Federal Reserve Bank of Chicago

Forecasting Economic Activity with Mixed Frequency Bayesian VARs

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

REVISED
July 12, 2018

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

July 12, 2018

Abstract

Mixed frequency Bayesian vector autoregressions (MF-BVARs) allow forecasters to incorporate a large number of time series observed at different intervals into forecasts of economic activity. This paper benchmarks the performance of MF-BVARs in forecasting U.S. real Gross Domestic Product growth relative to surveys of professional forecasters and documents the influence of certain specification choices. We find that a medium-large MF-BVAR provides an attractive alternative to surveys at the medium term forecast horizons of interest to central bankers and private sector analysts. Furthermore, we demonstrate that certain specification choices such as model size, prior selection mechanisms, and modeling in levels versus growth rates strongly influence its performance.

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

*This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. 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.

†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 Central Bank Forecasting Conference at the Federal Reserve Bank of St. Louis, the European Central Bank, the Federal Reserve Banks of Chicago, Cleveland, and St. Louis as well as the Federal Reserve Board of Governors. We are particularly indebted to Domenico Giannone for helpful comments at various stages of this project. We would also like to thank David Kelley for superb research assistance. Please address correspondence to: R. Andrew Butters, Business Economics and Public Policy, Kelley School of Business, Indiana University, 1309 E. Tenth Street, Bloomington, IN, 47405, USA. E-mail: rabutter@indiana.edu. Telephone: (+1)812-855-5768.

1 Introduction

Central bankers and private sector analysts share a demand for timely forecasts of economic activity. To be both encapsulating and reflective of the most recent events, these forecasts must blend information collected from a wide array of sources and observed at different intervals. To meet these demands, a considerable amount of effort has been spent by researchers 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. By combining the flexibility of a state-space representation with Bayesian shrinkage, the mixed frequency Bayesian vector autoregression (MF-BVAR) represents a distinctly well–suited methodology for this purpose.

In a pivotal contribution to the MF-BVAR methodology, Schorfheide and Song (2015) showed how to conduct exact posterior Bayesian inference with a data-driven approach to inform the degree of shrinkage. They then documented that an 11-variable MF-BVAR with monthly and quarterly variables achieves in real-time near-term forecasting gains over a traditional quarterly BVAR and medium-term forecasting gains relative to the Greenbook forecasts prepared by the staff of the Federal Reserve Board of Governors.[1] Their work, which focused on a small system of variables, left unanswered, however, how well the MF-BVAR approach may be expected to perform in the data-rich setting typically faced by central bankers and private sector analysts, and how robust this performance might be, for example, during periods of business cycle turning points.[2] In this paper, we close this gap by answering the questions: *How well does a medium-large MF-BVAR do in forecasting economic activity? And, how stable is this relative performance over time and across different specification choices?*

To this end, we estimate a 37-variable MF-BVAR to generate monthly nowcasts and forecasts up to four quarters ahead for U.S. real Gross Domestic Product (GDP) growth. We restrict our investigation to GDP, as it represents the most encompassing measure of economic activity and is produced quarterly with roughly a four month lag that makes the MF-BVAR particularly salient and of practical appeal. Furthermore, the wide array of monthly and quarterly real activity variables included in our model represents a large, mixed frequency dataset with a staggered real-time release pattern that the MF-BVAR methodology is well-suited to accommodate. We set up this data-rich environment by combining the proprietary archives of the Chicago Fed National Activity Index (CFNAI) with other traditional real-time macroeconomic datasets.[3] Together, they provide a unique opportunity to closely replicate the large number of U.S. macroeconomic time series typically within the information set of central bankers and private sector analysts.

We begin our analysis by assessing the out-of-sample forecasting performance of our medium-large MF-BVAR for real GDP growth over the period from 2004:Q3–2016:Q1 relative to both surveys of professional forecasters and the smaller scale model of Schorfheide and Song (2015). In benchmarking our model's performance against surveys of professional forecasters, we primarily rely on the Blue Chip Consensus mean forecasts (BCC). The BCC has the advantage of being conducted on a regular monthly schedule as opposed to quarterly or on an irregular basis like other common benchmarks (e.g. the Survey of Professional Forecasters or Greenbook). Convenience issues aside, however, BCC forecasts also tend to outperform other modeling approaches in forecasting real GDP growth.[4] Encouragingly, we find that our MF-BVAR forecasts are on par with or dominate the BCC predictions. At the nearest horizons (nowcasts and forecasts one to two quarters out)

[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.

[2]Because Greenbook forecasts are only publicly available after a five-year delay, the evaluation sample of Schorfheide and Song (2015) ended in December 2004, and, consequently, did not include the Great Recession and its subsequent recovery.

[3]Information on the CFNAI 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) to ours maintained by the Federal Reserve Bank of St. Louis. However, we include in our analysis several additional data series not found in FRED-MD.

[4]Chauvet and Potter (2013) find that the BCC forecasts outperform a large range of models (e.g. univariate methods, dynamic factor methods, dynamic stochastic general equilibrium models (DSGE)), in forecasting GDP growth one and two quarters out; and Reifschneider and Tulip (2007) find that BCC GDP growth forecasts compare favorably to a number of other public and private sector forecasts (e.g. the Greenbook, the Survey of Professional Forecasters, and the Congressional Budget Office).

differences between the two are not statistically distinguishable. However, three and four quarters out gains in predictive accuracy with our model can be as large as 10-15 percent and are statistically significant. To ensure the robustness of these findings, we also compare the performance of our model against the Survey of Professional Forecasters median forecasts and obtain similar results.

Turning to the comparison with the smaller (11-variable) model of Schorfheide and Song (2015), our medium-large MF-BVAR records large, statistically significant gains across all forecast horizons that hover around twenty percent. Our findings highlight how a richer information set than previously analyzed can enhance the ability of an MF-BVAR model to predict real GDP growth. On the stability of this relative performance over time, we find that while our medium-large MF-BVAR and the smaller scale model performed quite similarly prior to the Great Recession, large (and statistically significant) performance gains accrue with our model not only during the sharp contraction of the Great Recession, but also during the ensuing sluggish recovery. We interpret this result as complementary to similar work (Carriero et al. (2016)) that has found that the informational advantages from larger data sets are concentrated during periods of greater macroeconomic volatility.

After establishing the favorable performance of our baseline MF-BVAR to these alternatives, we then deconstruct its performance gains by examining how key specification choices affect its performance. For this part of the analysis, we look at the role of model size, elicitation of priors, and stationary transformations of the data. Here, we offer the following observations:

1. **Model size**: Broadly speaking, adding more real activity variables to the MF-BVAR in general improves its forecasting performance, although the magnitude of these improvements can vary by forecast horizon. Grouping variables by National Income and Product (NIPA) categories, the inclusion of series related to Personal Consumption Expenditures and Business Fixed Investment yields systematic improvements at most forecast horizons, while the gains from adding variables concerning Residential Investment mostly accrue at longer horizons.

2. **Choice of priors**: Our approach for eliciting priors extends the data-driven approach of Schorfheide and Song (2015) to include the residual variances as arguments in maximizing the marginal data density (MDD) (Giannone et al. (2015)). We find that using the marginal data density (MDD) to select model hyperparameters, including the prior on the residual variances, delivers performance gains of roughly 20 percent at the nowcast horizon versus the more traditional/default hyperparameter values (Carriero et al. (2015), Giannone et al. (2015)) with gains of roughly 10 percent alone coming from including the residual variances in the optimization of the hyperparameters.

3. **Levels vs. Growth Rates**: Modeling the medium-large MF-BVAR in levels outperforms an analogous specification in growth rates. Performance gains occur at all forecast horizons, with statistically significant gains ranging from roughly 10 to 15 percent for the one to four quarter ahead horizons. We interpret this result as evidence of the advantage that Bayesian shrinkage offers by facilitating the inclusion of the non-stationary components of many real activity variables that are valuable for forecasting real GDP growth.

The findings in our paper are distinct from, but also necessarily complementary to, the analysis of Schorfheide and Song (2015). Our main contribution is to provide evidence that medium-large MF-BVARs are a viable approach to incorporating a wide array of information on economic activity observed at different frequencies into the regularly produced forecasts at central banks and other institutions charged with tracking the economy in real-time. In fact, we find that it is only by including this wider information set that the MF-BVAR closes the gap on professional forecasters over very short horizons while still providing superior performance at further out horizons. Combining this result with the already established dominance of the MF-BVAR relative to traditional quarterly VARs in the near term (Schorfheide and Song (2015)),[5] suggests the MF-BVAR might

[5] In section (C.2) of the appendix, we report the relative performance of the medium-large MF-BVAR to a quarterly BVAR(1) alternative of the same sized model. We find almost identical patterns as reported by Schorfheide and Song (2015) for their 11-variable model. The MF-BVAR registers large gains over the quarterly BVAR at (especially) the nowcast and one quarter ahead horizon, with very similar forecast performance between the two models at further horizons.

offer a unique balance for practitioners. By efficiently incorporating a wide array of timely information, the MF-BVAR allows a forecaster of economic activity to realize the medium term gains that a model-based approach offers without suffering in the near term from the mixed frequency nature of the information. Furthermore, by being the first to document how the performance of MF-BVARs responds to key specification choices, our analysis offers further insights for practitioners seeking to apply this methodology.

In terms of related literature, similar methods have also been proposed to deal with data observed at different intervals, including dynamic factor models (Mariano and Murasawa (2003), Mariano and Murasawa (2010), Aruoba et al. (2009)), and the Mixed Data Sampling (MIDAS) regressions first proposed by Ghysels et al. (2004) and extended to the VAR case by Foroni et al. (2013), Ghysels (2016), and McCracken et al. (2015). We contribute to this literature by offering a comparison of the MF-BVAR methodology to the surveys of professional forecasters that have been used as a benchmark for comparison in applications of many of these other methodologies (Chauvet and Potter, 2013). Similarly, our work contributes to the growing literature investigating the role of specification choices on forecasting performance for traditional, single frequency BVARs. For example, past investigations have explored the effect of model size (Bańbura et al. (2010) Koop (2013)), the choice of prior (Koop (2013), Carriero et al. (2015), Giannone et al. (2015)), and specification choices more generally (Carriero et al. (2015)). Interestingly, our work for the mixed frequency case finds a somewhat larger sensitivity to some specification choices than what has been found in these studies for the single frequency case.

The rest of the paper is organized as follows. Section 2 briefly describes our framework for estimating MF-BVARs and discusses the specific modifications that we make to the data-driven methodology of Schorfheide and Song (2015) for eliciting the priors. Next, we outline the dataset that we use for our evaluation and make explicit the associated timing of the real-time information flow and the set of hyperparameters used in section 3. The forecasting results of our medium-large MF-BVAR relative to surveys of professional forecasters and the smaller scale model of Schorfheide and Song (2015) as well as other alternative specifications are then presented in sections 4 and 5, respectively. In section 6, we conclude and offer suggestions for future work.

2 Methodology

The MF-BVAR accommodates mixed frequency data by casting its system of equations into a state-space framework modeled at the level of the highest frequency variable. Within this framework, exact posterior inference is then conducted using a Gibbs sampling procedure to handle the latent values of lower frequency variables. Bayesian shrinkage operates on both the individual dynamics of every variable and the overall co-movement among them. To keep the exposition in this section succinct, we present here only the general mixed frequency state-space system of a MF-BVAR in section 2.1 and a brief description of our data-driven approach to eliciting priors for shrinkage in section 2.2. Given that other details of our estimation procedure parallel the existing literature, we relegate their discussion to A or to the relevant references.[6]

2.1 State-Space Representation of a MF-BVAR

Consider an n-dimensional vector y_t of time series observed at monthly (high) and quarterly (low) frequencies. To accommodate the mixed frequency nature of y_t, we adopt the convention of modeling the underlying base (monthly) frequency of each time series, stacked in the partially latent vector x_t. To match the realized values of those series in y_t observed at a quarterly frequency, the corresponding elements of x_t are aggregated with the appropriate accumulator (Harvey, 1989) depending on each time series' temporal aggregation properties (see A.1). For instance, the observed level of quarterly GDP in y_t corresponds to the three-month average of the corresponding element of x_t modeled at the monthly frequency.[7] We then denote the latent elements of the state vector

[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).

[7]We adopt the convention used by Mariano and Murasawa (2003, 2010) that the monthly realizations of (log) GDP are annualized, paralleling the quarterly series observed. Consequently, the temporal aggregation of quarterly (log) GDP is an *average* of the monthly realizations. This approach involves an approximation, but facilitates the

as $\{x_t^{latent}\}_{t=1}^T$, which stacks the monthly values of the quarterly time series as well as any missing observations for the monthly time series.

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

$$x_t = c + \phi_1 x_{t-1} + ... + \phi_p x_{t-p} + \epsilon_t; \quad \epsilon_t \sim i.i.d. \;\; MVN(0, \Sigma). \tag{1}$$

where each ϕ_l is an n-dimensional square matrix containing the coefficients associated with lag l (where all are subsumed into the composite parameter Φ), and ϵ_t is an n-dimensional vector of independent and identically multivariate normally distributed errors with covariance matrix Σ. This monthly VAR can be written in companion form and combined with a measurement equation for y_t to deliver the state-space representation of the MF-BVAR, 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

$$s_t' = \left[x_t', \; \dots, \; x_{t-p}', \; \zeta_t'\right],$$

which includes both lags of the time series at the monthly frequency and ζ_t, a vector of accumulators. Each accumulator maintains the appropriate combination of current and past x_t's to preserve the temporal aggregation of any quarterly time series in y_t.

