth rates as well.

Two additional results of this exercise are noteworthy. First, it is interesting that these improvements in RMSFE are larger than the 1 to 3 percent gains with similar comparisons reported by Carriero et al. [2015] and Giannone et al. [2015] for single frequency VARs. Second, comparing the hyperparameters in (8) with those selected with the marginal likelihood

(table 2) suggests that the increases in accuracy stem from the hyperparameter on the sum of coefficients, which at 0.15 is considerably lower than the value of 1 that is customary in the literature (see section A.4). We have verified that this is the case by re-running our benchmark MF-BVAR changing only this hyperparameter relative to the default values. In Appendix A.4 we further note that the marginal is sharply peaked with respect to this hyperparameter.

Levels vs. Growth Rates

Up until now, the results presented come from a MF-BVAR estimated with all of its indicators modeled in levels (or log levels). However, given that the ultimate forecast of interest is the *growth* rate of GDP, it would also be natural to work with a specification in growth rates instead. In this section, we assess the robustness of our results to data transformations and answer the questions: Do MF-BVAR specifications in growth rates perform better than in levels? And does this vary by forecast horizon?

The alternative specification includes the same series (and lags) as the benchmark; but, of course, using growth rates instead. To choose hyperparameters with the marginal data density, the prior must be modified to reflect the belief that growth rates are more likely (than levels) to be stationary. In particular, the co-persistence prior is shut down by setting $\lambda_4 = 0$, while the tightness (and decay) are selected with centers that shrink the individual first autoregressive lags (δ_i) toward zero for most series, as is customary.[32] The sum of coefficients is allowed to add up to 0 (or 1 for the ISM index and IS ratio), but the optimal value for this hyperparameter, λ_3, came in routinely at zero. Consequently, this form of shrinkage was not imposed.

The first two columns of table 4 compare this specification in growth rates to our bench-

[32]The transformation of two variables is retained from the levels specification, the ISM index and the inventory-sales (IS) ratio, since they do not exhibit random walk with drift behavior and their growth rates are quite volatile. For symmetry with the levels case, for these two variables δ_i is selected with the elements of Λ.

mark MF-BVAR in levels. The overall message from this table is that the benchmark specification in levels performs considerably better, with larger gains accruing at longer forecast horizons. More specifically, in terms of RMSFE gains, the benchmark MF-BVAR delivers improvements in forecast accuracy of about 8 percent on average across forecast horizons for q/q growth rates with larger gains accruing for y/y growth rate forecasts, the majority of which are statistically significant. Once again, differences in the nowcast are rather small, particularly for q/q growth rate forecasts.

Lag Length

The prior described in section 2.3 already shrinks coefficients on distant lags toward zero (with the strength of this prior controlled by the hyperparameter λ_3). However, this does not preclude the choice of lag length from impacting the predictive accuracy of the MF-BVAR. In this section, we explore the sensitivity of our benchmark MF-BVAR's forecast performance to the number of lags. For this exercise, our benchmark model (with 3 lags) is re-estimated with both four and five lags; results with longer lags (six and seven) provided qualitatively very similar results and are omitted simply for space considerations. Importantly, for each alternative lag length, the hyperparameters were re-estimated using the priors shown in table 2.

Table 5 presents the forecast performance results for alternative lag lengths relative to our benchmark three-lag MF-BVAR. Modest (but statistically significant) gains accrue compared to both of the longer lag specifications. However, broadly speaking it appears that the relative gains/losses in forecast performance across lag lengths are small, provided that the shrinkage parameter on distant lags is chosen optimally with each specification.[33] Most, if any, gains in RMSFE seem to be concentrated in the nowcast horizon and dissipate quickly at longer

[33] Not surprisingly the value of λ_3 selected with the marginal likelihood increases with the number of lags, implying more shrinkage.

horizons, particularly for q/q growth rates.

4.4 Comparison to a Quarterly BVAR

To further investigate how our benchmark MF-BVAR incorporates the monthly flow of information within a quarter, we contrast its forecast performance with a traditional quarterly BVAR using the same set of indicators. This comparison builds on the analysis of Schorfheide and Song [2015], who document performance gains from a MF-BVAR relative to a traditional quarterly BVAR in the near term but that die off beyond two quarters. However, our analysis of this issue differs from theirs in two important respects. First, we wish to understand if the waning benefits of working in mixed frequency hold if the object of interest is the year-over-year growth rate of GDP as opposed to the quarterly growth rate, which is commonly the case among policymakers. Second, and more technically, our quarterly BVAR must account for the changing flow of information given the staggered nature of data releases for our various indicators. For example, while the ISM index is published with merely a one month publication lag, manufacturing and trade sales are released with a three month delay. As such, the quarterly average for the latter series will be available two months later than the corresponding quarterly average for the ISM. This staggered pattern of missing values for quarterly data is not considered by Schorfheide and Song [2015] and necessitates the use of the Kalman filter for conditional forecasting (see Appendix 9).

Our comparison of the MF-BVAR and the quarterly BVAR is twofold. In the first part, we use exclusively the real-time vintage dataset described previously. This comparison benefits from its ability to best mimic the real-time flow of information available to professional forecasters. In the second part, we instead contrast the two models across a "pseudo real-time" dataset. This exercise involves evaluating forecasts over a longer sample period (January 1989 to July 2014), but with the caveat of using a dataset that does not replicate the exact real-

time information flow.[34] Both evaluations are performed on a monthly basis across forecast origins. For our real-time exercise, the monthly information updates for the quarterly BVAR concern revisions to past data and also occur when all monthly realizations for a given variable within a quarter are available. On this second issue, as explained in Appendix 9 we try to align the information flow across mixed frequency and quarterly models as close as possible by incorporating the monthly data as they complete a quarter. Finally, the quarterly BVAR is also estimated in levels, with a prior chosen with the marginal data density, two lags, and the same variables as our benchmark specification.[35]

The left column of figure 5 illustrates the relative forecast performance of our benchmark MF-BVAR and quarterly BVAR in real-time from the third quarter of 2004 through the third quarter of 2014. Overall, the MF-BVAR outperforms the traditional quarterly BVAR across forecast horizons and types of growth rates, achieving RMSFE gains of about 10 percent on average. Focusing on the quarterly growth rates (top left panel), performance gains accrue most heavily in the nowcast and one-quarter ahead horizon, with smaller gains at longer horizons. This pattern accords well with the results in Schorfheide and Song [2015]. However, the gains are more stable when one examines y/y growth rates (bottom left panel). Here, the MF-BVAR outperforms the traditional quarterly BVAR across all forecast horizons by 10-15 percent or more, with most of these gains being statistically significant.

The right column of figure 5 displays the relative pseudo real-time forecast performance of our benchmark MF-BVAR over the (longer) forecasting sample from January 1989 to July 2014. The predictive gains of the MF-BVAR over the traditional quarterly BVAR in this case are similar to those found in the real-time sample. Not surprisingly, given the longer history of forecasts, the results over this sample tend to more often be statistically significant. Focusing

[34] To create the "pseudo real-time" dataset, the final vintages from our real-time dataset were truncated recursively by one month going back through time.

[35] The quarterly BVAR with two lags performs slightly better than a quarterly specification with four lags instead.

on the quarterly growth rates (top right panel), most of the gains accrue in the nowcast and one-quarter ahead horizon, but now all forecast horizons experience statistically significant improvements. Similar to the real-time sample, the predictive benefits of the MF-BVAR for y/y growth rates remain considerable 1 year out (bottom right panel); here, too, all forecast horizons demonstrate statistically significant gains.

These results suggests that similar to Schorfheide and Song [2015] the performance gains of the MF-BVAR relative to the traditional quarterly BVAR for quarter-over-quarter growth rates are more concentrated in the near-term forecast horizons. That is, the ability of the MF-BVAR to incorporate the real-time flow of monthly information appears less critical to forecasting the quarter-over-quarter growth rate of GDP at longer horizons (albeit in our case gains remain substantial even one year out for the pseudo real-time comparison). In contrast, we find that the performance gains of the MF-BVAR relative to the traditional quarterly BVAR for year-over-year growth rates are more robust at longer forecast horizons.

5 Conclusion

We document the superior performance of a moderately sized MF-BVAR relative to surveys of professional forecasters for medium-term forecasts of U.S. real GDP growth. Gains in predictive accuracy over surveys are shown to be statistically significant, to accrue both conditionally and unconditionally, and to be larger for yearly as opposed to quarterly growth rates. When the information sets are closely aligned to the different timing of information across surveys the MF-BVAR also performs competitively at shorter horizons including the nowcast. The analysis leverages a novel dataset that includes a larger number of series available in real-time than what is usually considered in the literature. Still, the favorable comparison of the MF-BVAR to surveys is noteworthy considering that, relative to forecasters, we have confined

ourselves to only data for which real-time data vintages are currently available. Regarding the role of specification choices, model size, prior selection and data transformation were shown to have meaningful impacts on predictive accuracy.

6 Tables and Figures

Table 1: Summary of U.S. Macroeconomic Indicators

	Frequency	Lagged Vintage	Publication Lag (months)
Real Personal Consumption Expenditures	M	-	2
Industrial Production	M	x	2
Aggregate Weekly Hours Worked	M	x	2
ISM Manufacturing PMI	M	-	1
Real Manufacturing and Trade Sales	M	-	3
Real Manufacturers' Orders of Core Capital Goods	M	-	2
Capacity Utilization	M	x	2
Real Manufacturing and Trade Inventories	M	-	3
(Total) Business Inventories to (Total) Sales Ratio	M	x	3
Real Non-residential Private Construction Spending	M	-	2
Real Public Construction Spending	M	-	2
Real Retail Sales	M	x	2
Real Personal Income Less Transfers	M	-	2
Real Exports of Goods	M	-	2
Real Imports of Goods	M	-	2
GDP	Q	-	2-4
PCE: Nondurable Goods	Q	-	2-4
Business Fixed Investment	Q	-	2-4
Government Consumption and Gross Investment	Q	-	2-4
Exports of Goods and Services	Q	-	2-4
Imports of Goods and Services	Q	-	2-4

Notes: M–monthly, Q–quarterly

Table 2: Prior for Hyperparameters and Posterior Estimates

Λ	Description	Density	Mean	Std	[5,95] Prior Band	Optimal (Λ^*)
λ_1	Tightness	Gamma	4	3	[0.6, 9.85]	5.08
λ_2	Decay	Gamma	2	1.5	[0.3, 4.93]	1.11
λ_3	Sum of coefficients	Gamma	2	1.5	[0.3, 4.93]	0.10
λ_4	Co-persistence	Gamma	2	1.5	[0.3, 4.93]	0.89

Table 3: Percentage Gains in RMSFE of Benchmark Relative to Alternative–Model Size