As for the system matrices, the top n rows of each transition matrix T_t concatenate the coefficients associated with each lag. Notice that even if the VAR parameters are assumed to be time-invariant (as in our empirical application), the state-space system matrices are indexed by t due to the deterministic time variation required in calculating the accumulators, ζ_t. The remaining entries of the transition matrix, T_t, correspond to either 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 these intercepts are in rows associated with each accumulator. The rest of the elements in C_t are zeros. Finally, each R_t corresponds to the natural selection matrix augmented to accommodate the accumulators in the state.[8]

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 number of observables. In particular, in those months when the quarterly series are observed, it will have the full n selection rows. For the remaining periods in which only the monthly frequency series are observed, a subset of these selection rows will be included. Furthermore, in our empirical application, towards the end of the sample not all of the monthly series will be available. Depending on the specific release schedule and publication lag of each monthly series, a further subset of the selection rows will be dropped to accommodate these "jagged edges" of the data.

2.2 Inference and Prior Distributions

Exact posterior inference in this framework is made feasible by the Gibbs sampler and concerns both the VAR parameters (Φ, Σ) and the latent elements of the state vector $\{x_t^{latent}\}_{t=1}^T$. Here, we follow Schorfheide and Song (2015) in using a two-block Gibbs sampler that generates draws from the conditional posterior distributions for the VAR parameters and latent states, both conditional on the observed data (see A.2).

To address the curse of dimensionality, we use the following informative prior distributions on (Φ, Σ) belonging to the Normal-Inverse Wishart family that preserve conjugacy:

$$\Sigma \sim IW(\Psi, d)$$

$$\Phi | \Sigma \sim N(\Gamma, \Sigma \otimes \Omega).$$

continued use of a linear state-space model. For more information, see Mariano and Murasawa (2003, 2010).

[8]See A.1 for a more complete description of the transformation from the standard VAR system to the augmented state-space system that incorporates accumulators.

Following the convention of the Minnesota prior for single frequency BVARs, the matrix Γ consists solely of zeros and ones. A small set of hyperparameters (tightness-λ_1, decay-λ_2, sum of coefficients-λ_3, and co-persistence-λ_4) then contribute to the characterization of the covariance matrix Ω, and are collected in the hyperparameter vector Λ^p. Regarding the prior for Σ, the hyperparameter matrix Ψ is assumed to be diagonal, while the degrees of freedom hyperparameter, d, is chosen such that the prior for Σ is centered at Ψ/n, where n is the total number of series in the MF-BVAR and ψ_j is the corresponding entry on the main diagonal of Ψ.[9] These non-zero elements of the diagonal matrix Ψ are stacked into the n-dimensional hyperparameter vector Λ^σ. In summary, the prior for the VAR parameters is controlled by the vector of hyperparameters $\Lambda' = [\Lambda'^p, \Lambda'^\sigma]$ of dimension $n + 4$. To operationalize the prior, we use the data augmentation, or "shrinkage through dummy observations," approach introduced by Sims and Zha (1998) and often used in the single frequency BVAR context (e.g. Bańbura et al. (2010), see A.3). We select the set of hyperparameters that maximize the marginal data density (MDD), or, explicitly,

$$\Lambda^\star = \text{argmax } P(Y_{1:T}|Y_{-p+1:0}, \Lambda),$$

where $Y_{1:T}$ represents the full history of observations of vector y_t through time period T, and $Y_{-p+1:0}$ is a pre-sample set of observations to initialize the lags of the system. The MDD has been shown to be the sum of predictive densities, and thus summarizes the model's one-step ahead out-of-sample forecasting performance (Geweke (2001)). With conjugate priors, like ours, the marginal data density, $P(Y_{1:T}|Y_{-p+1:0}, \Lambda)$, is available analytically in the case of single frequency BVARs where all data are observed.[10] This allows for using optimization routines to find $\Lambda^\star$ quickly and efficiently (Giannone et al. (2015)).

With mixed frequency data, however, the MDD must be approximated to account for the unobserved higher frequency paths of lower frequency time series and the restrictions imposed on them by temporal aggregation. This approximation 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 grids over which to evaluate each possible combination of the hyperparameters (Carriero et al. (2012)). Adding to the computational burden, and contrary to Schorfheide and Song (2015), we infer the elements of Λ^σ, which renders the use of grids infeasible even for small models.

Given these computational demands, we instead pursue a two-step approach for choosing the hyperparameters that incorporates elements of Giannone et al. (2015) and Schorfheide and Song (2015). In the first step, we gauge the general patterns of the MDD by way of an approximation. This is constructed using the monthly series and interpolated estimates of our quarterly variables from separate bi-variate MF-BVARs. It is in this step, that we leverage the analytical form of the MDD for the single frequency case to recover the optimal hyperparmeters contained within Λ^σ. Then in the second step, we build an informed grid based on this initial exploration and use the Gibbs sampler to run all possible combinations of the grid elements of Λ^p, keeping Λ^σ fixed at the values deemed optimal in the first stage. We use the modified harmonic mean to infer the correct $P(Y_{1:T}|Y_{-p+1:0}, \Lambda)$ for each combination and select the one with the highest marginal data density. Further details and discussion of this approach can be found in A.4 and A.5.

3 Real-time Data and Forecasts

Here, we describe the salient features of our baseline MF-BVAR and the real-time data used to evaluate its forecast performance. We begin in section 3.1 by discussing the choice of variables defining our baseline model and the historical vintage data that we use to assess its out-of-sample forecast performance. Next, we detail in section 3.2 the timing of our forecasts and provide rationale for why surveys of professional forecasters provide a credible benchmark to judge the MF-BVAR's out-of-sample forecast performance. Then, we describe our out-of-sample forecasting exercise and formal evaluation of point forecast accuracy in section 3.3. Finally, in section 3.4 we discuss the

[9]This scaling of the mean for the covariance matrix is consistent with the number of observations used in the implementation of the prior via dummy observations, as explained in A.3.

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

results of our data-driven procedure for selecting hyperparameters, which then are ultimately used to generate the model's forecasts.

3.1 Choice of Variables

We included in our MF-BVAR any monthly variables that fit the following four criteria. First, we compiled a list of measures of real economic activity available in real-time that have been previously used to forecast GDP. Here, we leveraged the literature on forecasting with many predictors (e.g. Bańbura et al. (2010), Koop (2013)) in providing an initial set of key indicators, such as industrial production, payroll employment, etc. Our second criteria then involved limiting our search to indicators that correspond to headline numbers within subcomponents of GDP. This was done to avoid having an over representation of series with a diverse set of disaggregations available. As an example, in the case of industrial production or payroll employment, we include the aggregate series, but not the industry breakdown. For our two final criteria, we included only those variables with long histories and for which real-time vintages existed going back at least as far as the starting point for our most comprehensive data source.

For most monthly time series, applying this selection criteria was straightforward and full real-time vintage data was already available through Haver Analytics or publicly available sources such as the St. Louis and Philadelphia Federal Reserve Banks' ALFRED database and Real-time Dataset for Macroeconomists, respectively. For a handful of series (e.g. real nonresidential and residential private construction spending and real public construction spending), however, the only way to obtain a full history of real-time vintage data was to leverage the Federal Reserve Bank of Chicago's proprietary archives of the Chicago Fed National Activity Index (CFNAI).[11] The union of these databases provides a broad scope for us to estimate models of varying size, including specifications that encompass a large number of monthly variables that are commonly available to professional forecasters when predicting U.S. GDP. As such, it represents an ideal dataset with which to evaluate how forecasts from MF-BVARs perform in a real-time data-rich environment.

To this list, we then added quarterly time series for GDP and its major subcomponents. Based on these criteria, our baseline MF-BVAR includes 37 mixed frequency variables (30 monthly and 7 quarterly) in (log) levels, unless they are already expressed as percentage rates, in which case they are divided by one hundred to retain a comparable scale. Table 1 lists each variable (and a pneumonic for reporting purposes) categorized under a major subcomponent of GDP: Personal Consumption Expenditures, Business Fixed Investment, Residential Investment, Changes in the Valuation of Inventories, Government Spending and Net Exports. In section 5.1, we will use these groupings to organize comparisons across models of different sizes and to assess the informational contribution of different NIPA components of expenditure. Finally, the last column of table 1 highlights which of our real activity variables are also included in the smaller scale 11-variable MF-BVAR of Schorfheide and Song (2015), which serves as a model-based (as opposed to survey-based) benchmark for some of our results. Additional details on the construction of each variable and the source of its vintage data can be found in B.1.

3.2 Surveys and Forecast Origins

We compare the performance of our MF-BVAR against the Blue Chip Consensus (BCC) and, as a robustness check, the Survey of Professional Forecasters (SPF). For the BCC, historical forecasts were obtained from the Haver Analytics BLUECHIP database. In the context of evaluating a monthly frequency MF-BVAR, the BCC has the advantage of being conducted on a regular monthly schedule as opposed to quarterly (e.g. the Survey of Professional Forecasters), or on an irregular basis like other common benchmarks (e.g. the Greenbook). The BCC mean forecasts also offer a credible benchmark on absolute grounds of performance, as they have historically tended to outperform other modeling approaches in forecasting real GDP growth (Chauvet and Potter (2013)) and compare favorably to other professional forecasters (Reifschneider and Tulip (2007)) on this dimension.

[11]While some of these series can also be found in ALFRED, in every case the length of time for which vintage data is available is shorter than what is covered by the CFNAI archives; and, in many instances, fails to fully cover the Great Recession and its subsequent recovery.

Table 1: Summary of MF-BVAR Variables

	Frequency	SS (2015)
Real Gross Domestic Product (GDP)	Q	x
-***Personal Consumption Expenditures (PCE)***		
Total Nonfarm Payroll Employment (PAYROLL)	M	
Civilian Participation Rate (CIVPART)	M	
Initial Unemployment Insurance Claims (UICLAIM)	M	
Civilian Unemployment Rate (UNRATE)	M	x
Aggregate Weekly Hours Worked (HOURS)	M	x
Civilian Employment (LENA)	M	
Real Personal Consumption Expenditures (PCEM)	M	x
Light Vehicle Sales (VEHICLES)	M	
Real Retail & Food Service Sales (RSALES)	M	
Real Manufacturers' New Orders Consumer Goods & Materials (MOCGMC)	M	
Personal Savings Rate (SAVING)	M	
Real Personal Income Less Transfers (RPILLT)	M	
Univ. of Michigan Consumer Expectations (CEXP)	M	
-***Business Fixed Investment (BFI)****		
Real Business Fixed Investment (BFI)	Q	
Industrial Production (IP)	M	x
Capacity Utilization (CU)	M	
Real Manufacturing and Trade Sales (RMTS)	M	
Real Manufacturers' New Orders Core Capital Goods (RORDERS)	M	
ISM Manufacturing Index (ISM)	M	
Philly Fed Manufacturing Business Outlook Index (BOISM)	M	
Real Non-residential Private Construction Spending (CONSTPV)	M	
-***Residential Investment (RES)****		
Real Residential Investment (RES)	Q	
Real Residential Private Construction Spending (CONSTPVR)	M	
Housing Starts (HOUST)	M	
Housing Permits (PERMIT)	M	
-***Changes in the Valuation of Inventories (CIV)***		
Real Private Inventories (SH)	Q	
Real Manufacturing & Trade Inventories (RMTI)	M	
(Total) Business Inventories to (Total) Sales Ratio (ISRATIO)	M	
-***Government Expenditures (GOV)***		
Real Government Expenditures & Gross Investment (GOV)	Q	x
Real Public Construction Spending (CONSTPU)	M	
Real Federal Outlays (RFTO)	M	
-***Net Exports (NX)***		
Real Exports (EXP)	Q	
Real Imports (IMP)	Q	
Trade Balance (TRADE)	M	
Real Exports of Goods (GREXP)	M	
Real Imports of Goods (GRIMP)	M	

Notes: M–monthly, Q–quarterly. 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.

$\star$-Schorfheide and Song (2015) (SS (2015)) use Fixed Investment in their 11-variable MF-BVAR, which is the combination of business fixed investment and residential investment. The four other indicators used in Schorfheide and Song (2015) not listed here include the Consumer Price Index, Federal Funds Rate, 10-year Treasury Bond Yield, and S&P 500 Index.

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 facilitates keeping track of the information set available to the Blue Chip Consensus (BCC) forecasters. The first release of any quarter's GDP becomes available 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. For a further description of our forecast timing, see B.2.

To clarify the labeling of forecast origins, we use the second quarter's GDP as an example. At the end of April, the First release of the previous quarter's GDP (Q1) is published; making first quarter GDP information available to participants in the Blue Chip survey conducted in May, 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 further out horizons. 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.[12] The same pattern applies to the forecast origin and nowcasts for other quarters.

The staggered release of the monthly variables in our dataset (see B.3) and the uncertain timing of when survey responses are submitted create ambiguity in how to best align the information set of the BCC forecasters with our MF-BVAR at the time of each survey. We follow the timing convention of the CFNAI archives in our forecasts, using whatever data was available at the time the CFNAI was constructed, which tends to be closer to the middle of the month. To then assess the sensitivity of our results to this timing assumption, we also compare our forecasts to those from the Survey of Professional Forecasters, obtained from the Haver Analytics SURVEYS database. Although this survey is conducted only once a quarter, it uses information towards the middle of the second month (February, May, etc.), falling closer in line with the production schedule of the CFNAI in these months and, thereby, providing a valuable robustness check on our findings. Following our labeling of forecast origins, we compare the SPF median forecasts to our model forecasts made in R_1.[13]

3.3 Forecast Evaluation

Our out-of-sample forecasting exercise runs from the third quarter of 2004 through the first quarter of 2016. The beginning of the sample is imposed by the availability of the real-time vintages coming from the CFNAI archives. This results in an evaluation sample of 138 forecasts, each corresponding to a different real-time vintage.[14] The first vintage covers the period from January 1973 through October 2004, with the initial four years of data used to elicit a prior for the initial unobserved states conditional on the prior means of the VAR parameters.[15] Subsequent vintages add one additional month's worth of data resulting in a recursive sample design.