	Small		**Medium**	
Horizon	Y/Y	Q/Q	Y/Y	Q/Q
0	$6.2^{\star\star}/^{\bullet\bullet}/^{\dagger\dagger}/^{\ddagger\ddagger}$	$12.6^{\star\star\star}/^{\bullet\bullet\bullet}/^{\dagger\dagger\dagger}/^{\ddagger\ddagger\ddagger}$	-0.3	0.4
1	$17.5^{\star\star\star}/^{\bullet\bullet\bullet}/^{\dagger\dagger\dagger}/^{\ddagger\ddagger\ddagger}$	$13.7^{\star\star}/^{\bullet\bullet}/^{\dagger\dagger}/^{\ddagger\ddagger}$	$3.9^{\star\star\star}/^{\bullet\bullet\bullet}/^{\dagger\dagger\dagger}/^{\ddagger\ddagger\ddagger}$	1.4
2	$29.9^{\star\star\star}/^{\bullet\bullet\bullet}/^{\dagger\dagger\dagger}/^{\ddagger\ddagger\ddagger}$	$21.2^{\star\star}/^{\bullet\bullet}/^{\dagger\dagger}/^{\ddagger\ddagger}$	$6.6^{\star\star\star}/^{\bullet\bullet\bullet}/^{\dagger\dagger\dagger}/^{\ddagger\ddagger\ddagger}$	$2.8^{\star\star}/^{\bullet\bullet}/^{\dagger\dagger}/^{\ddagger\ddagger}$
3	$35.1^{\star\star}/^{\bullet\bullet}/^{\dagger\dagger}/^{\ddagger\ddagger}$	$17.2^{\star}/^{\bullet}/^{\dagger}/^{\ddagger}$	$8.4^{\star\star\star}/^{\bullet\bullet\bullet}/^{\dagger\dagger\dagger}/^{\ddagger\ddagger\ddagger}$	$1.5^{\star\star}/^{\bullet\bullet}/^{\dagger\dagger}/^{\ddagger\ddagger}$
4	$38.4^{\star\star}/^{\bullet\bullet}/^{\dagger\dagger}/^{\ddagger\ddagger}$	$15.7^{\star\star}/^{\bullet\bullet}/^{\dagger\dagger}/^{\ddagger\ddagger}$	$8.1^{\star\star\star}/^{\bullet\bullet\bullet}/^{\dagger\dagger\dagger}/^{\ddagger\ddagger\ddagger}$	$1.7^{\star\star}/^{\bullet\bullet}/^{\dagger\dagger}/^{\ddagger\ddagger}$

Notes: Entries in this table correspond to percentage gains in RMSFE for GDP growth at forecast horizons 0–4 quarters ahead (rows) for our benchmark MF-BVAR in levels with optimal hyperparameters set to maximize the marginal data density and 3 lags. Percentage gains are reported for both quarter-over-quarter (Q/Q) and year-over-year (Y/Y) growth rates. All evaluations use the Third Release of GDP to compute RMSFE, with the alternative model specification in the denominator of the ratio. Positive values indicate gains relative to the alternative specification. $\star/\bullet/\dagger/\ddagger$ denote statistical significance from one-sided Diebold and Mariano [1995] ($\star/\bullet$) and Harvey et al. [1997] ($\dagger/\ddagger$) tests using standard Normal and Student's t critical values at the 15, ($\star\star$) 10, ($\star\star\star$) and 5 percent level, respectively. HAC variances were computed with the Bartlett kernel and lag length equal to four months for the nowcast and an additional three months for each forecast horizon.

Table 4: Percentage gains in RMSFE of Benchmark Relative to Alternative–Transformation and Prior Selection

	Default Hyperparameters		**Growth Rates**	
Horizon	Y/Y	Q/Q	Y/Y	Q/Q
0	$6.9^{\star\star\star/\bullet\bullet/\dagger\dagger\dagger/\ddagger\ddagger}$	$6.1^{\star/\bullet/\dagger/\ddagger}$	$5.4^{\star\star/\bullet\bullet/\dagger\dagger/\ddagger\ddagger}$	1.5
1	$16.7^{\star\star/\bullet\bullet/\dagger\dagger/\ddagger\ddagger}$	12.3	2.8	$8.2^{\star\star\star/\bullet\bullet\bullet/\dagger\dagger\dagger/\ddagger\ddagger}$
2	$26.3^{\star/\bullet/\dagger/\ddagger}$	$16.0^{\star/\bullet/\dagger/\ddagger}$	$14.8^{\star\star/\bullet\bullet/\dagger\dagger/\ddagger\ddagger}$	$13.9^{\star\star\star/\bullet\bullet\bullet/\dagger\dagger\dagger/\ddagger\ddagger\ddagger}$
3	$28.1^{\star/\bullet/\dagger/\ddagger}$	10.9	$19.5^{\star/\bullet/\dagger/\ddagger}$	$8.7^{\star\star/\bullet\bullet/\dagger\dagger/\ddagger\ddagger}$
4	$28.5^{\star/\bullet/\dagger/\ddagger}$	9.0	$23.6^{\star\star\star/\bullet\bullet\bullet/\dagger\dagger\dagger/\ddagger\ddagger\ddagger}$	$7.7^{\star\star\star/\bullet\bullet\bullet/\dagger\dagger\dagger/\ddagger\ddagger\ddagger}$

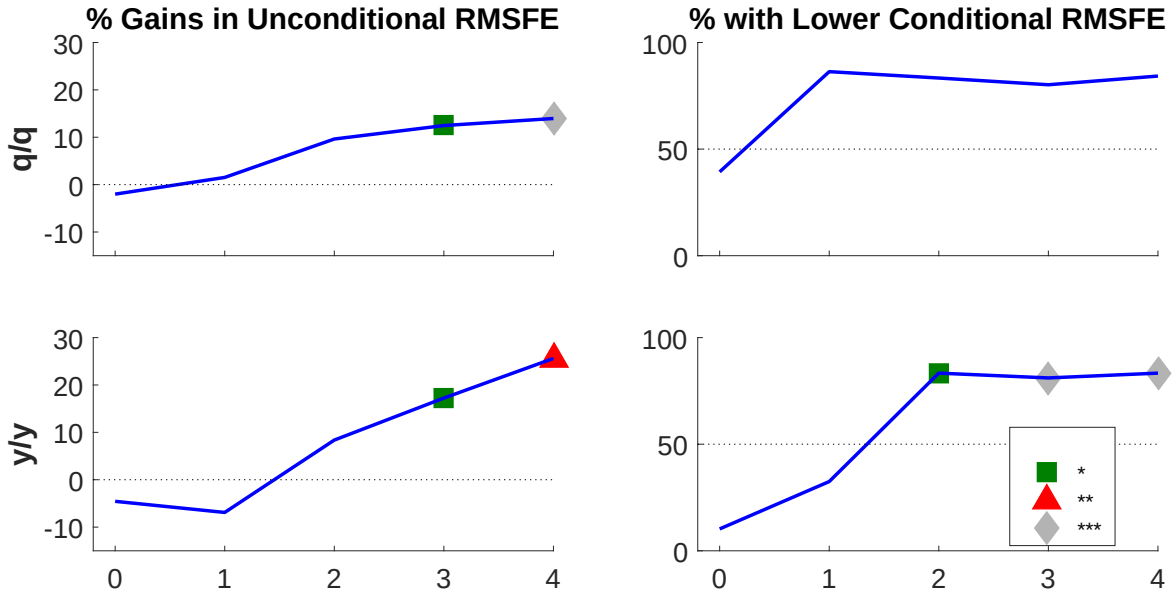
Notes: Entries in this table correspond to percentage gains in RMSFE for GDP growth at forecast horizons 0–4 quarters ahead (rows) for our benchmark MF-BVAR in levels with optimal hyperparameters set to maximize the marginal data density and 3 lags. Percentage gains are reported for both quarter-over-quarter (Q/Q) and year-over-year (Y/Y) growth rates. All evaluations the Third Release of GDP to compute RMSFE, with the alternative model specification in the denominator of the ratio. Positive values indicate gains relative to the alternative specification. $\star/\bullet/\dagger/\ddagger$ denote statistical significance from one-sided Diebold and Mariano [1995] ($\star/\bullet$) and Harvey et al. [1997] ($\dagger/\ddagger$) tests using standard Normal and Student's t critical values at the 15, ($\star\star$) 10, ($\star\star\star$) and 5 percent level, respectively. HAC variances were computed with the Bartlett kernel and lag length equal to four months for the nowcast and an additional three months for each forecast horizon.

Table 5: Percentage gains in RMSFE of Benchmark Relative to Alternative–Lag Length

	Four Lags		**Five Lags**	
Horizon	Y/Y	Q/Q	Y/Y	Q/Q
0	$1.7^{\star\star}/^{\bullet\bullet}/^{\dagger\dagger}/^{\ddagger\ddagger}$	$1.6^{\star}/^{\bullet}/^{\dagger}/^{\ddagger}$	$3.8^{\star\star\star}/^{\bullet\bullet\bullet}/^{\dagger\dagger\dagger}/^{\ddagger\ddagger\ddagger}$	$3.3^{\star\star}/^{\bullet\bullet}/^{\dagger\dagger}/^{\ddagger\ddagger}$
1	1.0	-0.4	-0.3	-1.0
2	1.3	$1.8^{\star\star\star}/^{\bullet\bullet\bullet}/^{\dagger\dagger\dagger}/^{\ddagger\ddagger\ddagger}$	$2.6^{\star}/^{\bullet}/^{\dagger}/^{\ddagger}$	$3.9^{\star\star\star}/^{\bullet\bullet\bullet}/^{\dagger\dagger\dagger}/^{\ddagger\ddagger\ddagger}$
3	$2.1^{\star\star}/^{\bullet\bullet}/^{\dagger}/^{\ddagger}$	0.2	$3.4^{\star\star}/^{\bullet\bullet}/^{\dagger\dagger}/^{\ddagger\ddagger}$	1.0
4	$2.0^{\star\star\star}/^{\bullet\bullet\bullet}/^{\dagger\dagger\dagger}/^{\ddagger\ddagger\ddagger}$	-0.8	$4.9^{\star\star\star}/^{\bullet\bullet\bullet}/^{\dagger\dagger\dagger}/^{\ddagger\ddagger\ddagger}$	-0.1