For each iteration of the Gibbs sampler, forecasts for real GDP growth are generated recursively from our baseline MF-BVAR for the current month up to one year into the future. We then compare

[12]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.

[13]To further err on the side of caution, we also adopt an alternative timing assumption in which for variables typically released after the Blue Chip survey is conducted we include only the previously available vintage (e.g. the previous month's release) in the MF-BVAR's information set. This alternative timing clearly puts us at a disadvantage relative to the SPF. See B.3 for details.

[14]For the four-quarter ahead forecast horizon, we lose 12 out-of-sample observations, leading to a total sample size of 126. In addition, 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 informational advantage. Results are almost identical if this quarter is included.

[15]More precisely, the first 6 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 period, for each data vintage to account for possible historical revisions or other changes to the data.

the median forecast at each horizon to realized GDP growth, where we gauge the sensitivity of our findings to using sequential real-time NIPA releases (First, Second, and Third) for realizations of GDP. To measure forecast gains over the Blue Chip Consensus and Survey of Professional Forecasters as well as the MF-BVAR of Schorfheide and Song (2015) in section 4, we report percentage root mean squared forecast error (RMSFE) gains/losses given by,

$$RMSFEG^{h}_{b,a} = 100 * \left(1 - \frac{RMSFE^{h}_{b}}{RMSFE^{h}_{a}}\right),$$

where for horizon $h = 0, .., 4$ (in quarters) the subscripts b and a denote our baseline MF-BVAR and alternatives, respectively. As such, positive (negative) values correspond to gains (losses) in point forecast accuracy for our baseline MF-BVAR. We then use this same procedure to evaluate performance differences between alternative specifications of our baseline MF-BVAR in section 5.

The statistical significance of any differences in unconditional predictive ability between our baseline MF-BVAR and alternative specifications (including surveys) 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/loss.[16] We report p-values using Student's t critical values and further incorporate a small-sample size correction recommended by Harvey et al. (1997). Heteroskedasticity and autocorrelation consistent (HAC) standard errors are constructed for this purpose using the Bartlett kernel with bandwith set equal to 50% of the sample size as a compromise between the standard bandwidth settings of either the number of lags plus forecast horizon or sample size as suggested by Kiefer and Vogelsang (2005). Inspection of the cross-correlograms of the mean-squared error differences of our tests suggests the former bandwidth is far too short, but results are robust to using 25% of the sample size as the bandwidth.

3.4 Hyperparameter Selection

To operationalize our two-step procedure for selecting the hyperparameters $(\Lambda^p, \Lambda^\sigma)$ of our prior, we first recursively estimated a bi-variate MF-VAR with a related monthly variable not included in our baseline MF-BVAR for each quarterly variable. Each bi-variate MF-VAR uses the same lag order (three) as our baseline model and proper, but fairly uninformative, priors.[17] A complete list of these related monthly time series can be found in B.1. Treating the posterior high frequency estimates of the quarterly variables as data along with the other monthly variables of the model, we then proceeded to optimize the MDD of this generated dataset (which is known in closed form) using numerical methods. This initial optimization has the dual function of facilitating our inference for the optimal hyperparameters, Λ^σ, and characterizing the contours surrounding the other hyperparameters, Λ^p, such that an informed grid can be set up to maximize the MDD in the next step and to detect possible identification issues.

Utilizing this informed grid, we then used the Gibbs sampler to run all possible combinations of the grid of elements for Λ^p and the modified harmonic mean to infer the correct $P(Y_{1:T}|Y_{-p+1:0}, \Lambda)$ for each combination, selecting the one with the highest marginal data density.[18] Given its size (equal to the number of variables), the hyperparameter vector Λ^σ is held fixed at its first stage estimate since the use of tensor grids would be computationally infeasible. Our approach, while feasible, is still more computationally demanding than the common practice of fixing the prior means for the innovation variances to a set of estimated residual variances of auxiliary univariate autoregressive regressions. As we will show in section 5.2, including Λ^σ in the initial optimization improves predictive accuracy at shorter forecast horizons, particularly for larger models.

To facilitate both of these steps in eliciting optimal hyperparameters, we impart a set of hyper-priors (Giannone et al. (2015)). In the case of λ_1 (tightness), we use a gamma density with a mean

[16] In the comparisons involving potentially nested models (see section 5), the Diebold-Mariano tests for the statistical significance of any RMSFE differences, strictly speaking, do not apply. Motivated by the Monte Carlo evidence reported by Clark and McCracken (2011a,b), we follow the conservative approach of Carriero et al. (2012) in reporting one-sided test results.

[17] See A.4 for further details. Results are quite robust to using alternative diffuse priors, possibly reflecting the greater number of monthly as opposed to quarterly series included in our model.

[18] 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 A.5 for a more detailed discussion.

Table 2: Prior and Posterior Estimates for Hyperparameters of Baseline MF-BVAR

Λ^p	Description	04Q3-07Q2	07Q3-10Q2	10Q3-13Q2	13Q3-16Q1
λ_1	Tightness	1.33	1.33	1.33	1.32
λ_2	Decay	1.00	1.00	1.00	1.00
λ_3	Sum of coefficients	1.15	0.96	0.95	0.90
λ_4	Co-persistence	1.51	1.40	1.43	1.42
Λ^σ	Innovation standard deviations				
GDP		0.03	0.03	0.03	0.03
-***PCE***					
PAYROLL		0.01	0.01	0.01	0.01
CIVPART		0.01	0.01	0.01	0.01
UICLAIM		0.31	0.33	0.32	0.31
UNRATE		0.01	0.01	0.01	0.01
HOURS		0.02	0.02	0.02	0.02
LENA		0.01	0.01	0.01	0.01
PCEM		0.02	0.02	0.02	0.02
VEHICLES		0.35	0.34	0.36	0.34
RSALES		0.05	0.05	0.05	0.05
MOCGMC		0.09	0.09	0.09	0.10
SAVING		0.04	0.04	0.05	0.04
RPILLT		0.04	0.04	0.04	0.04
CEXP		0.37	0.39	0.40	0.41
-***BFI***					
BFI		0.10	0.10	0.10	0.09
IP		0.01	0.01	0.01	0.01
CU		0.01	0.01	0.01	0.01
RMTS		0.04	0.04	0.04	0.04
RODERS		0.29	0.29	0.29	0.28
ISM		0.21	0.21	0.21	0.21
BOISM		0.19	0.19	0.19	0.20
CONSTPV		0.16	0.16	0.16	0.18
-***RES***					
RES		0.16	0.15	0.17	0.17
CONSTPVR		0.10	0.10	0.15	0.14
HOUST		0.40	0.40	0.42	0.42
PERMIT		0.35	0.35	0.36	0.36
-***CIV***					
SH		0.06	0.05	0.06	0.06
RMTI		0.02	0.02	0.02	0.02
ISRATIO		0.08	0.08	0.08	0.08
-***GOV***					
GOV		0.04	0.04	0.04	0.04
CONSTPU		0.23	0.22	0.22	0.22
RFTO		0.32	0.33	0.35	0.34
-***NX***					
EXP		0.09	0.09	0.09	0.08
IMP		0.09	0.09	0.10	0.10
TRADE		0.11	0.14	0.17	0.18
GREXP		0.17	0.16	0.16	0.14
GRIMP		0.23	0.22	0.20	0.21

Notes: This table reports the key hyperparameters used in the estimation of our baseline MF-BVAR. We report the optimal (posterior mode) hyperparameter for each sample from our data-driven approach of maximizing the marginal data density (MDD) outlined in section 3.4.

of 3, and standard deviation of 2. For both λ_3 (sum-of-coefficients) and λ_4 (co-persistence), we use a gamma density with a mean of 0.75, and a standard deviation of 0.25. We calibrate the λ_2 hyperparameter (rate of decay for lags) to 1. Finally, for each of the innovation standard deviations in Λ^σ we use a gamma density with a mean of 1 and a standard deviation of 0.5, reflecting the wide array of volatilities in our dataset. For the elements of Λ^p our hyperprior densities encompass settings commonly found in the literature for persistent variables (equal to 5 for the tightness and 1 for all other hyperparameters), while also allowing for smaller values (i.e. less shrinkage).[19]

In the results that follow in section 4, the hyperparameters are chosen using the first vintage in our real-time dataset and then re-optimized every three years. Table 2 reports our estimates (posterior modes) of $\Lambda^\star$ used over each of the three year windows of our out-of-sample evaluation period. The optimal hyperparameters in our case exhibit less tightness (λ_1) than is usual for models in (log) levels, while the other hyperparameters within Λ^p (λ_3, λ_4) are closer to their default values of 1. The wide array of volatilities in our dataset are reflected in the range of optimal hyperparameters for Λ^σ, reported in the bottom panel of table 2 as standard deviations for each series using their pneumonic (see table 1).

Another critical aspect of the optimal hyperparameters is their relative stability over the course of our out-of-sample evaluation period. Almost all of the hyperparameters, including those in Λ^σ, experience only modest changes across the four periods in which we re-optimize. Consequently, the performance of our model's forecasts is unlikely to be significantly impacted by variation in model hyperparameters over the evaluation sample, a result that is somewhat contrary to what has been found in other BVAR contexts (e.g. Clark et al. (2016)). The implications of our hyperparameter estimates for predictive accuracy are discussed further in section 5.2.

Conditional on these hyperparameters, we then use the Gibbs sampler to estimate the MF-BVAR for each data vintage. For an initial vintage, 36 parallel chains of 4,000 draws are used with the first 2,000 draws discarded. Subsequent vintages use this posterior as an initial guess, and are each run with the same number of draws and burn-in period. Potential scale reduction factors suggest convergence and relative numerical efficiencies indicate that the sampler mixes well (Gelman et al. (2014)).

4 Comparisons to Surveys and Schorfheide & Song

In discussing the results of our baseline MF-BVAR, we begin in section 4.1 by benchmarking our MF-BVAR's forecasting performance relative to the Blue Chip Consensus and Survey of Professional Forecasters. Next, we compare its predictive accuracy in section 4.2 against the smaller scale 11-variable MF-BVAR of Schorfheide and Song (2015) and highlight the relative ability of our MF-BVAR in capturing important turning points during our sample period. Additional survey and model based forecast comparisons can be found in C.

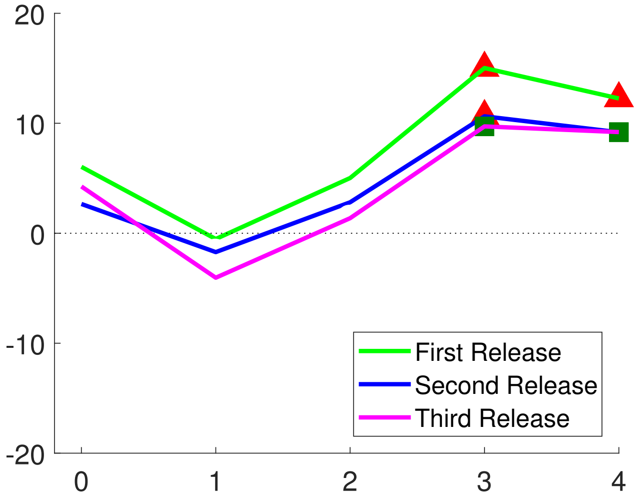
4.1 Benchmarking to Surveys

Figure 1 shows root mean-squared forecast error (RMSFE) percentage gains for our baseline MF-BVAR relative to the Blue Chip Consensus mean forecasts. We report RMSFE gains as well as indications of statistical significance at each forecast horizon out to four quarters ahead, where we pool all the forecast origins within a quarter. We then further gauge the sensitivity of our findings to using alternative real-time releases for GDP (First, Second, Third) to evaluate forecast accuracy.

The key insight from figure 1 is that *medium*-run predictions from our baseline MF-BVAR outperform the Blue Chip Consensus mean forecast in real-time over our sample period. The MF-BVAR delivers gains as large as 10 to 15 percent for quarter-over-quarter real GDP growth rates at the three and four quarter ahead horizons which are statistically significant at standard confidence levels of one-sided Diebold-Mariano tests. For shorter forecast horizons (nowcast to two quarters out), the performance of our baseline MF-BVAR is not statistically significantly different than the Blue Chip Consensus.

[19] The notation of Giannone et al. (2015) for the hyperparameters corresponds to the inverse of ours, such that an overall tightness of 4 in our context is equal to 0.25 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

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



Notes: This figure displays root mean-squared forecast error (RMSFE) gains as described in section 3.3 for our baseline MF-BVAR relative to the Blue Chip Consensus (BCC) mean forecasts for quarter-over-quarter real GDP growth over the 2004Q3-2016Q1 period at forecast horizons zero (nowcast) to four quarters ahead. We report separate results evaluated against the First, Second, and Third real-time releases of GDP. Positive values indicate gains relative to BCC. Markers denote statistical significance from one-sided Diebold and Mariano (1995) tests using Student's t critical values, the small-sample size correction suggested by Harvey et al. (1997), and HAC standard errors. (□) and (△) denote rejection of the null of equal mean-squared forecast error between the MF-BVAR and the BCC forecasts at the 10 and 5 percent level, respectively.