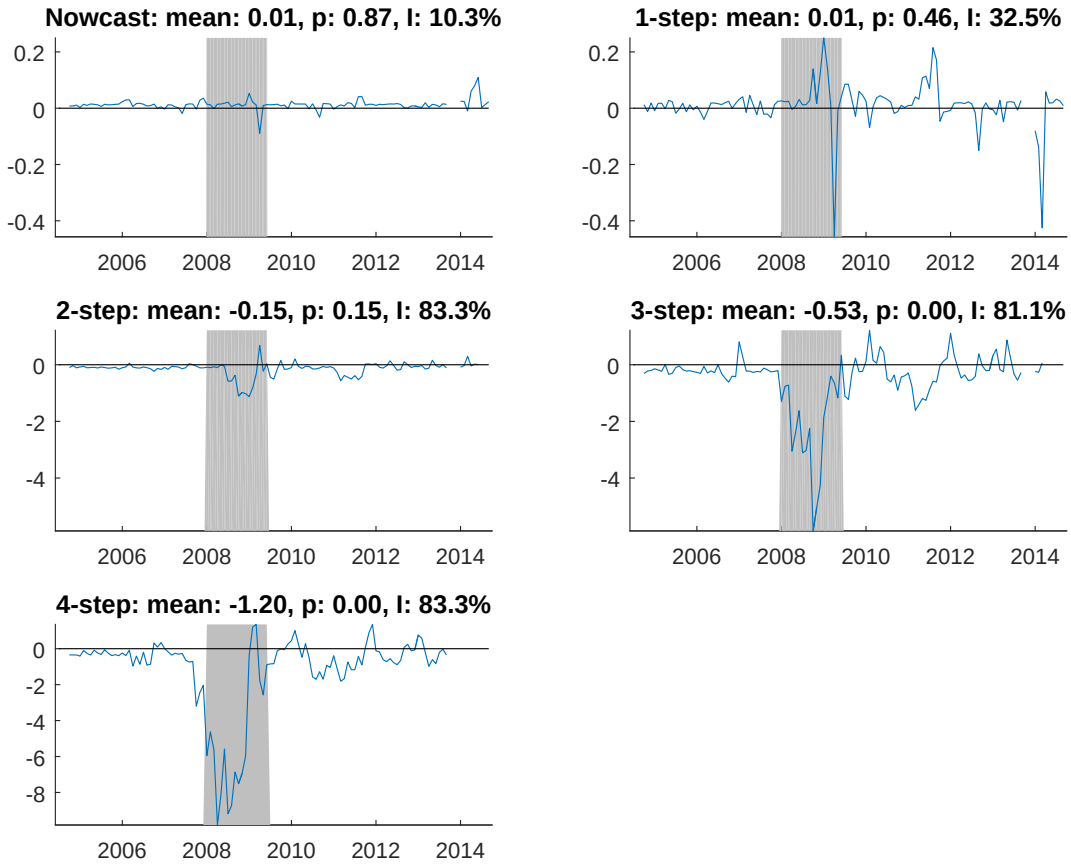
Notes: Entries in this table correspond to percentage gains in RMSFE for GDP growth at forecast horizons 0–4 quarters ahead (rows) for our benchmark MF-BVAR in levels with optimal hyperparameters set to maximize the marginal data density and 3 lags. Percentage gains are reported for both quarter-over-quarter (Q/Q) and year-over-year (Y/Y) growth rates. All evaluations use the Third Release of GDP to compute RMSFE, with the alternative model specification in the denominator of the ratio. Positive values indicate gains relative to the alternative specification. $\star/\bullet/\dagger/\ddagger$ denote statistical significance from one-sided Diebold and Mariano [1995] ($\star/\bullet$) and Harvey et al. [1997] ($\dagger/\ddagger$) tests using standard Normal and Student's t critical values at the 15, ($\star\star$) 10, ($\star\star\star$) and 5 percent level, respectively. HAC variances were computed with the Bartlett kernel and lag length equal to four months for the nowcast and an additional three months for each forecast horizon.

Figure 1: Percentage Gains in RMSFE Relative to Blue Chip Consensus



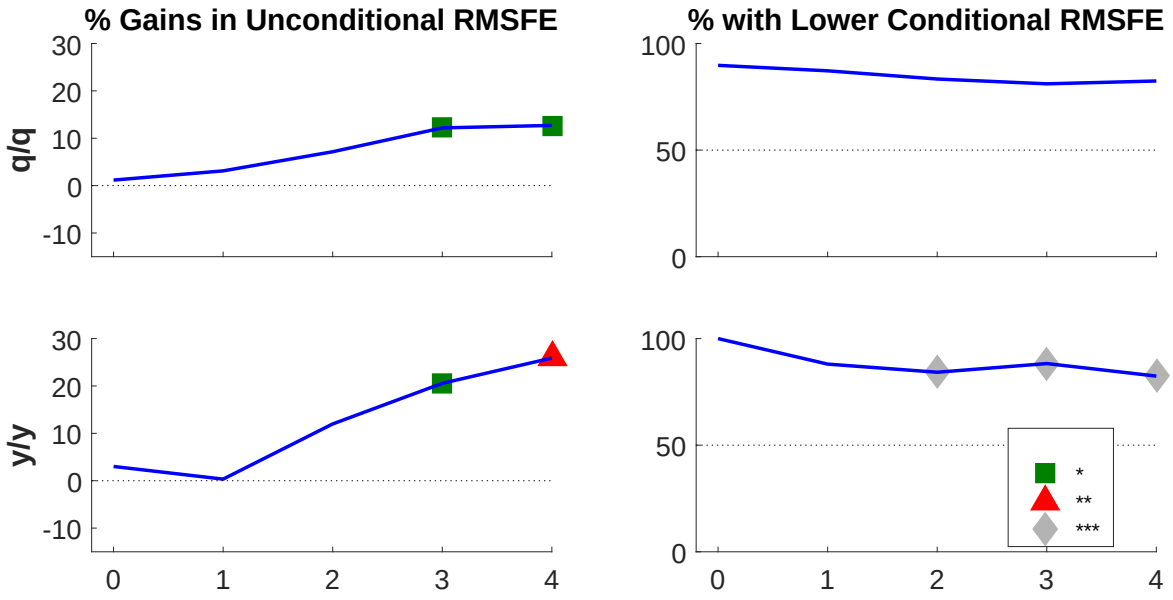
Notes: This figure displays RMSFE gains (both unconditional and conditional) for GDP growth of our benchmark MF-BVAR forecasts relative to the Blue Chip Consensus (BCC) mean forecasts under our *Baseline Timing* assumption discussed in section 3.2. In each panel, relative RMSFE gains are reported for forecast horizons 0 (nowcast) - 4 quarters ahead. All evaluations use the Third Release of GDP to compute RMSFE. Positive values indicate gains relative to BCC. Markers denote statistical significance from one-sided Diebold and Mariano [1995] tests for equal forecast accuracy using standard Normal critical values and the small-sample size correction suggested by Harvey et al. [1997] with HAC variances computed using the Bartlett kernel and lag length equal to four months for the nowcast and an additional three months for each subsequent forecast horizon. ($\star$) denotes rejection of the null of equal mean-squared forecast error between the MF-BVAR and the BCC forecasts at the 15, ($\star\star$) 10, and ($\star\star\star$) 5 percent level, respectively.

Figure 2: Predicted Squared Forecast Error Differences Relative to Blue Chip Consensus: Baseline Timing



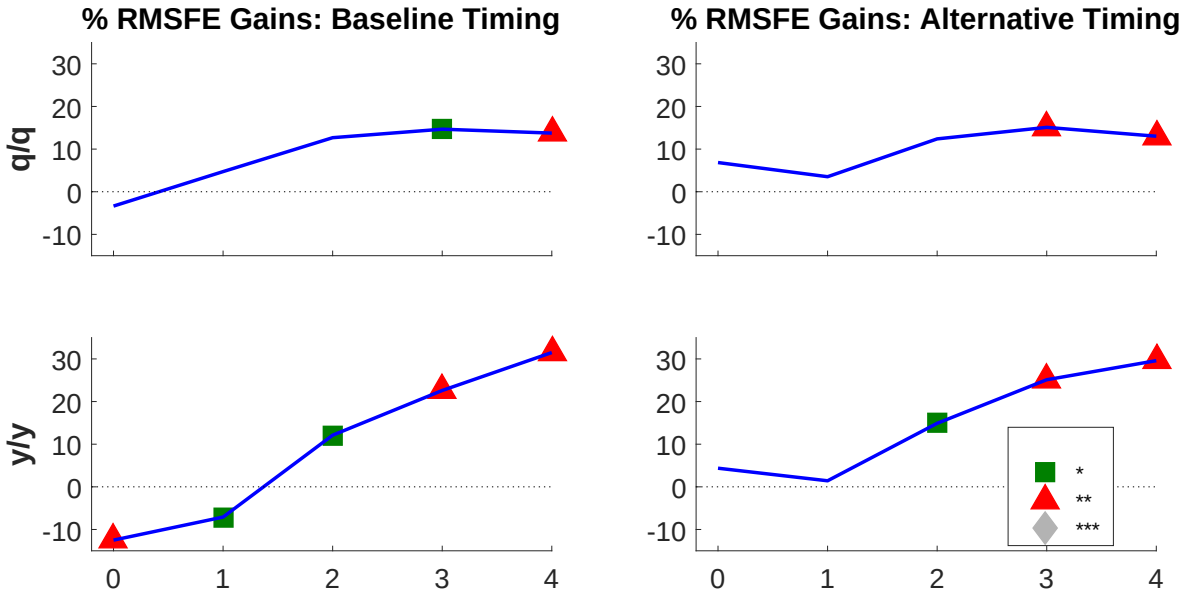
Notes: This figure displays predicted squared forecast error differences for year-over-year GDP growth between the benchmark MF-BVAR forecasts and the Blue Chip Consensus (BCC) mean forecasts under our *Baseline Timing* assumption discussed in section 3.2. The shaded period denotes the timing of the 2007-2009 U.S. recession according to the National Bureau of Economic Research. Negative values indicate gains relative to BCC. The top of each panel reports the average predicted squared forecast error difference (mean), its associated p-value from the Giacomini and White [2006] test of equal conditional forecast accuracy (p), and the number of forecasts where the MF-BVAR has a lower RMSFE conditional on the prior quarter's prediction (I), respectively, for forecast horizons 0 (nowcast) - 4 quarters ahead. All evaluations use the Third Release of GDP to compute forecast errors.