These results suggest that the MF-BVAR is a viable approach to nowcasting and forecasting GDP in real-time with the data flow from a wide array of real activity indicators. This is encouraging given that our insistence on using only the available real-time data vintages for these variables likely puts us at a disadvantage relative to the potentially much larger information set available to professional forecasters. Furthermore, as it is well-known, means or medians of professional surveys embed a combination of forecasts which are in general difficult to beat with a single model.

When we break down these results in table 3 by forecast origin using the Second real-time release of GDP to evaluate forecast accuracy (results are qualitatively similar using the First or Third releases instead), our findings are even more encouraging. For forecast origins $R1$ and $R2$, our baseline MF-BVAR outperforms the Blue Chip Consensus at the nowcast horizon and from two to four quarters out with roughly 5 to 12 percent RMSFE gains that are in many cases statistically significant. Only at the one quarter horizon do we find small but statistically insignificant RMSFE losses relative to BCC at all forecast origins.

When we pool across forecast origins in figure 1, the gains at the nowcast and two quarter out horizons are mitigated by losses at forecast origin $R3$. For the nowcast, in particular, this loss is large and statistically significant. It would appear then that for the nowcast to two quarter ahead horizons the $R3$ forecast origin (more appropriately called a "backcast") is where the informational constraints we impose on our baseline MF-BVAR are particularly binding. At horizons from three to four quarters out, however, the RMSFE gains we find by forecast origin remain of a similar magnitude and statistical significance to what we find in figure 1.

As a further robustness check on the informational content of our baseline MF-BVAR, figure 2 shows RMSFE percentage gains in comparison to the Survey of Professional Forecasters median forecasts. Given our timing assumptions, we evaluate against this survey using forecast origin $R1$ (second month of the quarter). Once again, we gauge how the assessment of forecast accuracy

accordingly to represent the same broad coverage of the hyperparameter domain.

Table 3: Percentage Gains in RMSFE Relative to Blue Chip Consensus by Forecast Origin

Forecast Origin	$R1$	$R2$	$R3$
Horizon			
0	5.97	$10.92^{\star}$	$-15.53^{\star\star}$
1	-0.32	-0.77	-4.48
2	7.25	4.55	-4.04
3	$11.82^{\star\star}$	$8.72^{\star\star}$	$11.41^{\star}$
4	$10.91^{\star\star}$	7.80	$8.77^{\star}$

Notes: Entries in this table correspond to root mean-squared forecast error (RMSFE) gains as described in section 3.3 for our baseline MF-BVAR relative to the Blue Chip Consensus (BCC) mean forecasts for quarter-over-quarter real GDP growth over the 2004Q3-2016Q1 period at forecast horizons zero (nowcast) to four quarters ahead by forecast origin ($R1$, $R2$, and $R3$) within a quarter. All evaluations use the Second real-time 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 using Student's t critical values, the small-sample size correction suggested by Harvey et al. (1997), and HAC standard errors. ($\star$) and ($\star\star$) denote rejection of the null of equal mean-squared forecast error between the MF-BVAR and the BCC forecasts at the 10 and 5 percent level, respectively.

varies across the three real-time releases for GDP. Our results here largely mirror what we found for the Blue Chip Consensus, with statistically significant RMSFE gains of similar size from three to four quarters out and comparable performance at shorter forecast horizons for real GDP growth.

These results are reassuring given that we already noted in section 3.2 that the timing assumption underlying our vintage data is likely to more closely align with the information set of the Survey of Professional Forecasters than the Blue Chip Consensus.[20] Finally, it is also worth noting that in comparing to either the BCC or the SPF, results are most favorable to the MF-BVAR when using the First real-time release of GDP, while almost identical with the other two. Therefore, to avoid overstating our findings, we present results for the remainder of the paper using the Second real-time release of GDP.

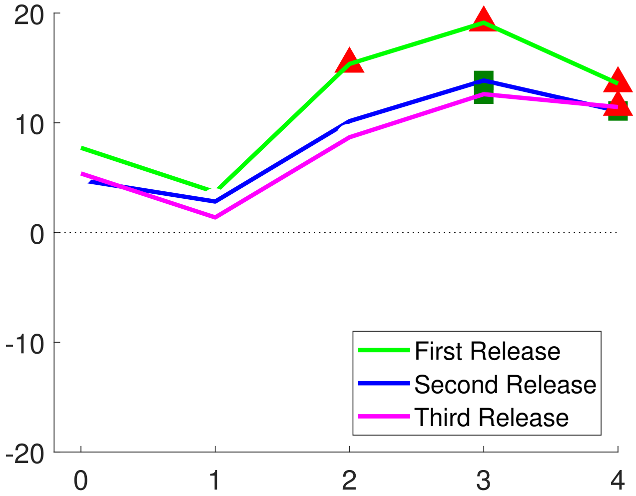
4.2 Benchmarking to Schorfheide & Song

Next, we compare the performance of our baseline MF-BVAR against the smaller scale 11-variable MF-BVAR of Schorfheide and Song (2015). We include this comparison to augment our findings relative to surveys with a model based forecast and, in particular, to emphasize the magnitude and timing of the gains that accrue from considering a larger set of monthly real activity variables in real-time than has been previously shown in the MF-BVAR context. Estimation and evaluation of both models are identical to what is described in section 3, with the hyperparameters obtained in both instances with our two-step procedure and re-optimized with the MDD every three years. The real variables of their model are listed in the last column of table 1 and are constructed as in their paper. To that list, we add the Consumer Price Index, Federal Funds Rate, 10-year Treasury Bond Yield and the S&P 500 Index in order to complete the series used in their original implementation, though due to differences in sample periods and hyperparameter selection strong differences still exist from the original implementation.

Figure 3 reports the RMSFE gains of our baseline MF-BVAR relative to the 11-variable MF-BVAR of Schorfheide and Song (2015). Across all forecast horizons, our baseline MF-BVAR outperforms their MF-BVAR by an average of 20 percent, achieving statistical significance in all cases. Thus, these results indicate that the forecast gains from our 37-variable MF-BVAR relative to surveys of professional forecasters must be manifested in part by the additionally incorporated

[20] To get at the issue of data availability in a slightly different way, C.1 repeats the results in this section using an alternative timing assumption which aims to more closely align the information set of our baseline MF-BVAR to the timing of the Blue Chip Consensus survey.

Figure 2: Percentage Gains in RMSFE Relative to Survey of Professional Forecasters



Notes: This figure displays root mean-squared forecast error (RMSFE) gains as described in section 3.3 for our baseline MF-BVAR relative to the Survey of Professional Forecasters (SPF) median forecasts for quarter-over-quarter real GDP growth over the 2004Q3-2016Q1 period at forecast horizons zero (nowcast) to four quarters ahead. We report separate results evaluated against the First, Second, and Third real-time releases of GDP. Positive values indicate gains relative to SPF. Markers denote statistical significance from one-sided Diebold and Mariano (1995) tests using Student's t critical values, the small-sample size correction suggested by Harvey et al. (1997), and HAC standard errors. (□) and (△) denote rejection of the null of equal mean-squared forecast error between the MF-BVAR and the SPF forecasts at the 10 and 5 percent level, respectively.

real activity series in our model relative to their smaller-scale model including prominent real, financial and price series,[21] where we discuss the types of real activity variables driving these information gains at various forecast horizons in the next section.

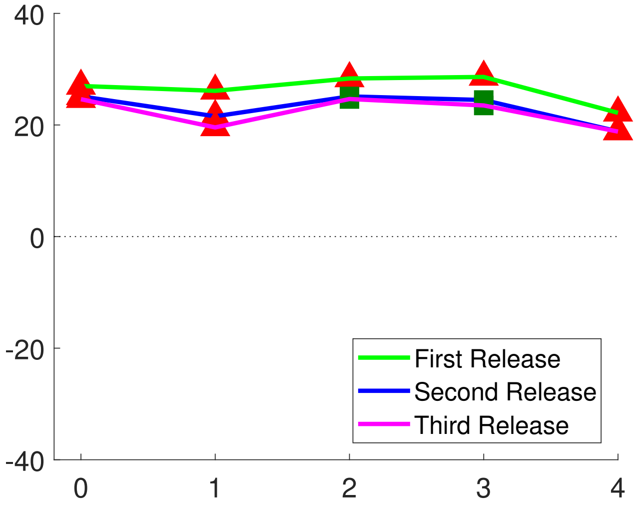
To provide further documentation of the relative performance gains of the medium-large MF-BVAR relative to the smaller scale MF-BVAR of Schorfheide and Song (2015), we also report the Fluctuation Test of Giacomini and Rossi (2009) as implemented by Rossi and Sekhposyan (2010) for each models' nowcasts. This test statistic is a re-scaled measure of the root mean squared forecast error for the two models over a rolling window. To the extent that the relative performance of the two models varies over time, it should reflect these turning points in relative performance and facilitate appropriate statistical inferences of these differences. Figure 4 displays the Fluctuation Test statistic as calculated over rolling windows of 36 months, together with the 90% confidence bands as calculated in Giacomini and Rossi (2009) using a HAC consistent estimate of the long run variance with a bandwidth size for the Bartlett kernel that corresponds to 50 percent of the forecast evaluation sample.[22] Negative values for the test statistic reflect windows where the root mean squared forecast error for the baseline MF-BVAR is lower than the 11-variable MF-BVAR of Schorfheide and Song (2015). To facilitate comparisons with the business cycle, we also shade the time periods associated with the Great Recession (as defined by NBER).

Several immediate conclusions can be drawn from figure 4. First, over any 36 month period in our evaluation period, our baseline MF-BVAR outperforms the smaller MF-BVAR. However, fluctuations in relative performance do exist between the two models over the evaluation period.

[21] In our sample, the BCC outperforms the Schorfheide and Song (2015) model from a high of 30 percent at the nowcast horizon to a low of 11 percent four quarters out, with all differences being statistically significant.

[22] The exact implementation of the Fluctuation Test of Giacomini and Rossi (2009) by Rossi and Sekhposyan (2010) uses a bandwidth selection of $T^{1/4}$. For consistency with results reported in other sections of the paper, we opted for the larger bandwidth selection, however, results are qualitatively similar for the bandwidth selection criterion used by Rossi and Sekhposyan (2010).

Figure 3: Percentage Gains in RMSFE Relative to Schorfheide and Song (2015)



Notes: This figure displays root mean-squared forecast error (RMSFE) gains as described in section 3.3 for our baseline MF-BVAR relative to the 11-variable MF-BVAR of Schorfheide and Song (2015)for forecasts of quarter-over-quarter real GDP growth over the 2004Q3-2016Q1 period at forecast horizons zero (nowcast) to four quarters ahead. We report separate results evaluated against the First, Second, and Third real-time releases of GDP. Positive values indicate gains relative to BCC. Markers denote statistical significance from one-sided Diebold and Mariano (1995) tests using Student's t critical values, the small-sample size correction suggested by Harvey et al. (1997), and HAC standard errors. (□) and (△) denote rejection of the null of equal mean-squared forecast error between the MF-BVAR and the BCC forecasts at the 10 and 5 percent level, respectively.

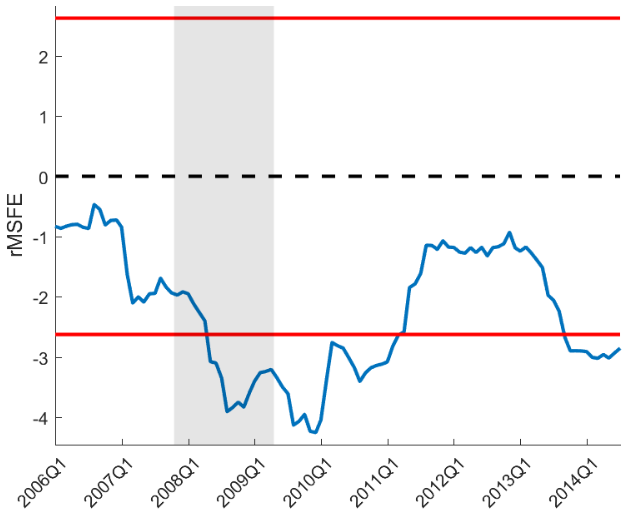
For instance, the relative gains experienced by our baseline MF-BVAR leading into the Great Recession are more modest and not statistically significant at the 90 percent confidence level as shown in figure 4. The performance of the models then diverges leading into the Great Recession and the subsequent recovery, where the baseline MF-BVAR achieves its largest relative performance gains over the smaller MF-BVAR.

Although both models experience their largest forecast errors during this period (not shown), the baseline MF-BVAR was relatively more accurate in predicting the depth of the downturn and weakness of the ensuing recovery. These results are consistent with a larger information set being more important during business cycle turning points, a result found in other modeling environments as well (Carriero et al. (2016)). However, as the figure also makes clear, these gains were not of the "one-off" variety or confined to the Great Recession, with the baseline MF-BVAR also achieving another statistically significant period of superior performance toward the end of our evaluation sample.

5 Deconstructing Performance Through Alternative Specifications

Motivated by the favorable comparison of our MF-BVAR to surveys of professional forecasters and to the smaller scale model of Schorfheide and Song (2015), we next examine the sensitivity of this performance to various specification choices faced by practitioners seeking to apply this methodology. To do so, we deconstruct our baseline MF-BVAR's forecast performance by comparing it against several alternative specifications including (i) a range of smaller model sizes (section 5.1) and (ii) a default set of hyperparameters for the degree of Bayesian shrinkage instead of our

Figure 4: Fluctuation Test of Giacomini and Rossi (2009): Baseline MF-BVAR vs. Schorfheide and Song (2015)



Notes: This figure shows for the nowcasts of the baseline MF-BVAR and the 11-variable MF-BVAR of Schorfheide and Song (2015) the Fluctuation Test test statistic (blue solid line) of Giacomini and Rossi (2009), expressed as re-scaled relative mean-squared forecast errors $\hat{\sigma}^{-1}m^{-1/2}rMSFE_t$, for a rolling window size of $m = 36$ months using a HAC estimate of the long-run variance $\hat{\sigma}$. The red solid lines correspond to 90% bands for testing the null hypothesis of equal forecast performance between the two models. Negative values of the test statistic indicate periods in which the baseline MF-BVAR outperforms the model of Schorfheide and Song (2015). Shaded regions correspond to U.S. recessions as defined by the National Bureau of Economic Research (NBER).