Figure 3: Percentage Gains Relative to Blue Chip Consensus: Alternative Timing



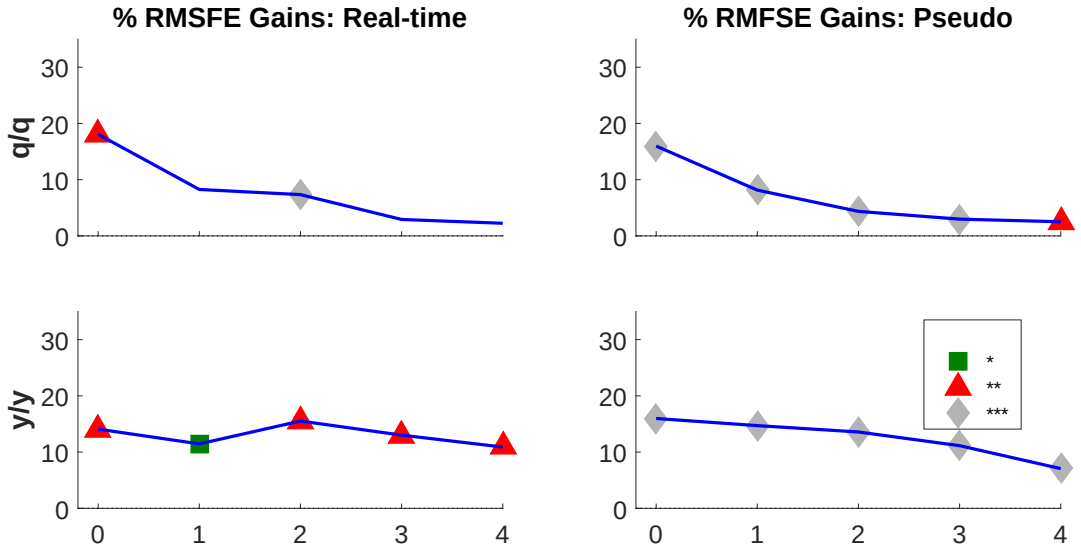
Notes: This figure displays RMSFE gains (both unconditional and conditional) for GDP growth of our benchmark MF-BVAR forecasts relative to the Blue Chip Consensus (BCC) mean forecasts under our *Alternative Timing* assumption discussed in section 3.2. In each panel, relative RMSFE gains are reported for forecast horizons 0 (nowcast) - 4 quarters ahead. All evaluations use the Third Release of GDP to compute RMSFE. Positive values indicate gains relative to BCC. Markers denote statistical significance from one-sided Diebold and Mariano [1995] tests for equal forecast accuracy using standard Normal critical values and the small-sample size correction suggested by Harvey et al. [1997] with HAC variances computed using the Bartlett kernel and lag length equal to four months for the nowcast and an additional three months for each subsequent forecast horizon. ($\star$) denotes rejection of the null of equal mean-squared forecast error between the MF-BVAR and the BCC forecasts at the 15, ($\star\star$) 10, and ($\star\star\star$) 5 percent level, respectively.

Figure 4: Percentage Gains in RMSFE Relative to Survey of Professional Forecasts



Notes: This figure displays RMSFE gains for GDP growth of our benchmark MF-BVAR forecasts relative to the Survey of Professional Forecasters (SPF) median forecasts for both the *Baseline* and *Alternative Timing* assumptions discussed in section 3.2. In each panel, relative RMSFE gains are reported for forecast horizons 0 (nowcast) - 4 quarters ahead. All evaluations use the Third Release of GDP to compute RMSFE. Positive values indicate gains relative to SPF. Markers denote statistical significance from one-sided Diebold and Mariano [1995] tests for equal forecast accuracy using standard Normal critical values and the small-sample size correction suggested by Harvey et al. [1997] with HAC variances computed using the Bartlett kernel and lag length equal to four months for the nowcast and an additional three months for each subsequent forecast horizon. ($\star$) denotes rejection of the null of equal mean-squared forecast error between the MF-BVAR and the SPF forecasts at the 15, ($\star\star$) 10, and ($\star\star\star$) 5 percent level, respectively.

Figure 5: Percentage Gains in RMSFE relative to Quarterly BVAR



Notes: This figure displays real-time and pseudo real-time RMSFE gains for GDP growth of our benchmark MF-BVAR forecasts relative to the quarterly BVAR discussed in section 4.4. In each panel, relative RMSFE gains are reported for forecast horizons 0 (nowcast) - 4 quarters ahead. Real-time evaluations use the Third Release of GDP to compute RMSFE, while pseudo real-time evaluations use the July 2014 vintage. Positive values indicate gains relative to the quarterly BVAR. Markers denote statistical significance from one-sided Diebold and Mariano [1995] tests for equal forecast accuracy using standard Normal critical values and the small-sample size correction suggested by Harvey et al. [1997] with HAC variances computed using the Bartlett kernel and lag length equal to four months for the nowcast and an additional three months for each subsequent forecast horizon. ($\star$) denotes rejection of the null of equal mean-squared forecast error between the MF-BVAR and the quarterly BVAR forecasts at the 15, ($\star\star$) 10, and ($\star\star\star$) 5 percent level, respectively.

References

S. Borağan Aruoba, Francis X. Diebold, and Chiara Scotti. Real-time measurement of business conditions. *Journal of Business & Economic Statistics*, 27(4):417–427, 2009.

Marta Bańbura, Domenico Giannone, and Lucrezia Reichlin. Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, 25(1):71–92, 2010.

Scott Brave and R. Andrew Butters. Chicago Fed National Activity Index turns ten - analyzing its first decade of performance. *Chicago Fed Letter*, (273), 2010.

Scott Brave and R. Andrew Butters. Diagnosing the Financial System: Financial Conditions and Financial Stress. *International Journal of Central Banking*, 8(2):191–239, June 2012.

Scott Brave and R. Andrew Butters. Nowcasting Using the Chicago Fed National Activity Index. *Economic Perspectives*, (Quarter I):19–37, 2014.

Scott Brave, R. Andrew Butters, and Alejandro Justiniano. A generalized Kalman filter and smoother with application to mixed-frequency data. Technical note, Federal Reserve Bank of Chicago, 2015.

Andrea Carriero, George Kapetanios, and Massimiliano Marcellino. Forecasting large datasets with Bayesian reduced rank multivariate models. *Journal of Applied Econometrics*, 26(5): 735–761, 2011.

Andrea Carriero, Todd E. Clark, and Massimiliano Marcellino. Common drifting volatility in large Bayesian VARs. Working Paper 1206, Federal Reserve Bank of Cleveland, 2012.

Andrea Carriero, Todd E. Clark, and Massimiliano Marcellino. Bayesian VARs: Specification choices and forecast accuracy. *Journal of Applied Econometrics*, 30(1):46–73, 2015.

Marcelle Chauvet and Simon Potter. Chapter 3 – Forecasting output. In Graham Elliott and Allan Timmermann, editors, *Handbook of Economic Forecasting*, volume 2, Part A of *Handbook of Economic Forecasting*, pages 141–194. Elsevier, 2013.

Todd Clark and Michael W. McCracken. Nested forecast model comparisons: a new approach to testing equal accuracy. Working paper, Federal Reserve Bank of St. Louis, 2011a.

Todd E. Clark and Michael W. McCracken. Testing for unconditional predictive ability. In Michael P. Clements and David F. Hendry, editors, *The Oxford Handbook of Economic Forecasting*. Oxford University Press, 2011b.

Marco Del Negro and Frank Schorfheide. Bayesian macroeconometrics. *The Oxford handbook of Bayesian econometrics*, pages 293–389, 2011.

Francis X Diebold and Roberto S Mariano. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263, July 1995.

Thomas Doan, Robert Litterman, and Christopher Sims. Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, 3(1):1–100, 1984.

James Durbin and Siem Jan Koopman. *Time Series Analysis by State Space Methods: Second Edition*. Oxford University Press, March 2012.

Federal Reserve Bank of Chicago. Chicago Fed National Activity Index (CFNAI), 2015. Available at `https://www.chicagofed.org/publications/cfnai/index`.

Claudia Foroni, Eric Ghysels, and Massimiliano Marcellino. Mixed-frequency vector autoregressive models, 2013.

John Geweke. Bayesian econometrics and forecasting. *Journal of Econometrics*, 100(1):11–15, January 2001.

Eric Ghysels. Macroeconomics and the reality of mixed frequency data. *Journal of Econometrics*, forthcoming, 2016.

Eric Ghysels, Pedro Santa-Clara, and Rossen Valkanov. The MIDAS touch: Mixed data sampling regression models. CIRANO Working Papers 2004s-20, CIRANO, May 2004.

Raffaella Giacomini and Halbert White. Tests of conditional predictive ability. *Econometrica*, 74(6):1545–1578, 2006.

Domenico Giannone, Michele Lenza, and Giorgio E. Primiceri. Prior Selection for Vector Autoregressions. *The Review of Economics and Statistics*, 2(97):436–451, May 2015.

Andrew C. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambrindge University Press, 1989.

David Harvey, Stephen Leybourne, and Paul Newbold. Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2):281–291, June 1997.

Sune Karlsson. Chapter 15 – Forecasting with Bayesian vector autoregression. In Graham Elliott and Allan Timmermann, editors, *Handbook of Economic Forecasting*, volume 2, Part B, pages 791–897. Elsevier, 2013.

Gary M. Koop. Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics*, 28(2):177–203, 2013.

Robert B. Litterman. Forecasting with Bayesian vector autoregressions: Five years of experience. *Journal of Business & Economic Statistics*, 4(1):25–38, 1986.

Roberto S. Mariano and Yasutomo Murasawa. A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, 18(4):427–443, 2003.

Roberto S. Mariano and Yasutomo Murasawa. A Coincident Index, Common Factors, and Monthly Real GDP. *Oxford Bulletin of Economics and Statistics*, 72(1):27–46, 02 2010.

Michael McCracken, Michael Owyang, and Tatevik Sekhposyan. Real-time forecasting with a large, mixed-frequency bayesian VAR. Working paper, St. Louis Federal Reserve Bank, October 1 2015.

Michael W. McCracken and Serena Ng. FRED-MD: A monthly database for macroeconomic research. Working Papers 2015-12, Federal Reserve Bank of St. Louis, June 2015. URL `https://ideas.repec.org/p/fip/fedlwp/2015-012.html`.

Tommaso Proietti. Temporal disaggregation by state space methods: Dynamic regression methods revisited. *The Econometrics Journal*, 9(3):357–372, 2006.

Frank Schorfheide and Dongho Song. Real-time forecasting with a mixed-frequency VAR. *Journal of Business & Economic Statistics*, pages 1–30, 2015.

Christopher A. Sims. Using a likelihood perspective to sharpen econometric discourse: Three examples. *Journal of Econometrics*, 95(2):443–462, April 2000.

Christopher A Sims and Tao Zha. Bayesian methods for dynamic multivariate models. *International Economic Review*, 39(4):949–968, November 1998.

A Methodology: Additional Details

In section 2, we provided the general empirical approach to the estimation of the MF-BVAR and the subsequent evaluation of its forecasts. In this section, we develop in more detail the construction of the state-space system and provide the details of the interpolation procedure involved in finding the optimal hyperparameters. For clarity, some equations from within the text are reproduced in this section.

A.1 Building the State-Space System

What follows is a more detailed discussion of our state-space framework accommodating monthly and quarterly time series. A more general description of how one might build a state-space system with other forms of mixed frequency data and the subsequent use of the Kalman filter and smoother is provided by Brave et al. [2015].

As in section 2.1, we consider an n-dimensional vector y_t of macroeconomic time series of differing frequencies (e.g. some monthly indicators and some quarterly indicators). Due to the mixed frequency nature of the series in y_t, all the variables within y_t will not be observed every period. To this end, partition $y_t^{'} = \begin{bmatrix} y_t^{q'} & , & y_t^{m'} \end{bmatrix}$ such that the first n_q elements collects the vector y_t^q of quarterly variables, such as Gross Domestic Product, which are observed only once every three periods in a monthly model. In turn, let y_t^m be comprised solely of monthly indicators, such as Industrial Production, with dimension $n_m = n - n_q$.