Table 4: Percentage Gains in RMSFE Relative to Alternative Model Sizes

Model	SS-7	+PCE	+BFI	+RES	+CIV	+GOV
Number of Variables	(7)	(17)	(23)	(27)	(30)	(32)
Horizon						
0	24.69⋆⋆	15.51⋆⋆	5.63⋆⋆	8.56⋆⋆	2.90⋆	2.51
1	16.98⋆⋆	8.98⋆	5.19⋆	1.28	−0.06	0.26
2	18.15⋆⋆	9.69⋆⋆	7.65⋆⋆	4.06⋆⋆	1.06	2.46
3	17.73⋆⋆	13.59⋆⋆	11.66⋆⋆	6.06⋆⋆	1.77⋆	4.00⋆⋆
4	12.75⋆⋆	13.33⋆⋆	10.01⋆⋆	4.29⋆⋆	−0.72	1.69

Notes: Entries in this table correspond to root mean-squared forecast error (RMSFE) gains as described in section 3.3 for our baseline 37-variable MF-BVAR relative to alternative model sizes for quarter-over-quarter real GDP growth over the 2004Q3-2016Q1 period at forecast horizons zero (nowcast) to four quarters ahead. The smallest alternative model that we consider (SS-7) is based on the seven real activity variables found in the MF-BVAR of Schorfheide and Song (2015). Additional alternative models are then labeled according to NIPA expenditure conventions, with iterative additions to SS-7 by Personal Consumption Expenditures (PCE), Business Fixed Investment (BFI), Residential Investment (RES), Changes in the Valuation of Inventories (CIV), and Government Expenditures (GOV) variables, leaving Net Exports (NX) variables as the only omitted subcomponent of GDP in our baseline MF-BVAR. All models are specified in levels with optimal hyperparameters set to maximize the marginal data density and three lags. All evaluations use the Second real-time release of GDP to compute RMSFE. Positive values indicate gains relative to the alternative model sizes. Markers denote statistical significance from one-sided Diebold and Mariano (1995) tests using Student's t critical values, the small-sample size correction suggested by Harvey et al. (1997), and HAC standard errors. (⋆) and (⋆⋆) denote rejection of the null of equal mean-squared forecast error between our baseline MF-BVAR and alternative models at the 10 and 5 percent level, respectively.

optimally chosen values and a model specified in growth rates as opposed to levels (section 5.2).

5.1 Model size

Our baseline MF-BVAR was designed to include a comprehensive set of real activity variables encompassing NIPA expenditure categories. But given the computational demands of larger systems like ours, it is reasonable to question whether or not this is necessary or if considerably smaller models informed by a few variables could perform just as well. To answer this question, we proceed in two steps. First, we identify a suitable "small-scale" model to serve as the counterpoint for our baseline MF-BVAR.[23] Second, we iteratively expand the size of this model by incorporating an additional set of variables from one of the subcomponents of GDP. By parsing the model size comparisons along sub-components of GDP, we identify heterogeneous gains for certain types of variables across forecast horizons. While the ordering of subcomponents used here is arbitrary, qualitatively the results that we report hold across alternative orderings as well.

Table 4 reports the gains in RMSFE with our baseline 37-variable MF-BVAR relative to each of the smaller scale models obtained by sequentially expanding the set of real activity variables according to the expenditure categories in table 1. Entries in the table are almost uniformly positive, suggesting that forecast performance gains accrue as model size increases. The first column (SS-7) begins with the seven real activity variables in the MF-BVAR of Schorfheide and Song (2015), listed in the second column of table 1, and includes some of the most widely cited and readily available indicators of activity: industrial production, personal consumption expenditures, nonfarm payroll employment, and the unemployment rate. Still, our baseline MF-BVAR delivers

[23] One could also interpret this "small-scale" model comparison as providing reassurances that the performance differences between the 11-variable model of Schorfheide and Song (2015) and the medium-large baseline MF-BVAR already reported were not driven by the additional price and financial series included in that comparison that might generate noise in forecasts for economic activity.

large and statistically significant RMSFE gains at all horizons, ranging from 13 to 25 percent.

The magnitude of these gains makes clear that the favorable performance of our model relative to surveys of professional forecasters documented in the previous section stems almost entirely from the information embedded in the additional real activity indicators not typically considered in smaller scale models. Interestingly, though, the gains that we find from these indicators are heterogeneous across the different subcomponents of GDP.

To begin to uncover the variables responsible for these performance gains, the second column (+PCE) of the table adds ten additional variables related to the labor market and personal consumption expenditures. While the MF-BVAR again records statistically significant gains at all forecast horizons, comparing across adjacent columns reveals that this additional information reduces the gap in performance more so at shorter horizons (nowcast through two quarters ahead) than for medium-term horizons (three and four quarters ahead). Similarly, the further inclusion of six business fixed investment variables (+BFI) further shrinks the gains of the baseline MF-BVAR at shorter horizons with smaller effects at horizons of two quarters or more.

It is not until the addition of variables related to residential investment (+RES) that the gap in forecast performance with our baseline MF-BVAR begins to appreciably shrink at medium-term horizons as well. This occurs, however, at the same time as a small deterioration in performance for the nowcast. Inspecting the forecast paths from this and previous models suggests that data on residential investment helps better capture the slow recovery following the Great Recession while adding some noise to short-run predictions. Expanding the model to cover inventories (+CIV) restores some of the performance gains at the nowcast horizon lost by the model with the additional residential investment indicators, and essentially delivers forecasts that are very similar to our baseline MF-BVAR at other forecast horizons. Finally, the addition of variables related to government expenditures (+GOV) does not add much in the way of additional performance gains, which speaks to the contribution of the real exports and imports variables (the last category that recovers our baseline).[24]

5.2 Priors and stationary transformations

The estimation strategy to this point has involved a data-driven methodology for selecting hyperparameters by maximizing the MDD (see section 3.4). This approach has the advantage of maximizing the one-step ahead forecast performance of the model (Geweke (2001)). How this approach performs along different forecast horizons is less clear. Furthermore, computational costs in optimizing the MDD are quite substantial in models with latent variables. Since a set of "default" hyperparameters are available in the Bayesian forecasting literature and have been shown to perform well in single frequency VARs, this begs the question: how large are the gains from a data-driven approach to choosing hyperparameters for a medium-large sized model in the mixed frequency setting?

To answer this question, we estimate two alternative specifications of our baseline MF-BVAR (each with the same variables in (log) levels). Both specifications keep the decay hyperparameter (λ_2) fixed at 1, just as in the baseline. In contrast, the innovation variances collected in the vector Λ^σ are no longer included in the optimization of the MDD. Instead, they are chosen as is typical in the literature by using univariate auto-regressions with sample data.[25] In general, the optimal hyperparameters for the residual variances tend to be quite similar to values implied by the traditional univariate auto-regression approach, at least for the monthly variables which do not suffer from an aggregation issue (see footnote 25). While estimates of the optimal hyperparameters for the residual variances come in both above and below the default values, the average difference between the two across all the monthly series and across each of the four re-optimized vintages is 8 percent. Consequently, any differences in forecast performance arising from optimizing the residual variances are likely to manifest primarily from optimizing the residual variances of the

[24]Clearly, for instance, the size of the gap and statistical significance of RMSFE gains with the baseline relative to the last column is dependent on the excluded category. If Net Exports were ordered prior to CIV, making the latter the excluded category, all entries would be positive and statistically significant.

[25]For monthly variables, we run AR(3) models on the first vintage's estimation sample. For quarterly variables, we use the estimated residual variance of an AR(1) at a quarterly frequency and scale it by three. For all variables, the residual variables are scaled accordingly by the number of series entering into the BVAR to reflect the implementation of the prior.

Table 5: Alternative Hyperparameters for MF-BVAR

		Specification		
Λ^p	Description	Fixed	Hybrid	Optimal
λ_1	Tightness	5	1.54	1.33
λ_2	Decay	1	1	1
λ_3	Sum of coefficients	1	2.00	1.15
λ_4	Co-persistence	1	1.92	1.51
Λ^σ	Innovation variances	1st-vintage AR(3)	1st-vintage AR(3)	MDD

Notes: This table reports the hyperparameters of two alternative specifications used in the estimation of the MF-BVAR together with the optimal hyperparameters from the first vintage of the baseline MF-BVAR. Λ^p collects the hyperparameters for the Normal prior of the autoregressive parameters, while Λ^σ reports the method for obtaining the diagonal entries of the prior mean for the Inverse Wishart, i.e. the prior individual innovation variances. **Fixed**: Λ^p set to default values in the literature, innovation variances fixed values estimated with an AR(3) using data from the first vintage. **Hybrid**: Λ^p estimated with the marginal data density (MDD) using the same priors as in table 2, innovation variances estimated with an AR(3) using data from the first vintage. **Optimal**: reproduced from table 2 (first column), both $(\Lambda^p, \Lambda^\sigma)$ estimated with the MDD. For all the cases, posterior mode estimates are reported using the first data vintage.

quarterly series, which of course the single frequency Bayesian literature has yet to suggest a default hyperparameter setting for a *monthly* based VAR.

In the first specification, labeled *Fixed* in table 5, the hyperparameters are set to the default values used in Carriero et al. (2015) and Giannone et al. (2015) in the context of single frequency BVARs. In the second specification, labeled *Hybrid*, we optimize over the hyperparameters corresponding to the *tightness* (λ_1), *sum of coefficients* (λ_3), and *co-persistence* (λ_4), using the same hyperpriors as in section 3.4. Posterior estimates of the hyperparameters under the *hybrid* approach are reported in table 5. For comparison purposes, the last column of the table reproduces the hyperparameters of the baseline model reported in table 2 (first column) for the first vintage of the evaluation sample.

The RMSFE gains of the baseline MF-BVAR relative to these two settings of the hyperparameters are shown in table 6. The results reveal that a data-driven method for choosing hyperparameters yields considerable improvements in the accuracy of nowcasts, with a substantially smaller role for other forecast horizons. Relative to the *Hybrid* case, the baseline specification records a 12 percent gain in forecast accuracy at the nowcast horizon when the innovation variances are included in the optimization of the MDD. In contrast, the gains climb to 21 percent relative to a specification that does not use the MDD. Comparing the hyperparameters in table 5 suggests that the increases in accuracy at the nowcast horizon with the baseline are likely tied to the hyperparameter on the tightness of the prior, which at 1.33 is considerably lower than the value of 5 that is customary in the literature and slightly lower than the 1.54 in the *Hybrid* case (see A.5 for more discussion).

Finally, in both cases at further out horizons the prior has less influence on relative forecasting performance, apart from a small, but statistically significant, gain at the three quarter ahead horizon for the baseline relative to the *Hybrid* settings. Here, it seems that the slightly lower values that we estimate for the co-persistence and sum of coefficients hyperparameters help to improve forecast performance when the innovation variances are chosen according to the MDD. In this case, both hyperparameters are also much closer to the default values in the literature, hence the reason why differences in forecast accuracy at further out horizons are so small between our baseline and the *Fixed* settings.

The co-persistence and sum of coefficients hyperparameters inherently address the level of integration and cointegration in the MF-BVAR, which has been necessary so far because all of the MF-BVAR's non-stationary variables are modeled in levels (or log levels). However, given that the

Table 6: Percentage Gains in RMSFE Relative to Alternative Specifications

	Specification		
	Levels		Growth Rates
Hyperparameters	Fixed	Hybrid	Optimal
Horizon			
0	20.71**	12.02**	9.13
1	1.12	−0.24	10.01**
2	−0.72	1.58	12.96**
3	0.69	4.19*	14.78**
4	−0.75	3.87	9.75*

Notes: Entries in this table correspond to root mean-squared forecast error (RMSFE) gains for forecasts of quarter-over-quarter real GDP growth over the 2004Q3-2016Q1 period as described in section 3.3 from our baseline MF-BVAR in levels with optimal hyperparameters set to maximize the marginal data density relative to the alternative *Fixed* and *Hybrid* prior specifications for the hyperparameters in table 5 and the same model specified in growth rates instead. All evaluations use the Second Release of GDP to compute RMSFE for quarter-over-quarter real GDP growth at forecast horizons zero (nowcast) to four quarters ahead. Positive values indicate gains relative to the alternative prior or data transformation specifications. Markers denote statistical significance from one-sided Diebold and Mariano (1995) tests using Student's t critical values, the small-sample size correction suggested by Harvey et al. (1997), and HAC standard errors. ($\star$) and ($\star\star$) denote rejection of the null of equal mean-squared forecast error between our baseline MF-BVAR and the alternative prior specifications at the 10 and 5 percent level, respectively.

ultimate forecast of interest is the *growth* rate of real GDP, it would also be natural to work with a stationary specification in growth rates. We estimate this alternative specification by including the same variables (and lags) as our baseline MF-BVAR, but transforming all of the time series to growth rates (log first differences or first differences for percentage variables) instead of levels.