To describe the monthly dynamics of this system, let x_t^q denote the monthly latent variables underlying the quarterly series, y_t^q. We combine these latent variables with the indicators observed at a monthly frequency in $x_t^{'} = \begin{bmatrix} x_t^{q'} & , & x_t^{m'} \end{bmatrix}$. Clearly, each element of x_t^m corresponds to the element of y_t^m when observed. In contrast, some aggregated combination of past x_t^q monthly realizations will equal y_t^q when the quarterly variables are observed. In general,

the aggregation for some series i is deterministic and given by:

$$y_t^q(i) = G_i(x_t^q(i), x_{t-1}^q(i), ..., x_{t-s}^q(i))$$

for some pre-determined horizon s.[36] An example of $G_i(\cdot)$, common for measures of economic activity in levels, is the three-month average of x_t^q, such that

$$y_t^q(i) = \frac{x_t^q(i) + x_{t-1}^q(i) + x_{t-2}^q(i)}{3}. \tag{9}$$

When working with growth rates (Δy_t^q), x_t corresponds to the first difference instead of the level, and an alternative accumulator, the "triangle accumulator," is used. The triangle accumulator specifies the quarterly growth rate of GDP as given by:[37]

$$\Delta y_t^q(i) \equiv y_t^q(i) - y_{t-3}^q(i) = \frac{x_t^q(i) + 2x_{t-1}^q(i) + 3x_{t-2}^q(i) + 2x_{t-3}^q(i) + x_{t-4}^q}{3}. \tag{10}$$

With the mapping of x_t to y_t determined, the vector x_t and its monthly dynamics are summarized by the vector autoregression of order p given by

$$x_t = c + \Phi_1 x_{t-1} + ... + \Phi_p x_{t-p} + \epsilon_t; \quad \epsilon_t \sim i.i.d.N(0, \Sigma), \tag{11}$$

where each Φ_l is an n-dimensional square matrix containing the coefficients associated with lag l. The companion form of this monthly VAR together with a measurement equation for y_t delivers the common two equation state-space system given by

[36] We follow the approach of Mariano and Murasawa [2003] and treat the quarterly observations of GDP and its subcomponents as the quarterly *average* of the monthly realizations. This leads to the interpretation that the underlying monthly variable is annualized.

[37] The triangle accumulator is an approximate aggregation that preserves the linearity of the system. Mariano and Murasawa [2010] use this approximation in their examination of mixed frequency factor models.

$$y_t = Z_t s_t \tag{12}$$

$$s_t = C_t + T_t s_{t-1} + R_t \epsilon_t, \tag{13}$$

with the vector of observables, y_t, defined as above, and the state vector, s_t, defined as in the text as

$$s_t' = \left[x_t',\ \ldots,\ x_{t-p}',\ \Psi_t'\right],$$

which includes both lags of the time series at the monthly frequency, and Ψ_t, a vector of accumulators. For GDP, the accumulator used for the benchmark (levels) MF-BVAR is defined by equation 9, while the accumulator used for the MF-BVAR in growth rates is given by equation 10.

Given the additional variables in the state (the accumulators, Ψ_t), the transition matrix is an $n * p + n_q$ square matrix. In the transition matrix, the entries of the first n rows are the concatenation of the coefficients associated with each lag $\Phi = [\Phi_1, \Phi_2, ..., \Phi_p]$. The last n_q rows are made up of two separate components. The first component involves a (time-varying) scaled version of the coefficients associated with the quarterly time series and corresponds to the current monthly contribution to the "accumulator" series. The second component involves a deterministic series of fractions (e.g. 0, 1/2 and 1/3 for the regular average) that loads onto the lagged value of the accumulator and corresponds to a running total of past contributions of monthly realizations within the current quarter. The remaining entries of this matrix correspond to ones and zeros to preserve the lag structure. The VAR intercepts sit at the top of C_t, while scaled versions of intercepts are in rows associated with each accumulator. The rest of C_t has zeros. Finally, each R_t corresponds to the natural selection matrix, using the

same deterministic series of fractions used in T_t augmented to accommodate the additional accumulator variables in the state.

In periods in which all of the indicators are observed, the selection matrix Z_t is comprised solely of n selection rows made up of zeros and ones. Specifically, for these periods the Z_t matrix is given by:

$$Z_t = \begin{bmatrix} 0 & 0 & \dots & I_{n_q} \\ 0 & I_{n_m} & \dots & 0 \end{bmatrix},$$

where the identity matrix in the first n_q rows of Z_t corresponds to the mapping of the accumulators to the quarterly variables, and the identity matrix in the last n_m rows of Z_t corresponds to the mapping of the monthly (base frequency) time series and their observed counterparts in y_t.

The row dimension of Z_t varies over time due to the changing dimensionality of the observables. For the months in which only monthly time series are observed, the last n_m rows of Z_t will be included. Furthermore, towards the end of the sample not all of the monthly indicators will be available, depending on their release schedule; and, hence, a further subset of these last n_m rows will not be used.

A.2 Priors through dummy observations

We consider four forms of shrinkage implemented through dummy observations appended to the actual data and given by

$$
Y_d = \begin{pmatrix} \lambda_1 \text{diag}(\bar{y}_1\sigma_1\delta_1, \ldots, \bar{y}_n\sigma_n\delta_n) \\ 0_{n(p-1)\times n} \\ \ldots \\ \text{diag}(\sigma_1, \ldots, \sigma_n) \\ \ldots \\ 0_{1\times n} \\ \ldots \\ \lambda_3 \text{diag}(\omega_1\bar{y}_1, \ldots, \omega_n\bar{y}_n) \\ \ldots \\ \lambda_4\bar{y} \end{pmatrix} \quad X_d = \begin{pmatrix} \lambda_1 J_p \otimes \text{diag}(\bar{y}_1\sigma_1, \ldots, \bar{y}_n\sigma_n) & 0_{np\times 1} \\ \ldots & \ldots \\ 0_{n\times np} & 0_{n\times 1} \\ \ldots & \ldots \\ 0_{1\times np} & \alpha \\ \ldots & \ldots \\ (1_{1\times p}) \otimes \lambda_3 \text{diag}(\omega_1\bar{y}_1, \ldots, \omega_n\bar{y}_n) & 0_{n\times 1} \\ \ldots & \ldots \\ (1_{1\times p}) \otimes \lambda_4\bar{y} & \lambda_4 \end{pmatrix}. \tag{14}
$$

The first block corresponds to the *tightness* and *decay* components of the prior governed by $\lambda_1$and λ_2, respectively; and where $J_p = \text{diag}(1^{\lambda_2}, \ldots, p^{\lambda_2})$, $\bar{y}$ is an n-dimensional vector of pre-sample means, while the n-dimensional vector $\bar{\sigma} = (\sigma_1, \ldots, \sigma_n)'$ has as its i-th element the residual variance for each series from a univariate p-lag autoregression on a pre-sample. The series specific scalars δ_i reflect the center of the prior for the first order own-lag autoregressive coefficients and are usually set to 1, 0.8 or 0. The second block implements the prior for the residual variances, while the third one represents the diffuse prior for the intercepts with α a small number (1e-5). Prior information regarding the sum of coefficients is governed by λ_3, where once again the series-specific scalars ω_i correspond to the centers of the prior and are set to 1 (or 0.8 in the case of the ISM index and the inventory-to-sales (IS) ratio). Finally, λ_4 controls beliefs regarding the co-persistence of the system.

A.3 Interpolation Model

Solving for the set of hyperparameters that maximize the marginal data density cannot be accomplished analytically due to the presence of latent variables (see section 2.4). Consequently, to find the optimal hyperparamters a grid search is required. Before this grid is constructed, we evaluate the optimal hyperparameters (available in analytical form) of an approximation to the marginal likelihood. This approximation involves interpolating the quarterly time series using the procedure described in this section. This allows us to explore the broad patterns of the marginal data density and to construct an informed grid with which to optimize the correct marginal density, as disccused in Appendix A.4. The rest of this section describes the state-space system based off the work of Proietti [2006] that is used to estimate the interpolated series.

The goal of the interpolation procedure is to generate a monthly time series, y_t, for an observed quarterly series, Y_t. We impose a temporal aggregation constraint such that the implied quarterly aggregates of the interpolated monthly series match exactly the quarterly time series observed. Moreover, a set of related (monthly) series, R_t, not already incorporated into our MF-BVAR are used to inform the month-to-month variation in the interpolated series.

This framework lends itself naturally to a state-space system, where the interpolated series, y_t, is modeled as an unobserved state variable. A fairly general interpolation model is given by the following system:

$$\begin{bmatrix} R_t \\ Y_t \end{bmatrix} = \begin{bmatrix} \beta_0 \\ 0 \end{bmatrix} + \begin{bmatrix} \beta_1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_t \\ \Psi_t \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \end{bmatrix} \tag{15}$$

$$\begin{bmatrix} y_t \\ \Psi_t \end{bmatrix} = T_t \begin{bmatrix} y_{t-1} \\ \Psi_{t-1} \end{bmatrix} + R_t \eta_t \tag{16}$$

$$\epsilon_t \sim N(0, \sigma_\epsilon) \quad \eta_t \sim N(0, \sigma_\eta), \tag{17}$$

where any potential AR(p) coefficients are embedded in the system matrix T_t and an accumulator (Ψ_t) is used to preserve the appropriate temporal aggregation properties of the monthly time series. In our empirical analysis we specify a first order autoregression with coefficient ρ. The complete vector of parameters $\Theta = (\beta_0, \beta_1, \rho, \sigma_\epsilon, \sigma_\eta)$ can be estimated using maximum likelihood methods, aided by the Kalman filter which allows for an easy calculation of the log-likelihood function. Once estimated, we generate a smoothed estimate of y_t conditional on the inferred parameters and the full history of the quarterly series (Y_t) and the related monthly series (R_t).

A.4 Contours of the Approximate and Correct Marginal Likelihood

To select the hyperparameters governing the priors we make use of the marginal data density (see section section 2.4). To provide an informed grid over which to search, an approximation to this marginal data density is initially explored. This approximation primarily involves interpolating the quarterly series as described in the previous section (see section A.3), which can be used to obtain an analytically convenient approximation to the true marginal data density.[38] Exploring the general countours of this approximate marginal data density allows us to set up an informed grid for each hyperparameter, and run the Gibbs sampler for all possible combinations of the grid elements. In each case, the modified harmonic mean is used to estimate the correct marginal $P(Y_{0:T}|\Lambda)$, and the set of hyperparameters attaining the highest value for this density is selected.

Clearly, the loosely speaking "approximate" marginal density does not correspond to the

[38] For all results reported in this paper, 120 different starting values were generated at random from the prior for the hyperparameters described in table 2 and put through different optimization routines using only the first data vintage.

correct marginal of the MF-BVAR as it does not account for the latent states. Nonetheless, since it is orders of magnitude easier to compute and maximize, it can help in guiding the intialization of the more computationally demanding grid search. Moreover, it can help gauge the peakedness of the marginal data density and possible identification issues. However, the usefulness of this initial exploration depends on the similarity between the "approximate" and correct data densities.

Figures 6 and 7 shed light on these issues by showing aspects of the marginal data density from the first and second step, respectively. For each figure, the top panels provide the surface (left) and contours (right) over a domain for λ_1 (controlling the overall tightness) and λ_3 (governing the sum of coefficients) for a fixed value of λ_2, and λ_4 (the remaining optimal hyperparameters; see table 2). The bottom panels provide slices of the marginal data density for λ_1 and λ_3. A few patterns emerge from the comparison of both figures. First, broadly speaking the two surfaces display similar shapes both in terms of where they peak as well as what combination of hyperparameters constitute level sets. Second, both marginal data densities are more sharply peaked around the optimal value of λ_3 (0.1), and less so for λ_1 (5.08) . This distinction is interesting given that the optimal hyperparameter for λ_3 for the benchmark MF-BVAR is the only one that deviates considerably from standard values in the literature (which are 1 for λ_3 and 5 for λ_1). As mentioned in the text, this investigation provides complementary evidence of the role of the sum of coefficients prior on the sensitivity of the forecasting performance, as evidenced with the benchmark MF-BVAR in section 4.3.

Furthermore, note that the contours of the correct data density in figure 7 are noisier than the interpolated ones, which is particularly evident in the slices for λ_1 (but considerably less so for $\lambda 3$). This reflects the simulation error from the Gibbs sampler estimate. Finally, the marginal data densities do not coincide in magnitudes, as expected, due in part to the adjustment for missing observations in the correct density for the mixed frequency case.

Figure 6: Contours of the "Approximate" Marginal Data Density (Conditional on Interpolated data)

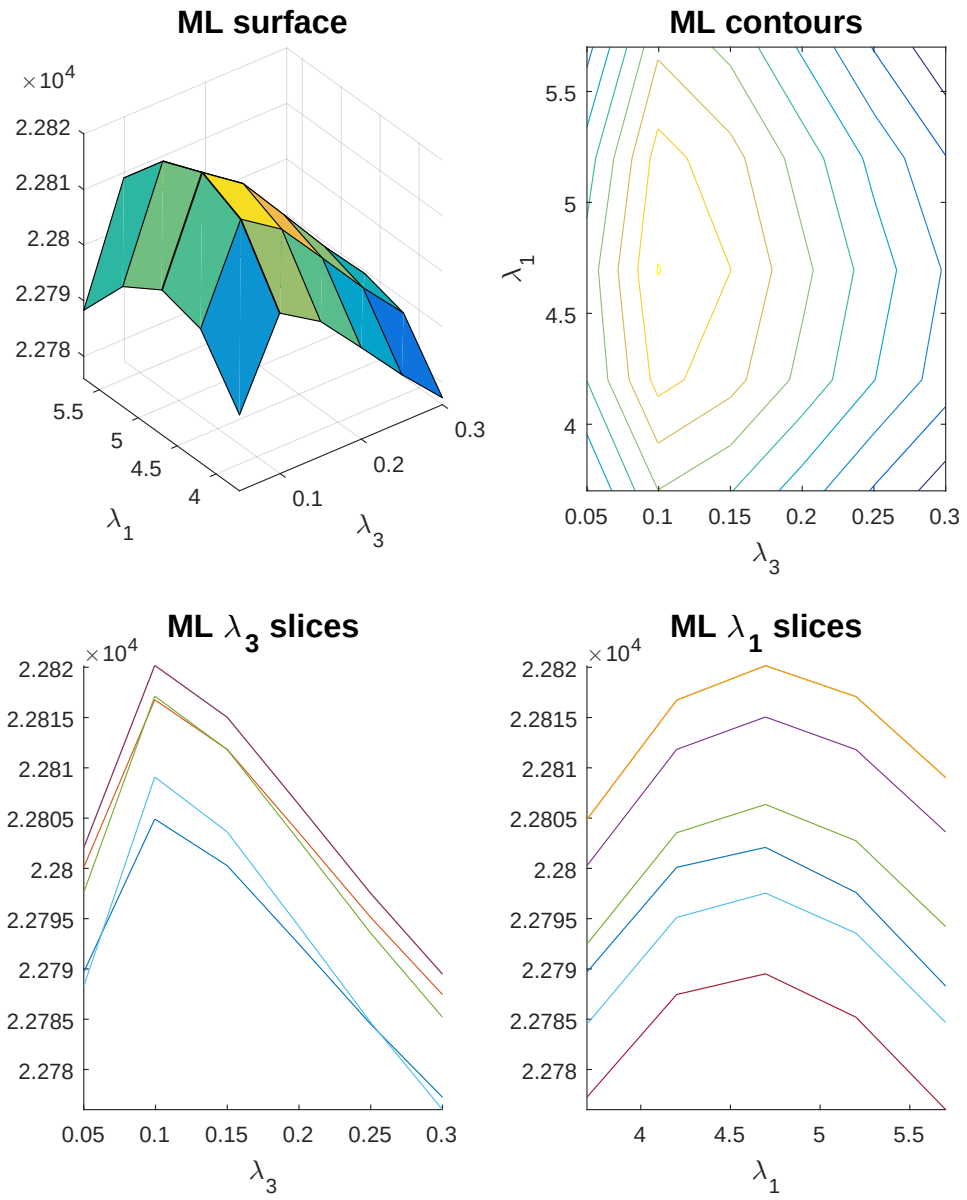
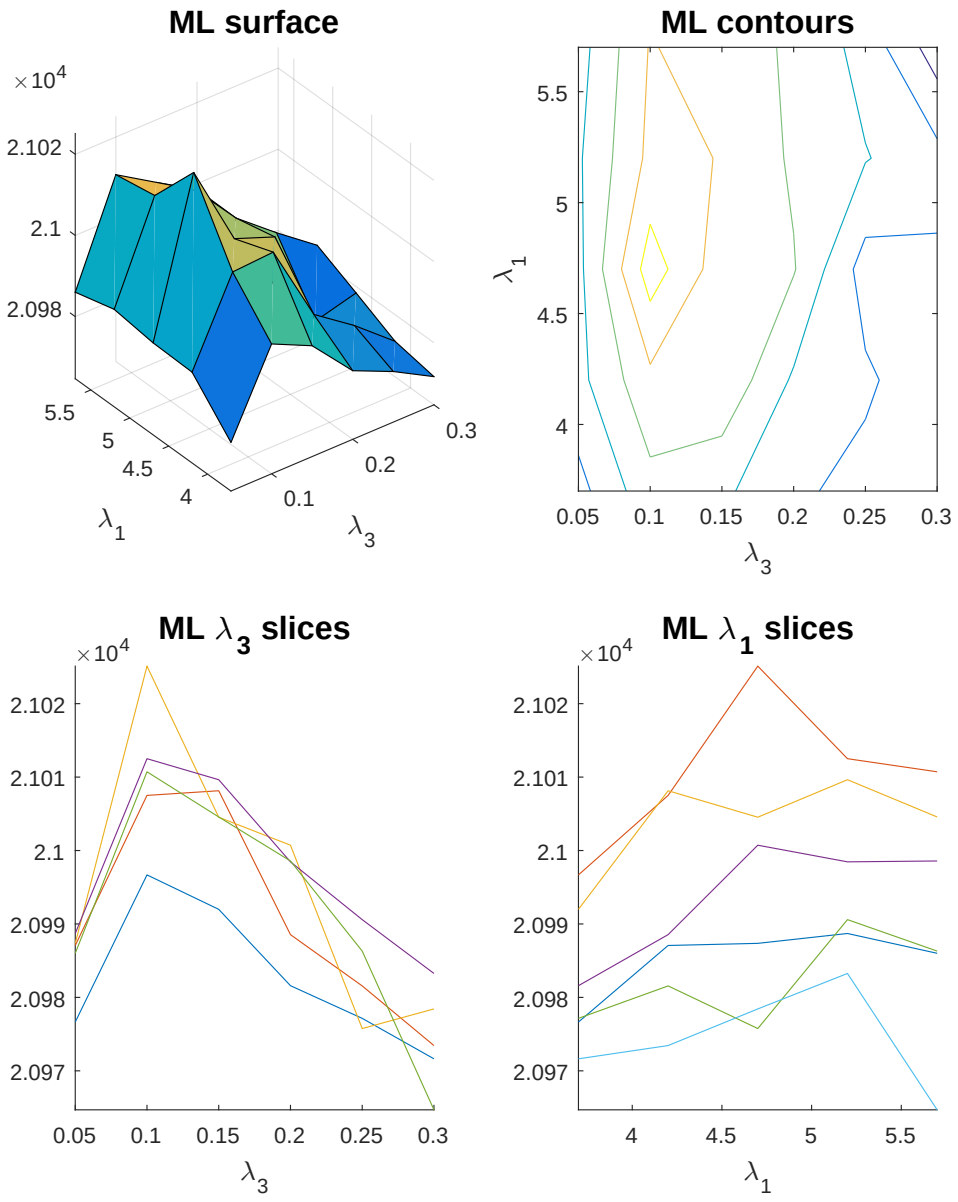


Figure 7: Contours of the MF-BVAR Marginal Data Density Obtained via Modified Harmonic Mean and Gibbs Sampler