To choose hyperparameters with the marginal data density for this alternative model, the prior must be modified in order 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 is selected with centers that shrink the individual first autoregressive lags toward zero for all series, as is customary for stationary variables. In auxiliary runs, the sum of coefficients prior was revised to be centered at 0 but the optimal value for this hyperparameter, λ_3, came in routinely at zero. Consequently, this form of shrinkage was shut down by setting $\lambda_3 = 0$. As with our baseline MF-BVAR in levels, all the other hyperparameters (including the Λ^{σ}) are elicited using the MDD.

The last column of table 6 reports RMSFE percentage gains with our baseline MF-BVAR in levels relative to the same model specified in growth rates. The overall message from this table is that there are substantial gains to be had from working in levels to address the non-stationarity of the data as opposed to transforming the data. On average across all of the forecast horizons that we consider, modeling in levels outperforms the growth rate specification by about 9 to 15 percent, with improvements being statistically significant at all but the nowcast horizon. We interpret these results as underscoring the importance of the prior specification, as it is the use of Bayesian inference and shrinkage on the underlying dynamics of each variable that facilitates capturing the non-stationary behavior common to many indicators of economic activity.

6 Conclusion

In this paper, we documented the performance of a MF-BVAR in a data-rich environment for near to medium-term forecasts of U.S. real GDP growth. We found that a medium-large MF-BVAR compares favorably to surveys of professional forecasters in terms of predictive accuracy, especially in the medium term. We also provided evidence that this favorable performance can be linked to the inclusion of a broad set of real activity variables. Furthermore, other modeling choices inherent to the MF-BVAR framework, such as the elicitation of priors and the decision to model in levels vs. growth rates each have sizeable impacts on the forecasting performance of the MF-BVAR.

Our findings suggest that MF-BVARs are a viable approach to incorporating the wide array of information observed at different frequencies into the regularly produced forecasts at central banks and other institutions charged with tracking the economy. Future work should seek to determine the scalability of these findings to other key series of interest to central bankers and private sector analysts (e.g. inflation). Additionally, it would also be valuable to provide direct evaluations of how MF-BVARs compare in predictive accuracy to other popular forecasting methods in mixed-frequency data-rich environments.

References

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

Bańbura, M., D. Giannone, and L. Reichlin (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics 25*(1), 71–92.

Brave, S. and R. A. Butters (2010). Chicago Fed National Activity Index turns ten - analyzing its first decade of performance. *Chicago Fed Letter April*(273), 1–4.

Brave, S. and R. A. Butters (2014). Nowcasting using the Chicago Fed National Activity Index. *Economic Perspectives 38*(Quarter I), 19–37.

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

Carriero, A., T. E. Clark, and M. Marcellino (2012). Common drifting volatility in large Bayesian VARs. Working Paper 1206, Federal Reserve Bank of Cleveland.

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

Carriero, A., A. B. Galvao, and G. Kapetanios (2016). A comprehensive evaluation of macroeconomic forecasting methods. EMF Research Papers 10, Economic Modelling and Forecasting Group.

Carriero, A., G. Kapetanios, and M. Marcellino (2012). Forecasting government bond yields with large Bayesian vector autoregressions. *Journal of Banking & Finance 36*(7), 2026 – 2047.

Chauvet, M. and S. Potter (2013). Chapter 3 – Forecasting output. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2, Part A of *Handbook of Economic Forecasting*, pp. 141–194. Elsevier.

Clark, T. and M. W. McCracken (2011a). Nested forecast model comparisons: A new approach to testing equal accuracy. Working paper, Federal Reserve Bank of St. Louis.

Clark, T. E., A. Carriero, and M. Marcellino (2016, June). Large Vector Autoregressions with Stochastic Volatility and Flexible Priors. Working Paper 1617, Federal Reserve Bank of Cleveland.

Clark, T. E. and M. W. McCracken (2011b). Testing for unconditional predictive ability. In M. P. Clements and D. F. Hendry (Eds.), *The Oxford Handbook of Economic Forecasting*. Oxford University Press.

Del Negro, M. and F. Schorfheide (2011). Bayesian macroeconometrics. In J. Geweke, G. Kop, and H. V. Dijk (Eds.), *The Oxford handbook of Bayesian econometrics*, pp. 293–389. Oxford University Press.

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

Doan, T., R. Litterman, and C. Sims (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews 3*(1), 1–100.

Durbin, J. and S. J. Koopman (2012, March). *Time Series Analysis by State Space Methods: Second Edition.* Oxford University Press.

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

Foroni, C., E. Ghysels, and M. Marcellino (2013). *Mixed-Frequency Vector Autoregressive Models*, Chapter 7, pp. 247–272. Emerald Group Publishing Limited.

Gelman, A., J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin (2014, November). *Bayesian Data Analysis, Third Edition (Chapman & Hall/CRC Texts in Statistical Science)* (Third ed.). London: Chapman and Hall/CRC.

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

Ghysels, E. (2016). Macroeconomics and the reality of mixed frequency data. *Journal of Econometrics 193*(2), 294 – 314. The Econometric Analysis of Mixed Frequency Data Sampling.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2004, May). The MIDAS touch: Mixed data sampling regression models. CIRANO Working Papers 2004s-20, CIRANO.

Giacomini, R. and B. Rossi (2009). Detecting and predicting forecast breakdowns. *The Review of Economic Studies 76*(2), 669–705.

Giannone, D., M. Lenza, and G. E. Primiceri (2015, May). Prior selection for vector autoregressions. *The Review of Economics and Statistics 2*(97), 436–451.

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

Harvey, D., S. Leybourne, and P. Newbold (1997, June). Testing the equality of prediction mean squared errors. *International Journal of Forecasting 13*(2), 281–291.

Karlsson, S. (2013). Chapter 15 – Forecasting with Bayesian vector autoregression. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2, Part B, pp. 791–897. Elsevier.

Kiefer, N. M. and T. J. Vogelsang (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory 21*(6), 11301164.

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

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

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

Mariano, R. S. and Y. Murasawa (2010, 02). A coincident index, common factors, and monthly real GDP. *Oxford Bulletin of Economics and Statistics 72*(1), 27–46.

McCracken, M., M. Owyang, and T. Sekhposyan (2015, October 1). Real-time forecasting with a large, mixed-frequency Bayesian VAR. Working paper, St. Louis Federal Reserve Bank.

McCracken, M. W. and S. Ng (2015, June). FRED-MD: A monthly database for macroeconomic research. Working Papers 2015-12, Federal Reserve Bank of St. Louis.

Reifschneider, D. L. and P. Tulip (2007). Gauging the uncertainty of the economic outlook from historical forecasting errors. Finance and Economics Discussion Series 2007-60, Board of Governors of the Federal Reserve System (U.S.).

Rossi, B. and T. Sekhposyan (2010). Have economic models' forecasting performance for US output growth and inflation changed over time, and when? *International Journal of Forecasting 26*(4), 808 – 835.

Schorfheide, F. and D. Song (2015). Real-time forecasting with a mixed-frequency VAR. *Journal of Business & Economic Statistics 33*(3), 366–380.

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

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

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 details on the interpolation procedure used in finding the informed grid to search for optimal hyperparameters. For the sake of clarity, some equations from within the text are reproduced below.

A.1 Building the State-Space System

We consider an n-dimensional vector y_t of macroeconomic time series of differing frequencies (e.g. some monthly and some quarterly). 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 $y_t^{'} = \left[y_t^{q'} \; y_t^{m'} \right]$ such that the first n_q elements collects the vector y_t^q of quarterly variables, such as real Gross Domestic Product (GDP), which are observed only once every three periods in a monthly model. In turn, let y_t^m be comprised solely of monthly variables, 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 $x_t^{'} = \left[x_t^{q'} \; x_t^{m'} \right]$. 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.[26] 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

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

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:[27]

$$\Delta y_t^q(i) \equiv y_t^q(i) - y_{t-3}^q(i) = \frac{x_t^q(i) + 2x_{t-1}^q(i) + 3x_{t-2}^q(i) + 2x_{t-3}^q(i) + x_{t-4}^q}{3}. \tag{5}$$

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

$$x_t = c + \phi_1 x_{t-1} + ... + \phi_p x_{t-p} + \epsilon_t; \quad \epsilon_t \sim i.i.d. N(0, \Sigma), \tag{6}$$

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

$$y_t = Z_t s_t \tag{7}$$

$$s_t = C_t + T_t s_{t-1} + R_t \epsilon_t, \tag{8}$$

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

[26] 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.

[27] The triangle accumulator is an approximate aggregation that preserves the linearity of the system. Mariano and Murasawa (2003, 2010) use this approximation in their examination of mixed frequency factor models.

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

which includes both lags of the time series at the monthly frequency and ζ_t, a vector of accumulators. For GDP, the accumulator used for the benchmark (levels) MF-BVAR is defined by equation 4, while the accumulator used for the MF-BVAR in growth rates is given by equation 5.

Given the additional variables in the state (the accumulators, ζ_t), the transition matrix is an $np+n_q$ 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.[28]

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:

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

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 a further subset of these last n_m rows will be used.

A.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 $\Theta = \{\Phi, c, \Sigma, \{x_t^{latent}, \zeta_t\}_{t=1}^T\}$. 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 $\{x_t^{latent}\}_{t=1}^T$ (of the quarterly time series as well as any missing monthly variables), and the accumulators $\{\zeta\}_{t=1}^T$ 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:

$$P(\Phi, c, \Sigma | X_{1:T}, Y_{-p+1:T}) \tag{9}$$

and

$$P(X_{1:T} | Y_{-p+1:T}, \Phi, c, \Sigma), \tag{10}$$

where we stack $\{x_t^{latent}, \zeta_t\}_{t=1}^T$ into the matrix $X_{0:T}$.

The first density, given in (9), 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 (10),

[28] See Brave et al. (2015) for further details.

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.

A.3 Shrinkage through Dummy Observations

To overcome the curse of dimensionality in equation (9), we impose prior information regarding the parameters of the VAR.[29] 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).

Four of the hyperparameters that govern these priors are collected in the vector Λ^p. 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 (6). The first two priors involving the *tightness* and *decay* primarily govern the first two moments of the set of coefficients

$$E[\phi_l(i,j)|\Sigma] = \begin{cases} \delta_i & j=i, l=1 \\ 0 & \text{otherwise} \end{cases} \quad V[\phi_l(i,j)|\Sigma] = \left\{ \left(\frac{1}{\lambda_1 l^{\lambda_2}} \frac{\sigma_i}{\psi_j}\right)^2 \right. \quad , \tag{11}$$

which shrink the VAR system toward independent random walks when $\delta_i = 1$ or white noise when $\delta_i = 0$.[30] The σ_i corresponds to the i-th diagonal element of the residual covariance matrix Σ that the moments (and prior for the Φ coefficients) are conditioned on, while the ψ_j's are the hyperparameters for this matrix and are discussed more explicitly in what follows (Giannone et al. (2015)).[31]

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.

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.[32] Second, rather than assuming that Σ is known, a prior is placed on this parameter matrix.

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 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 system that strictly adheres to the prior.

[29]The benchmark model includes 37 variables and 3 lags and consequently requires 112 coefficients to be estimated for each variable on a maximum sample size of 470 monthly observations.

[30]For the intercepts, c, we adopt a fairly diffuse prior as is customary in BVARs.

[31]We tried allowing 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 included 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. For the model in (log) levels, we set the $\delta_i = 0$ for the Real Federal Outlays (RFTO) variable given its univariate properties. Additionally, for the specification of the MF-BVAR in growth rates, all of the δ_i are set equal to zero to reflect the convention that these series are stationary in growth rates.

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

To operationalize these four different priors, we use the data augmentation approach introduced by Sims and Zha (1998) and often used in the BVAR context (Bańbura et al. (2010), Schorfheide and Song (2015)). The set of dummy observations that implements the four forms of shrinkage we consider are given by

$$Y_d = \begin{pmatrix} \lambda_1 \text{diag}(\psi_1\delta_1, \ldots, \psi_n\delta_n) \\ 0_{n(p-1)\times n} \\ \cdots \\ \text{diag}(\psi_1, \ldots, \psi_n) \\ \cdots \\ 0_{1\times n} \\ \cdots \\ \lambda_3 \text{diag}(\delta_1\bar{y}_1, \ldots, \delta_n\bar{y}_n) \\ \cdots \\ \lambda_4 \bar{y} \end{pmatrix} \quad X_d = \begin{pmatrix} \lambda_1 J_p \otimes \text{diag}(\psi_1, \ldots, \psi_n) & 0_{np\times 1} \\ & \\ \cdots & \cdots \\ 0_{n\times np} & 0_{n\times 1} \\ \cdots & \cdots \\ 0_{1\times np} & \alpha \\ \cdots & \cdots \\ (1_{1\times p}) \otimes \lambda_3 \text{diag}(\delta_1\bar{y}_1, \ldots, \delta_n\bar{y}_n) & 0_{n\times 1} \\ \cdots & \cdots \\ (1_{1\times p}) \otimes \lambda_4 \bar{y} & \lambda_4 \end{pmatrix}. \tag{12}$$

The first block corresponds to the *tightness* and *decay* components of the prior governed by λ_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 $\overline{\psi} = (\psi_1, \ldots, \psi_n)'$ has as its i-th element a non-negative number that is proportional to the residual variance for that series. The series-specific scalars δ_i reflect the center of the prior for the first order own-lag autoregressive coefficients and are set to 1 or 0. The second block ensures the prior for the residual variances is appropriately centered, while the third block 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 δ_i correspond to the centers of the prior for the first autoregressive coefficient and are set to 1 for the levels specification and 0 for the growth rates specification. Finally, λ_4 controls beliefs regarding the *co-persistence* of the system.

Regarding the covariance matrix, Σ, it can be shown that these dummy observations combined with an improper prior of the form $P(\Sigma) \propto |\Sigma|^{-\frac{n+1}{2}}$ imply an Inverse-Wishart prior density centered at $\frac{\Psi}{n}$, where the diagonal matrix Ψ is fully characterized by the vector $\overline{\psi}$ described above. This is because the degrees of freedom resulting from multiplying that prior and the "likelihood" of the dummy observations is given by $T^* - k$, where T^* is the number of dummy observations and is equal to $k + n + 1 + n$ and $k = np + 1$. As such, the mean of the IW is equal to

$$\frac{\Psi}{T^* - k - n - 1} = \frac{\Psi}{n}.$$

This suggests caution should be exercised if one wishes to center the prior of the residual variances at residual variance estimates of previously estimated univariate autoregressions, which must be scaled by n when using the full complement of priors. It is also noteworthy that if not all four priors are active then the mean of the implied inverse-Wishart prior may not even be well defined.

A.4 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 3.4). Consequently, to find the optimal hyperparamters a grid search is required (Carriero et al. (2012), Schorfheide and Song (2015)). Before this grid is constructed, we infer the optimal hyperparameters (available in analytical form) from an "approximate" marginal likelihood. This approximation involves interpolating each of the quarterly time series using a bi-variate MF-BVAR estimated with a related series that is not in the baseline model (e.g. Manufacturing Industrial Production for real GDP). We set the number of lags to three and specify loose priors for the VAR parameters by setting

$$\lambda_1 = 0.3, \lambda_2 = 1.0, \lambda_3 = \lambda_4 = 0,$$

with prior variances centered at estimates from individual autoregressive processes. The interpolated estimate of each quarterly series corresponds to the mean of the monthly elements of the latent state vector, subject to a temporal aggregation constraint. Treating these estimates as data, one proceeds to optimize the hyperparameters using numerical methods (initialized at 96 independent random starting values from an over-dispersed grid relative to the prior). Alternative settings of the prior, e.g. λ_1 set to 0.1 or 0.5, did not have any material effects on the hyperparameters inferred for this initial approximation.

A.5 Contours of the Approximate and Correct Marginal Likelihood

Exploring the general contours of the "approximate" marginal data density described in the previous section 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 $P(Y_{1:T}|Y_{-p+1:0}, \Lambda)$, 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 correct marginal of the MF-BVAR, as it does not account for the latent states and the temporal aggregation constraint imposed on them. Nonetheless, since it is orders of magnitude easier to compute and maximize, it can help in guiding the initialization 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 5 and 6 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 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. This is the reason why we find the first step in our approach so valuable. Second, both marginal data densities are sharply peaked around the optimal value of λ_1, which is around 1.3 in both figures. In contrast, both densities are considerably flatter in λ_3, though the peaks once again coincide around 1. Third, 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.

Finally, it is important to note that the contours of the correct data density in figure 6 were obtained by averaging ten independent runs of the Gibbs sampler for each element of the grid. Otherwise, the surface will be considerably noisier than the interpolated one, reflecting the simulation error from the Gibbs sampler estimate.

Figure 5: Contours of the "Approximate" Marginal Data Density (Conditional on Interpolated data)

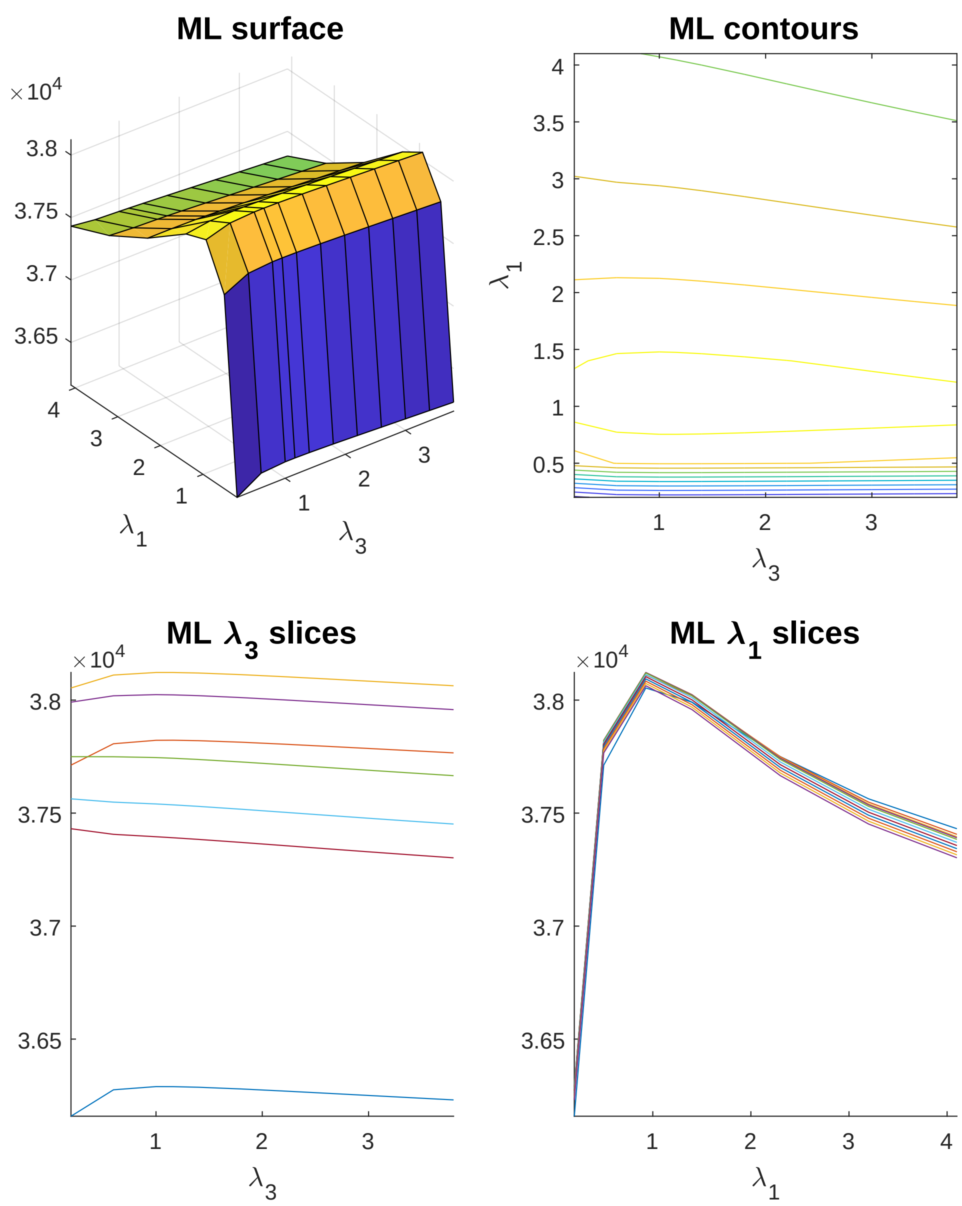
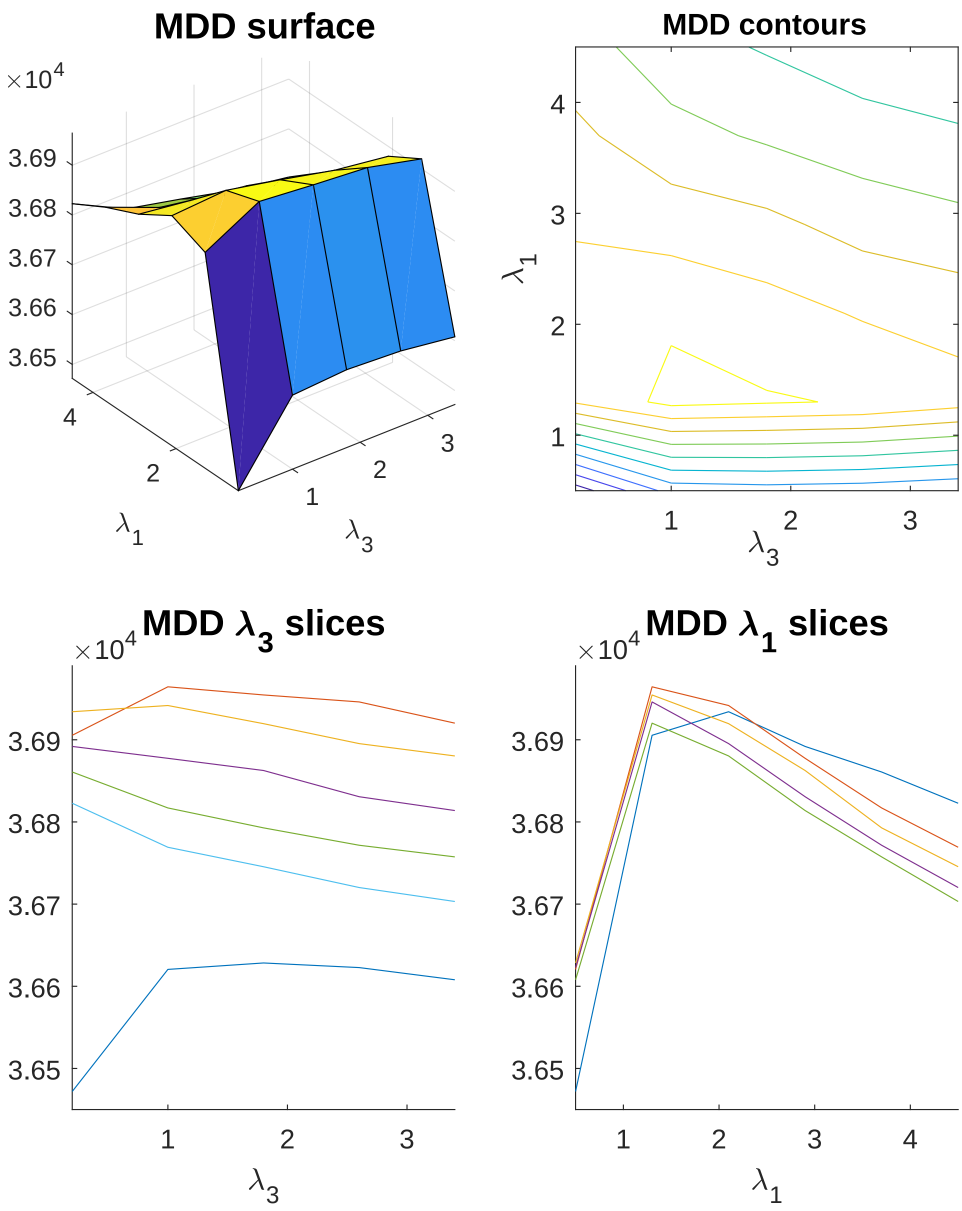


Figure 6: Contours of the MF-BVAR Marginal Data Density Obtained via Modified Harmonic Mean and Gibbs Sampler



B Real-time Data and Forecasts: Additional Details

B.1 Vintage Data

Figure 7 lists the 37 variables in our baseline MF-BVAR by subcomponent of GDP along with the 7 related series that we use in our interpolation model for the quarterly variables. ALFRED, Haver Analytics, or CFNAI mnemonics are shown for each series with the exception of the Philly Fed Manufacturing Business Outlook Survey. The source of this data is shown in the Notes column of the table. The Notes column also indicates for CFNAI Archives variables whether or not vintage data is available prior to the late-2004 starting point for our sample in either ALFRED or the Haver ASREPGDP database. In some instances, vintage data from ALFRED were either spliced onto similar data from Haver Analytics to extend their time series back further in time well prior to the beginning of our forecast sample or were deflated using Haver ASREPGDP deflators. This is also noted for these series in the Notes column of the table.

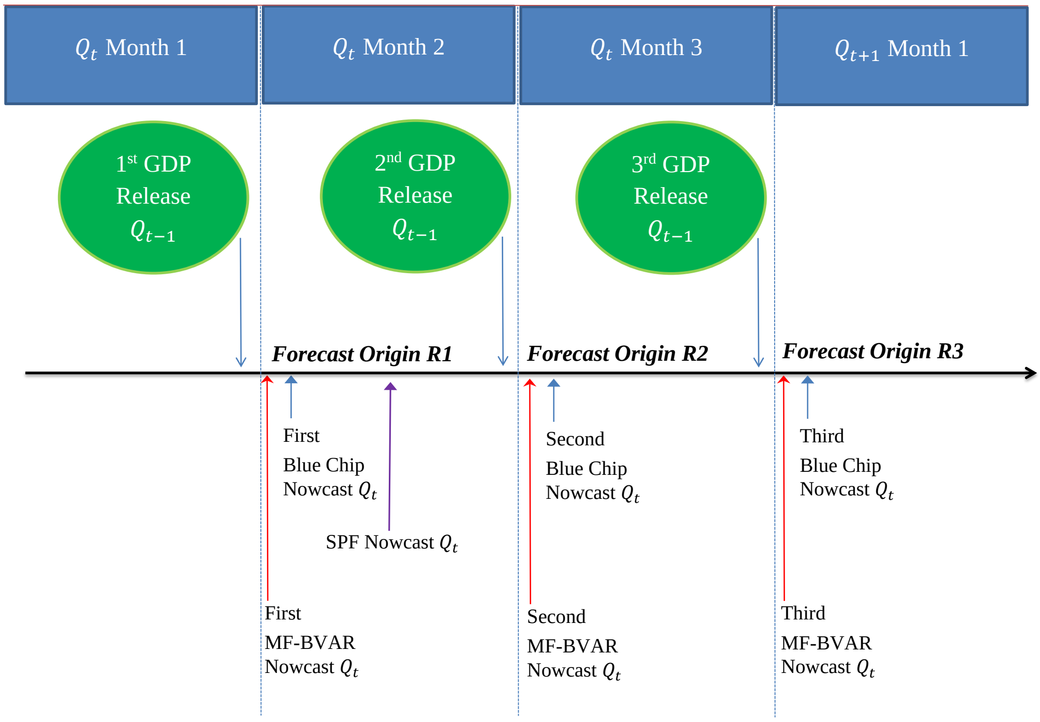
B.2 Forecast Timing

Figure 8 details our labeling of forecast origins and the timing of the professional surveys that we compare our MF-BVAR forecasts against for a generic quarter, Q_t. The top of the figure reports calendar time (e.g. Month 1 is January for the first quarter). At the end of each month NIPA releases for the previous quarter become available and are, hence, available to Blue Chip Consensus (BCC) survey respondents at the beginning and Survey of Professional Forecasters (SPF) survey respondents in the middle of the next month. This is how we index forecast origins for the nowcasts and forecasts of our MF-BVAR.