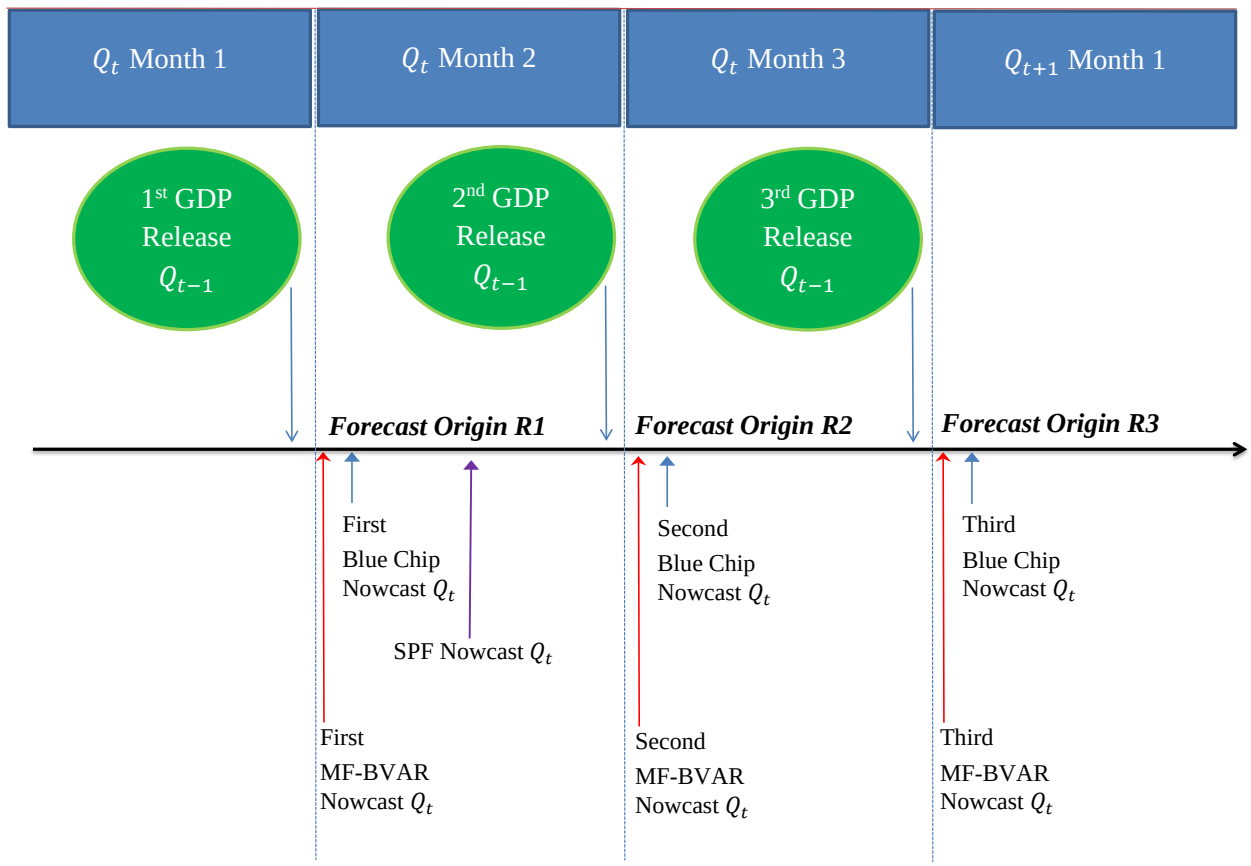


A.5 Forecast Origins and Flow of Information

Figure 8 details the labeling of forecast origins and information flow under the *Baseline Timing* assumption discussed in section **3** for a generic quarter, Q_t. The top of the figure reports calendar time (e.g. Month 1 is April for second quarter). At the end of each month NIPA releases for the previous quarter become available and are hence available to survey respondents at the beginning of the next month (e.g. the first Blue Chip survey with information on first quarter's first release, Q_{t-1}, is conducted at the beginning of May). This is how we index forecast origins for the nowcasts and forecasts, as seen in figure 8.

Our *Baseline Timing* assumption purposefully vintage lags information that may have been available to the Blue Chip respondents at the beginning of the month. This informational disadvantage of our MF-BVBAR is particularly evident with the the Survey of Professional Forecasters, which is conducted in the middle of the second month of the quarter, corresponding to Forecast Origin 1. These respondents, for instance, have access to the Employment Situation report. The *Alternative Timing* assumption uses information available through the middle of each month and hence better aligns with the SPF. This explains the differences in nowcasting performance of the MF-BVAR relative to this survey across information assumptions documented in section **3**.

Figure 8: Forecast Origins and Timing of Information for the Nowcasts of Quarter Q_t under Baseline Timing Assumption.



A.6 A Schorfheide and Song (2015) Inspired Model

Section 3 documents considerable forecasting gains by expanding the number of monthly indicators from the small-scale model to our benchmark. For completeness, we consider a specification that includes the series used in Shorheide and Song (2015). The monthly indicators in this MF-BVAR are Industrial Production, monthly PCE, Hours, the Federal Funds rate, 10 year Treasury Bond yield and the S&P 500 index. The quarterly series correspond to GDP, Fixed Investment and Government Expenditures. All data are real-time and the evaluation is carried out using the *Baseline Timing* assumption, which entails that Industrial Production, Hours and the Unemployment rate are vintage lagged. The estimation and evaluation samples are identical to those described in section 1. Hyperparameters for this specification are obtained with our two step procedure, and we report results for both three and six lags. It is important to strongly emphasize that we do not claim this specification replicates the results in Shorheide and Song (2015), particularly given differences in samples and selected hyperparameters. Instead, we include this specification to further note the gains that accrue from considering a larger set of monthly indicators than is standard in the literature owing to our novel real-time dataset.

Table 6 reports the RMSFE gains with our benchmark model relative to the three and six lag specifications using the data just described. Results are fairly similar across these lags and convey large and statistically significant gains for our benchmark MF-BVAR at all horizons, both for y/y and q/q growth rates of GDP.

A.7 Quarterly Conditional Forecasts

Aligning the information set for the monthly variables across the mixed frequency and quarterly models demands attention to details given the staggered nature of releases. To illustrate this point, the left panels in table 9 present the flow of information in our MF-BVAR for three

Table 6: Percentage Gains in RMSFE of benchmark Relative to Schorfheide-Song inspired dataset

	Three Lags		**Six Lags**	
Horizon	Y/Y	Q/Q	Y/Y	Q/Q
0	10.74 **/••/††/‡‡	16.05 ***/•••/†††/‡‡‡	12.45 **/••/††/‡‡	18.54 **/••/††/‡‡
1	30.16 ***/••/††/‡‡	28.91 **/••/††/‡‡	29.32 ***/•••/†††/‡‡‡	26.24 ***/•••/†††/‡‡‡
2	47.18 ***/•••/†††/‡‡	31.42 **/••/††/‡‡	45.94 ***/•••/†††/‡‡‡	29.70 ***/•••/†††/‡‡‡
3	48.26 ***/•••/†††/‡‡‡	23.50 **/••/††/‡‡	46.74 ***/•••/†††/‡‡‡	20.16 ***/•••/†††/‡‡
4	48.495 ***/•••/†††/‡‡‡	21.16 ***/•••/†††/‡‡‡	46.76 ***/•••/†††/‡‡‡	19.99 ***/•••/†††/‡‡‡

Notes: Entries in this table correspond to percentage gains in RMSFE for GDP growth at forecast horizons 0–4 quarters ahead (rows) for our benchmark MF-BVAR in levels with optimal hyperparameters set to maximize the marginal data density and 3 lags. The alternative model uses our vintage data for the same series included by Schorfheide and Song (2015), also in levels, with optimal hyperparameters. Three and six lags versions of this alternative specification are considered. Percentage gains are reported for both quarter-over-quarter (Q/Q) and year-over-year (Y/Y) growth rates. All evaluations use the Third Release of GDP to compute RMSFE, with the alternative model specification in the denominator of the ratio. Positive values indicate gains relative to the alternative specification. */•/†/‡ denote statistical significance from one-sided Diebold and Mariano [1995] (*/•) and Harvey et al. [1997] (†/‡) tests using standard Normal and Student's t critical values at the 15, (**) 10, (***) and 5 percent level, respectively. HAC variances were computed with the Bartlett kernel and lag length equal to four months for the nowcast and an additional three months for each forecast horizon.

representative series that cover the three timings of staggered data releases in our dataset: 1 month delay, 2 months delay, and 3 months delay. As an illustration, Panel A shows the information available for the second quarter's first forecast origin (Release 1/R1), which corresponds to the May Blue Chip Survey. Under our *Baseline Timing Assumption*, the index of activity constructed by the Institute for Supply Management (ISM) is the only series in our dataset for which last month's reading in calendar time (April) is available. All remaining series have at least a further one month delay in publication. In the case of PCE (also IP, among others), for example, only the March number is known by the beginning of May. In turn, Real Manufacturing and Trade Sales (RMTS) has a three-month publication delay so the latest available observation at that point in time is for February.

The right panels in table 9 show the corresponding data availability for the quarterly BVAR. In designing an equivalent information set, we adopt the series-specific rule that the information for a quarter is used only if all monthly readings for that quarter are available. This is why for the same forecast origin we treat ISM as missing for the current quarter despite having the April reading (Panel B). By the same token, first quarter data for RMTS is also missing since the March number is not yet known. In contrast, since all three months of PCE are available, the first quarter average of this series is included.

As a result, generating nowcasts and forecasts with the quarterly BVAR requires conditioning on the staggered flow of information. In the first forecast origin example, for instance, we run the Kalman filter of the quarterly model to complete the first quarter data for RMTS and other variables with a three-month publication delay in calendar time. The inferred state then becomes the jumping point for the nowcast, in this case a one step ahead prediction, and subsequent forecasts.

As the set of monthly indicators becomes complete, we update the information accordingly for the quarterly model. Consider the third forecast origin in Panels E and F for the mixed

frequency and quarterly models, respectively. By July, all readings of the ISM index are available for Q2, so this information is included in the quarterly model. In this case, the nowcast is no longer a one-period ahead prediction, but instead is equal to the smooth state obtained with the Kalman filter at the end of the sample.

While computationally more involved, we believe that conditional forecasting puts the two models on more equal footing. An alternative would have been to assume that the quarterly nowcast is always a one-step ahead forecast and hence to disregard all monthly data for the current quarter. At the other extreme, in going from a monthly to a quarterly model we could have "plugged" the information with all available monthly data for the current quarter, even if incomplete. For instance using the single month of PCE as a plug for Q2 in the second forecast origin (June), and so on. However, under this approach the nowcast in the quarterly model never requires a one-step ahead prediction since the only missing values correspond to the quarterly NIPA series. We view both of these alternative assumptions as extreme, as they disregard the staggered nature of data releases.

Figure 9: Data availability in the mixed frequency and quarterly models

Mixed Frequency BVAR

Panel A: First Forecast Origin (May)

Quarter	Month	ISM	PCE	RMTS	GDP
Q1	Feb	Y	Y	Y	R1
Q1	Mar	Y	Y	N	
Q2	Apr	Y	N	N	N

Panel C: Second Forecast Origin (June)

Quarter	Month	ISM	PCE	RMTS	GDP
Q1	Feb	Y	Y	Y	R2
Q1	Mar	Y	Y	Y	
Q2	Apr	Y	Y	N	N
Q2	May	Y	N	N	

Panel E: Third Forecast Origin (July)

Quarter	Month	ISM	PCE	RMTS	GDP
Q1	Feb	Y	Y	Y	R3
Q1	Mar	Y	Y	Y	
Q2	Apr	Y	Y	Y	
Q2	May	Y	Y	N	N
Q2	June	Y	N	N	

Quarterly BVAR

Panel B: First Forecast Origin (May)

Quarter	ISM	PCE	RMTS	GDP
Q1	Y	Y	N	R1
Q2	N	N	N	N

Panel D: Second Forecast Origin (June)

Quarter	ISM	PCE	RMTS	GDP
Q1	Y	Y	Y	R2
Q2	N	N	N	N

Panel F: Third Forecast Origin (July)

Quarter	ISM	PCE	RMTS	GDP
Q1	Y	Y	Y	R3
Q2	Y	N	N	N

Notes: An example of all three months and forecast origins for second quarter nowcasts. Data available? **Y(es)** or **N(o).** Table shows data availability for three different series in all three forecast origins, using Q2 as an example. ISM is the indicator published by the Institute for Supply Management, PCE corresponds to Personal Consumption Expenditures, and RMTS to Real Manufacturing Trade and Sales. For instance, in May (panel A) the April number for the ISM is known, March is the last release for PCE and the most recent RMTS is for February; in this month the first release of Q1 GDP is known as well. In going from a mixed frequency (left panels) to quarterly model (right panels) monthly data are aggregated to a quarter only if all months in the quarter are available. This is why the first quarter number for RMST is missing in May (Panel B) but available in June (Panel D) once the March number completes the quarter (panel C).