Figure 7: MF-BVAR Variables and Related Series

	Series Name	ALFRED	Haver Analytics	CFNAI Archives	Notes
GDP	Real Gross Domestic Product	GDPC1	GDPH@ASREPGDP		
	Total Nonfarm Payroll Employment	PAYEMS			
	Civilian Participation Rate	CIVPART			
	Initial Unemployment Insurance Claims			LCUN	ALFRED vintages start in 2009.
	Civilian Unemployment Rate	UNRATE			
	Aggregate Weekly Hours Worked	AWHI			
	Civilian Employment	CE16OV			
PCE	Real Personal Consumption Expenditures	PCEC96	CBHM@ASREPGDP		Spliced at times with CBHM from the CFNAI archives.
	Light Vehicle Sales	ALTSALES			Spliced with TLVAR@USECON prior to 1976.
	Real Retail & Food Service Sales	RSAFS/CPIAUCSL			Spliced with TRST@USECON prior to 1992.
	Real Manufacturer's New Orders of Consumer Goods & Materials			MOCGMC	
	Personal Savings Rate	PSAVERT			
	Real Personal Income Less Transfers			YPLTPMH	ALFRED vintages start in 2009. Haver ASREPGDP is missing vintages.
	Univ. of Michigan Consumer Expectations		U0M083@BCI		Final values are not revised and not seasonally adjusted.
	Real Business Fixed Investment	PNFIC1	FNH@ASREPGDP		
	Industrial Production	INDPRO			
	Capacity Utilization	TCU			
BFI	Real Manufacturing & Trade Sales			TSTH	ALFRED vintages start in 2013.
	Real Manufacturer's New Orders of Core Capital Goods	NEWORDER			Spliced with MOCNX@USECON prior to 1997. Deflated with JFNE@ASREPGDP.
	ISM Manufacturing Index			NAPMC	
	Philly Fed Manufacturing Business Outlook Index				From Philly Fed website. Seasonally adjusted and put on an ISM basis.
	Real Nonresidential Private Construction Spending			CONSTPV	ALFRED vintages start in 2011 (TLNRESCONS). Deflated by JFNS@ASREPGDP.
	Real Residential Investment	PRFIC1	FRH@ASREPGDP		
RES	Real Residential Private Construction Spending			CPVR	ALFRED vintages start in 2011 (TLRESCONS). Deflated by JFR@ASREPGDP.
	Housing Starts	HOUST			
	Housing Permits	PERMIT			
	Real Private Inventories	CBIC1			Accumulated over time to obtain a level matching SH@USNA.
CIV	Real Manufacturing & Trade Inventories			TITH	ALFRED vintages start in 2013.
	(Total) Business Inventories to (Total) Sales Ratio	ISRATIO			
	Real Government Expenditures & Gross Investment	GCEC1	GH@ASREPGDP		
GOV	Real Public Construction Spending			CONSTPU	ALFRED vintages start in 2011 (TLPBLCONS).
	Real Federal Outlays		(FTBO+FTOFF)@USECON		Unrevised but seasonally adjusted and deflated by JG@ASREPGDP.
	Real Exports	EXPGSC1	XH@ASREPGDP		
	Real Imports	IMPGSC1	MH@ASREPGDP		
NX	Trade Balance	BOPGSTB			Spliced with TMBCA@USECON prior to 1997.
	Real Exports of Goods	BOPGEXP			Spliced with TMXCA@USECON prior to 1997. Deflated by JX@ASREPGDP .
	Real Imports of Goods	BOPGIMP			Spliced with TMMCA@USECON prior to 1997. Deflated by JM@ASREPGDP.
	Manufacturing Industrial Production			IPMFG	ALFRED vintages start in 2007.
	Real Manufacturers' New Orders of Nondefense Capital Goods	MOCNC			ALFRED vintages start in 2011 (ANDENO).
	Total Business Inventories		NTI@USECON		
Related Series	New Single-family Home Sales	HSN1F			
	Government Payroll Employment	USGOVT			
	Exports of Goods	BOPGEXP			Spliced with TMXCA@USECON prior to 1997.
	Imports of Goods	BOPGIMP			Spliced with TMMCA@USECON prior to 1997.

Figure 8: Forecast Origins and Timing of Information for the Nowcasts of Quarter Q_t



B.3 The Real-time Data Flow

The staggered release of our real-time data throughout the months of a quarter leads our real-time vintages to have a "jagged edge" quality to them. This feature of our dataset is summarized in figure 9, using the example forecast timing shown in figure 8 when Q_t Month 1 is January. Under the timing assumption used in the text, the Blue Chip Consensus mean and Survey of Professional Forecasters median forecasts for February of Q1 real GDP growth and some monthly variables (Type I) through January are already observed (marked by "X") at forecast origin R1. Quarterly variables, however, are only available through calendar quarter Q4, as are what we call Type II monthly variables. In addition, there are a handful of monthly variables that have yet to be released with full coverage of Q4 (noted by their $\varnothing$ marker in Q4 M12 (December)). All of these variables must be projected forward in time to March to obtain a Q1 forecast. This pattern is then maintained as we roll through the calendar months from February to March and April.

Figure 9: The Real-time Data Flow

Forecast Origin R1: February

Release Date		Survey Forecasts		Monthly Indicators			Quarterly Indicators
		BCC	SPF	Type I	Type 2	Type 3	NIPA
Q4	M11			X	X	X	Ø
Q4	M12			X	X	Ø	X
Q1	M1	Feb/Q1	Feb/Q1	X	Ø	Ø	Ø

Forecast Origin R2: March

Release Date		Survey Forecasts		Monthly Indicators			Quarterly Indicators
		BCC	SPF	Type I	Type 2	Type 3	NIPA
Q4	M11			X	X	X	Ø
Q4	M12			X	X	X	X
Q1	M1	Feb/Q1		X	X	Ø	Ø
Q1	M2	Mar/Q1		X	Ø	Ø	Ø

Forecast Origin R3: April

Release Date		Survey Forecasts		Monthly Indicators			Quarterly Indicators
		BCC	SPF	Type I	Type 2	Type 3	NIPA
Q4	M11			X	X	X	Ø
Q4	M12			X	X	X	X
Q1	M1	Feb/Q1	Feb/Q1	X	X	X	Ø
Q1	M2	Mar/Q1		X	X	Ø	Ø
Q1	M3	Apr/Q1		X	Ø	Ø	Ø

C Results: Additional Comparisons

C.1 Additional Survey Comparisons

To err on the side of caution, we also adopt an alternative timing assumption in which for variables typically released after the Blue Chip Consensus survey is conducted 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 the MF-BVAR's information set rather than the most recent one. The only difference in the real-time data flow, shown in figure 9, that results from this change is that some monthly variables go from being classified as Type I to being Type II (e.g. Industrial Production, Capacity Utilization, Real Retail & Food Service Sales, Real Federal Outlays, Initial Unemployment Insurance Claims, Housing Starts & Permits, and the Philly Fed Manufacturing Business Outlook Index) or Type II to Type III (e.g. Business Inventories to Sales Ratio) instead.

Table 7 reports RMSFE gains under this alternative timing assumption for our baseline MF-BVAR relative to the Blue Chip Consensus mean and Survey of Professional Forecasters median forecasts pooled across forecast origins and using all three (First, Second, Third) real-time releases of GDP to evaluate forecast accuracy. The near-term comparisons to both surveys are generally weaker under this timing assumption, although the differences are not always statistically significant. In contrast, the RMSFE gains at medium-term forecast horizons noted in the text are of a similar magnitude (10 to 15 percent) and remain statistically significant.

C.2 Additional Model Comparisons

As an alternative benchmark model for comparison, we follow Schorfheide and Song (2015) and estimate a Quarterly BVAR using our baseline 37-variable specification. In particular, we estimate a VAR(1) on quarterly real-time data over the same evaluation period (2004:Q3-2016:Q1) described in the main text to construct nowcasts to four-quarter ahead forecasts of real GDP growth. We average the real-time vintages of any monthly variables to the quarterly frequency for this purpose, such that any within-quarter information is not used in the construction of nowcasts.

To select the hyperparameters of our Quarterly BVAR, we follow the procedure described in Giannone et al. (2015), using the gamma density with mean 0.32 and standard deviation of 0.20

Table 7: Percentage Gains in RMSFE Relative to BCC and SPF under Alternative Timing

Survey	BCC			SPF		
GDP Release	First	Second	Third	First	Second	Third
Horizon						
0	−6.25	−9.22	−6.38	−10.28	−10.20⋆	−10.36⋆⋆
1	−9.38⋆⋆	−8.20⋆⋆	−9.08⋆⋆	−4.51	−3.38	−3.94
2	5.41	3.79	2.83	12.39⋆	7.90	6.24
3	14.49⋆⋆	9.88⋆	8.62⋆	17.83⋆⋆	12.52⋆	11.03
4	11.30⋆⋆	8.08⋆	7.78⋆	11.76⋆	8.98	8.84

Notes: Entries in this table correspond to root mean-squared forecast error (RMSFE) gains as described in section 3.3 for our baseline MF-BVAR relative to the Blue Chip Consensus (BCC) mean and Survey of Professional Forecasters (SPF) median forecasts for quarter-over-quarter real GDP growth at forecast horizons zero (nowcast) to four quarters ahead across real-time GDP releases (First, Second, Third). Positive values indicate gains relative to BCC or SPF. Markers denote statistical significance from one-sided Diebold and Mariano (1995) tests using Student's t critical values, the small-sample size correction suggested by Harvey et al. (1997), and HAC standard errors. (⋆) and (⋆⋆) denote rejection of the null of equal mean-squared forecast error between the MF-BVAR and the BCC or SPF forecasts at the 10 and 5 percent level, respectively..

Table 8: Percentage Gains in RMSFE Relative to Quarterly BVAR

GDP Release	First	Second	Third
Horizon			
0	42.80	39.00	37.75
1	12.22	7.23	4.52
2	−3.98	−1.45	−2.21
3	2.26	2.16	2.18
4	1.52	1.60	1.02

Notes: Entries in this table correspond to root mean-squared forecast error (RMSFE) gains as described in section 3.3 for our baseline MF-BVAR relative to the Quarterly BVAR described in section (C.2) for quarter-over-quarter real GDP growth at forecast horizons zero (nowcast) to four quarters ahead across real-time GDP releases (First, Second, and Third). Positive values indicate gains relative to the Quarterly BVAR.

for the tightness (λ_1) hyperprior, calibrating the decay hyperparameter to 1, and using the gamma density with mean 1.08 and standard deviation of 0.30 for both the sum-of-coefficients (λ_3) and co-persistence (λ_4) hyperpriors; while the hyperprior for the residual variances is set to the inverse gamma density with scale and shape parameters both equal to $(0.02)^2$. To match the selection of hyperparameters for our MF-BVAR, we update the hyperparameter estimates for the Quarterly BVAR at identical three-year intervals, selecting the values which optimize the MDD.

Results for this exercise are shown in table 8. Perhaps not surprisingly given its exclusive use of within-quarter information, the MF-BVAR displays performance gains in RMSFE of roughly 40 percent over its quarterly counterpart at the nowcast horizon. These gains then shrink noticeably to a range of 5-15 percent by the one-quarter ahead horizon and are either marginal or slightly negative for horizons from 2-4 quarters ahead. This pattern mirrors that found by Schorfheide and Song (2015). An important implication of these results in the context of those presented earlier is that the superior performance of the medium-large MF-BVAR relative to surveys at medium-term forecast horizons is also likely to hold in the traditional single frequency (quarterly) case.

C.3 Pseudo real-time evaluation (1988:Q4–2016Q1)

We also performed a pseudo real-time evaluation over the extended period of time from 1988:Q4-2016:Q1 in comparison to the smaller scale 11-variable model of Schorfheide and Song (2015). This longer out-of-sample comparison serves to further determine if the gains from the larger size of our baseline MF-BVAR relative to a smaller scale model are similar both leading up to and following the Great Recession. To create our pseudo real-time dataset, the final vintages from our real-time dataset were truncated recursively by one month going back through time. Our two-step procedure was then used for hyperparameter selection as before, with the diagonal elements of the prior innovation variance-covariance matrix fixed at the first step and hyperparameters re-optimized with the MDD every three years. Even on this extended sample, our baseline MF-BVAR still dominates the smaller scale model of Schorfheide and Song (2015), although the average gain is smaller, particularly for the nowcast. This indicates that the predictive gains with the baseline are concentrated during the Great Recession and the ensuing slow recovery as the body of the text suggests. As such, it would seem that the value in tracking the current state of the economy from including of a diverse set of real activity variables may have increased as of late.

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.

Working Paper Series *(continued)*

Working Paper Series *(continued)*

Working Paper Series *(continued)*

Working Paper Series *(continued)*