Working Paper Series

A series of research studies on regional economic issues relating to the Seventh Federal Reserve District, and on financial and economic topics.

The Urban Density Premium across Establishments **WP-13-01**
R. Jason Faberman and Matthew Freedman

Why Do Borrowers Make Mortgage Refinancing Mistakes? **WP-13-02**
Sumit Agarwal, Richard J. Rosen, and Vincent Yao

Bank Panics, Government Guarantees, and the Long-Run Size of the Financial Sector:
Evidence from Free-Banking America **WP-13-03**
Benjamin Chabot and Charles C. Moul

Fiscal Consequences of Paying Interest on Reserves **WP-13-04**
Marco Bassetto and Todd Messer

Properties of the Vacancy Statistic in the Discrete Circle Covering Problem **WP-13-05**
Gadi Barlevy and H. N. Nagaraja

Credit Crunches and Credit Allocation in a Model of Entrepreneurship **WP-13-06**
Marco Bassetto, Marco Cagetti, and Mariacristina De Nardi

Financial Incentives and Educational Investment:
The Impact of Performance-Based Scholarships on Student Time Use **WP-13-07**
Lisa Barrow and Cecilia Elena Rouse

The Global Welfare Impact of China: Trade Integration and Technological Change **WP-13-08**
Julian di Giovanni, Andrei A. Levchenko, and Jing Zhang

Structural Change in an Open Economy **WP-13-09**
Timothy Uy, Kei-Mu Yi, and Jing Zhang

The Global Labor Market Impact of Emerging Giants: a Quantitative Assessment **WP-13-10**
Andrei A. Levchenko and Jing Zhang

Size-Dependent Regulations, Firm Size Distribution, and Reallocation **WP-13-11**
François Gourio and Nicolas Roys

Modeling the Evolution of Expectations and Uncertainty in General Equilibrium **WP-13-12**
Francesco Bianchi and Leonardo Melosi

Rushing into the American Dream? House Prices, the Timing of Homeownership,
and the Adjustment of Consumer Credit **WP-13-13**
Sumit Agarwal, Luojia Hu, and Xing Huang

Working Paper Series *(continued)*

The Earned Income Tax Credit and Food Consumption Patterns **WP-13-14**
Leslie McGranahan and Diane W. Schanzenbach

Agglomeration in the European automobile supplier industry **WP-13-15**
Thomas Klier and Dan McMillen

Human Capital and Long-Run Labor Income Risk **WP-13-16**
Luca Benzoni and Olena Chyruk

The Effects of the Saving and Banking Glut on the U.S. Economy **WP-13-17**
Alejandro Justiniano, Giorgio E. Primiceri, and Andrea Tambalotti

A Portfolio-Balance Approach to the Nominal Term Structure **WP-13-18**
Thomas B. King

Gross Migration, Housing and Urban Population Dynamics **WP-13-19**
Morris A. Davis, Jonas D.M. Fisher, and Marcelo Veracierto

Very Simple Markov-Perfect Industry Dynamics **WP-13-20**
Jaap H. Abbring, Jeffrey R. Campbell, Jan Tilly, and Nan Yang

Bubbles and Leverage: A Simple and Unified Approach **WP-13-21**
Robert Barsky and Theodore Bogusz

The scarcity value of Treasury collateral:
Repo market effects of security-specific supply and demand factors **WP-13-22**
Stefania D'Amico, Roger Fan, and Yuriy Kitsul

Gambling for Dollars: Strategic Hedge Fund Manager Investment **WP-13-23**
Dan Bernhardt and Ed Nosal

Cash-in-the-Market Pricing in a Model with Money and
Over-the-Counter Financial Markets **WP-13-24**
Fabrizio Mattesini and Ed Nosal

An Interview with Neil Wallace **WP-13-25**
David Altig and Ed Nosal

Firm Dynamics and the Minimum Wage: A Putty-Clay Approach **WP-13-26**
Daniel Aaronson, Eric French, and Isaac Sorkin

Policy Intervention in Debt Renegotiation:
Evidence from the Home Affordable Modification Program **WP-13-27**
Sumit Agarwal, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, Tomasz Piskorski, and Amit Seru

Working Paper Series *(continued)*

The Effects of the Massachusetts Health Reform on Financial Distress **WP-14-01**
Bhashkar Mazumder and Sarah Miller

Can Intangible Capital Explain Cyclical Movements in the Labor Wedge? **WP-14-02**
François Gourio and Leena Rudanko

Early Public Banks **WP-14-03**
William Roberds and François R. Velde

Mandatory Disclosure and Financial Contagion **WP-14-04**
Fernando Alvarez and Gadi Barlevy

The Stock of External Sovereign Debt: Can We Take the Data at 'Face Value'? **WP-14-05**
Daniel A. Dias, Christine Richmond, and Mark L. J. Wright

Interpreting the *Pari Passu* Clause in Sovereign Bond Contracts:
It's All Hebrew (and Aramaic) to Me **WP-14-06**
Mark L. J. Wright

AIG in Hindsight **WP-14-07**
Robert McDonald and Anna Paulson

On the Structural Interpretation of the Smets-Wouters "Risk Premium" Shock **WP-14-08**
Jonas D.M. Fisher

Human Capital Risk, Contract Enforcement, and the Macroeconomy **WP-14-09**
Tom Krebs, Moritz Kuhn, and Mark L. J. Wright

Adverse Selection, Risk Sharing and Business Cycles **WP-14-10**
Marcelo Veracierto

Core and 'Crust': Consumer Prices and the Term Structure of Interest Rates **WP-14-11**
Andrea Ajello, Luca Benzoni, and Olena Chyruk

The Evolution of Comparative Advantage: Measurement and Implications **WP-14-12**
Andrei A. Levchenko and Jing Zhang

Saving Europe?: The Unpleasant Arithmetic of Fiscal Austerity in Integrated Economies **WP-14-13**
Enrique G. Mendoza, Linda L. Tesar, and Jing Zhang

Liquidity Traps and Monetary Policy: Managing a Credit Crunch **WP-14-14**
Francisco Buera and Juan Pablo Nicolini

Quantitative Easing in Joseph's Egypt with Keynesian Producers **WP-14-15**
Jeffrey R. Campbell

Working Paper Series *(continued)*

Constrained Discretion and Central Bank Transparency **WP-14-16**
Francesco Bianchi and Leonardo Melosi

Escaping the Great Recession **WP-14-17**
Francesco Bianchi and Leonardo Melosi

More on Middlemen: Equilibrium Entry and Efficiency in Intermediated Markets **WP-14-18**
Ed Nosal, Yuet-Yee Wong, and Randall Wright

Preventing Bank Runs **WP-14-19**
David Andolfatto, Ed Nosal, and Bruno Sultanum

The Impact of Chicago's Small High School Initiative **WP-14-20**
Lisa Barrow, Diane Whitmore Schanzenbach, and Amy Claessens

Credit Supply and the Housing Boom **WP-14-21**
Alejandro Justiniano, Giorgio E. Primiceri, and Andrea Tambalotti

The Effect of Vehicle Fuel Economy Standards on Technology Adoption **WP-14-22**
Thomas Klier and Joshua Linn

What Drives Bank Funding Spreads? **WP-14-23**
Thomas B. King and Kurt F. Lewis

Inflation Uncertainty and Disagreement in Bond Risk Premia **WP-14-24**
Stefania D'Amico and Athanasios Orphanides

Access to Refinancing and Mortgage Interest Rates:
HARPing on the Importance of Competition **WP-14-25**
Gene Amromin and Caitlin Kearns

Private Takings **WP-14-26**
Alessandro Marchesiani and Ed Nosal

Momentum Trading, Return Chasing, and Predictable Crashes **WP-14-27**
Benjamin Chabot, Eric Ghysels, and Ravi Jagannathan

Early Life Environment and Racial Inequality in Education and Earnings
in the United States **WP-14-28**
Kenneth Y. Chay, Jonathan Guryan, and Bhashkar Mazumder

Poor (Wo)man's Bootstrap **WP-15-01**
Bo E. Honoré and Luojia Hu

Revisiting the Role of Home Production in Life-Cycle Labor Supply **WP-15-02**
R. Jason Faberman

Working Paper Series *(continued)*

Risk Management for Monetary Policy Near the Zero Lower Bound **WP-15-03**
Charles Evans, Jonas Fisher, François Gourio, and Spencer Krane

Estimating the Intergenerational Elasticity and Rank Association in the US:
Overcoming the Current Limitations of Tax Data **WP-15-04**
Bhashkar Mazumder

External and Public Debt Crises **WP-15-05**
Cristina Arellano, Andrew Atkeson, and Mark Wright

The Value and Risk of Human Capital **WP-15-06**
Luca Benzoni and Olena Chyruk

Simpler Bootstrap Estimation of the Asymptotic Variance of U-statistic Based Estimators **WP-15-07**
Bo E. Honoré and Luojia Hu

Bad Investments and Missed Opportunities?
Postwar Capital Flows to Asia and Latin America **WP-15-08**
Lee E. Ohanian, Paulina Restrepo-Echavarria, and Mark L. J. Wright

Backtesting Systemic Risk Measures During Historical Bank Runs **WP-15-09**
Christian Brownlees, Ben Chabot, Eric Ghysels, and Christopher Kurz

What Does Anticipated Monetary Policy Do? **WP-15-10**
Stefania D'Amico and Thomas B. King

Firm Entry and Macroeconomic Dynamics: A State-level Analysis **WP-16-01**
François Gourio, Todd Messer, and Michael Siemer

Measuring Interest Rate Risk in the Life Insurance Sector: the U.S. and the U.K. **WP-16-02**
Daniel Hartley, Anna Paulson, and Richard J. Rosen

Allocating Effort and Talent in Professional Labor Markets **WP-16-03**
Gadi Barlevy and Derek Neal

The Life Insurance Industry and Systemic Risk: A Bond Market Perspective **WP-16-04**
Anna Paulson and Richard Rosen

Forecasting Economic Activity with Mixed Frequency Bayesian VARs **WP-16-05**
Scott A. Brave, R. Andrew Butters, and Alejandro Justiniano